DECEMBER 2022

Computer Vision News

& Medical Imaging News

The Magazine of the Algorithm Community

Editorial

This photo was taken in peaceful, lovely and brave Odessa, Ukraine.

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editors:
Marica Muffoletto
Ioannis Valasakis

Publisher:
RSIP Vision

Contact us

Free subscription

Read Previous Magazines

Follow Us

Dear reader,

As a busy year draws to a close, I think we should all take a moment to reflect on just how much has been achieved. We've seen some awesome events and so much innovation in 2022. It is probably not less important that our community was able to meet again, be it at **CVPR**, at **MICCAI**, at **ECCV** or at any of the countless events we have joined during the year. It was such a relief... More is going to happen in 2023 and we look forward to it!

Computer Vision News has continued to grow this year, with almost 9,000 active subscribers! Our new supplement launched one year ago, **Medical Imaging News**, has found a stable place and a consistent audience of followers and readers. You can find it this month starting on page 28 with impressive new AI features for single-port robotic surgery; a wonderful report about the new definition of interpretability; and an exciting guide by Marica Muffoletto to Mayavi - a great tool for 3D visualization of scientific data in Python.

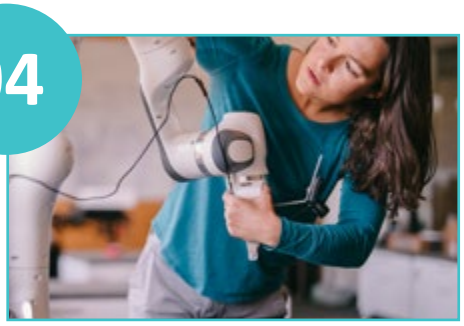
Also this month, we speak to **Chelsea Finn**, one of the raising stars in the field of machine learning and robotics. As she puts it, she's "***really fascinated by this question of how we might allow agents, including robots, to develop broadly intelligent behavior in the real world***. *I think that machine learning and interaction with the world is a key component of doing that!*" You don't want to miss her exclusive interview on page 4!

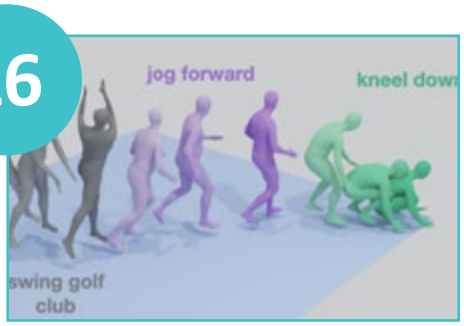
We have prepared an exclusive section of what happened at the **GeoMedIA workshop**, a MICCAI-endorsed event held less than 2 weeks ago in Amsterdam: you will read about a summary of the acclaimed **keynote speech by Emma Robinson** and the review of **GeoMorph**, the brilliant **Best Paper Award winning work**. Find that on page 34. This is not the only Best Paper that we feature this month: we continue this winning streak with Eleonora Giunchiglia, winner of the **Best Student Paper prize at IJCLR 2022** - watch her video interview on page 12.

We have much more to offer this month, so enjoy the reading, and remember to share our link with your friends and colleagues so they can **subscribe for free!**

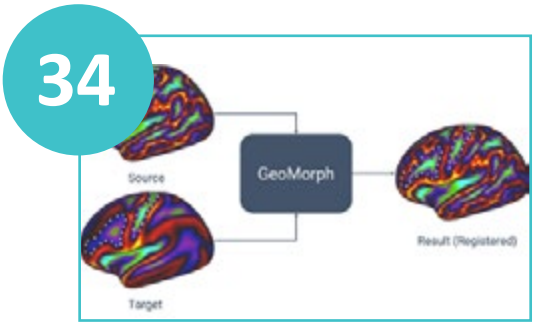
Ralph Anzarouth,
Editor, **Computer Vision News,**
Marketing Manager, **RSIP Vision**

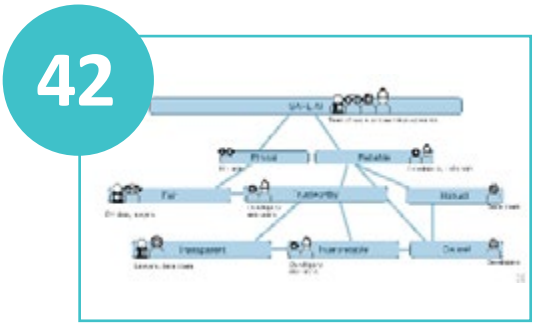
Computer Vision News





Medical Imaging News





Chelsea Finn is an Assistant Professor at Stanford University in the Computer Science and Electrical Engineering Departments. She is also a part-time Research Scientist at Google Brain.

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

Chelsea, tell us about your work!

I do research in machine learning and robotics. I'm really fascinated by this question of how we might allow agents, including robots, to develop broadly intelligent behavior in the real world. I think that machine learning and interaction with the world is a key component of doing that. My work has looked at questions that I think are important in trying to solve that problem, including the ability to quickly learn new things by leveraging previous experience rather than learning from scratch. Also, the ability to generalize broadly by using a wide variety of data and by training models to be more robust. Also, it has involved applying machine learning to real robots and seeing if they can learn manipulation skills in the real world. That entails both perception and action for robots to be able to see and to be able to act in the world.

When did you discover this passion?

It's something that kind of happens over time. I don't think there's any one moment in which you're like, *"Oh, this is the one thing that I love doing!"* I've always enjoyed solving puzzles and trying to figure out answers to questions. I think that this question of developing intelligent robots is a really fascinating one to work on. I did a little bit of robotics and computer science in middle school. I enjoyed working on it, but I didn't really think of it as something that I would dedicate my career to at that point. And then later on, when I was in college, it kind of became clear to me that computer science would open a lot of doors toward solving really interesting problems. I then dabbled a little bit in computational biology as well as other areas like robotics and computer vision and ended up really liking the problem of studying intelligence and robotics, especially robotics, because it's very much going all the way to a very tangible and real system.

You are asking questions and from time to time you are also getting answers…

Yeah, definitely.

What answer has satisfied you the most?

That's a really hard question. I think that there's been lots of work that's been quite satisfying. I feel like any research project provides some answers but also opens a lot more questions. Some of the things that I've been satisfied with are when we actually see robots doing things in the real world. In some of the early projects that I did, we were able to train robots to do tasks that are pretty basic for people but are fairly complex for robots, like being able to screw a cap onto a bottle or being able to use a patch flow to lift an object. Seeing the robot do those things was really fascinating and interesting. But at the same time, it opened up questions like, well, it can do that sort of skill in one environment. Now, can we have robots execute that sort of behavior in a variety of environments with a variety of different objects? From there, I've been kind of satisfied by results where we've been able to allow robots to do things with novel objects. There was one paper a few years ago where we would give a robot some objects, and we would tell it a goal. The goal is to move a couple of pieces of trash over to the side of a bin, and it would figure out that it can pick up this object that it has never seen before and use that to sweep the objects to the side of the bin without having being told that it

should use that object as a tool. That's one example.

Do you give more credit to the robot itself or to the people who program this robot?

All of it comes down to the software and the algorithms that you design and put into the system. We buy robots all the time, and we just buy them off the shelf. While it's very useful to be able to buy robots off the shelf, the hardware itself is not at all capable unless you have algorithms that can power it to do the kinds of things that we're trying to get them to do.

You still have decades to work on robots. What would be your dream result to see robots accomplish?

One really basic example of something that I would love to see is where you can put a robot in a kitchen that it has never been in before and tell it to make a bowl of cereal, and it will be able to make a bowl of cereal. It sounds really basic like people can do this when they're half asleep, but it involves opening up a package of cereal, opening up a fridge, opening up some milk, getting out a bowl, and pouring.

All in the right order?

Well, the right order actually isn't the hard part. The hard part is the dexterity of opening up packages and closing packages and pouring and all that and being able to do it in a way that's general with packages that maybe you haven't quite seen that exact package before or you haven't seen exactly that kitchen before.

Why should the robot close the package? The instructions were only to prepare a bowl of cereal.

Even task specification is really challenging, as you point out. I mean, I would like my robot to clean up after itself even after I give it a command. If you say to make a bowl of cereal, implicitly, you also want it to clean up after making a bowl of cereal by putting away the milk because you don't want the milk to go bad. And so, actually, even this problem of specifying what the task is to a robot is challenging in and of itself beyond the problem of actually executing that task.

I don't like to ask negative questions, but there is one that I need to ask. When you tried to make something happen, and it just didn't work, what was the most frustrating moment?

In research, frustration and failures come up all the time. I don't even think it's necessarily a negative thing. It's just kind of part of what happens, and the times in which the robot fails and the algorithms fail are the times in which you learn. In some ways, this can actually be a positive experience. In terms of examples of frustration, it's too many to count. Along with the task specification that we just talked about, there have been times in which you give a robot a demonstration that illustrates the task, and you'll try to learn a reward function underlying that demonstration, and it won't actually give you the right result. It will give you a reward function that is consistent with the behavior in that one scenario but doesn't generalize to a new scenario. There have been times in which I've worked on inverse reinforcement learning, where you try to learn rewards underlying demonstrations, and it ends up being a very challenging problem. There have been times in which we've spent months trying to do something. Another example is a lot of reinforcement learning algorithms work beautifully in a simulation where you try to have the robot learn how to run, and so it runs, and then it falls down. And so then, in simulation, you

have to have it try to run again. But in the real world, you can't just reset the robot back to where it started. If it fell down, you need to actually pick it back up. We've been trying to develop algorithms that actually allow robots to learn autonomously, where if they fall down, they can get back up. Or if they push an object into the corner of the workspace, they're able to get it back to the main part of the workspace. This is one problem that we've been looking at, and at first, it can be very tedious to train algorithms with the reset, but when we try to actually develop algorithms that work without the resets, there have been times where we will run an algorithm in simulation, and it's working beautifully in simulation. Then we put it on the robot, and it's just not working or learning at all, and it gets stuck. It's hard to tell what the difference is between the simulation, what's the learning process in simulation, and the learning process in the real world to try to understand why it's not working in the real world.

You said before that the credit goes to the software when it works. Should we also credit the software when it does not work?

Yeah, absolutely. I think that there are a number of mismatches between simulations in the real world that could be the cause. One could be that when you read camera images in simulation, you can get camera images instantaneously. Whereas in the real world, there's latency, there's a delay. If your algorithm isn't prepared to handle that delay from the camera images, then it might not work, and it'll fail in a way. It won't tell you that the camera images are delayed; therefore, it's not working. You kind of need to figure out what the issue is.

The public imagines a robot, like a humanoid with arms and a head, and eyes, which isn't always true. What kind of machines are you working on?

My lab primarily works on robotic manipulation problems, and that means that we're mostly working with arms. The robot arms that we work with, we work with some that are small, some that are larger. Oftentimes it's actually just one arm doing the task rather than having two arms, although we have some projects that are done by manual manipulation where there are two arms trying to complete the task. Typically, the arms have similar degrees of freedom where they can move around in a similar way as a person's arm. So that's primarily what the robots that we work on are, and in terms of sensors, we will usually stick a camera somewhere. It might be near the arm, or it might be opposite the arm. Sometimes we also put a camera on the wrist of the robot as well. And then we've also worked with other sensors as well, like audios, although cameras are kind of the simplest to work with, they are cheap and work well. I've also worked a little bit with legged robots before as well, but I primarily work with arms.

Cameras are like the eyes of a robot. Do you ever feel the urge to re-engineer human beings? Maybe we shouldn't have the eyes here on our head; should we have them on our arms? Have you ever thought that we would be much more efficient if we functioned like that robot?

Humans are remarkable at fine motor skills, object manipulation, and so forth. In many ways, we have evolved to be remarkable at these things because if we couldn't use our hands to do things, then we probably wouldn't survive and operate in the same way that we do. I don't have any complaints about how human hardware works because we are already so amazing

at these sorts of things. And if anything, I wish that we could have robots that have the sort of tactile sensing that people have. We have extremely high-resolution tactile sensing all over our bodies, especially on our hands. There are some tactile sensors out there that are starting to do better, but usually, they're either expensive, or they're somewhat brittle. So if you run lots of experiments, they'll degrade over time or they're very low resolution and they don't give you the very detailed feedback that humans can get. That's actually one of the reasons why for robots because we don't have these really awesome tactile sensors that humans have, we often actually just put cameras on the hand of the robot, and then you can get actually pretty nice sensory information in the same place that humans have sensory information.

Is that what you call haptics?

Yes, tactile sensing is kind of a form of haptics. But we often don't use haptics because the sensors are not quite there.

Not yet. If I understand correctly, the main problem is the hardware, and the hardware is mostly not in your hands. If I look at the whiteboard behind you, most of the work that you do is work with formulas, software, and mathematics, and not with the hardware itself. If I could give you better hardware than what you have in stock, what would that be?

The main thing that I love on the hardware side is the robots that we have right now, but ten to 100 times cheaper. That would be great! Because then we could buy lots more of them. I mean, we still already have probably more than ten robot arms in our lab, but we might be able to buy 100 if they were ten times cheaper. Then my second thing would be really good tactile sensors

that are high resolution, that don't break or change over time, and that are also fairly cheap.

What prevents the hardware from becoming cheaper or more sensitive?

The price of robots has been going down, and in many ways, that's because of motors. In the nicest robots, the motors are really kind of driving the cost of the robot. On some of the cheaper ones, there are these new motors called Dynamixel motors that are pretty nice and are a lot cheaper than the kind of motors that you find in industrial robots. But they also have their limitations. For example, a lot of the robots that we work with have parallel jaw grippers, which means they have two fingers open and close, kind of like a pincher. And the reason for that is if you want to build a more dexterous hand, like

a five-fingered hand, you need motors to control each of the joints. If you just try to stick a motor in each joint of the finger, you would have this pretty giant hand. Humans don't have motors, we are tendon-driven, and so there are also tendon-driven designs of hands, but those also tend to be rather challenging to work with. So, yeah, those are some of the limitations. Then in terms of tactile sensing, there's nothing that is computationally like skin. So the ways that people have tried to create technology that is like high-resolution tactile sensors are actually quite different from the way that skin works.

I only talked to you for a few minutes, but it sounds obvious to me that you are doing exactly what you want, and you are exactly in the place where you would like to be. How difficult was it for you to find your way and get to a point where you are really doing what you want to do? How did you get so close to what seems to be 100% fit for you?

It's a combination of hard work and a little bit of luck and hard work to create more luck as well. I found a topic that I was really excited about. I was lucky to have some mentors that were super supportive and helped me learn a lot and grow a lot, and do great work.

Can you name them?

Yeah! Starting with undergrad, my advisor Seth Teller was really passionate about research and was one of the first people who introduced me to research and showed

me why it's exciting. Then during my PhD, I worked a lot with Pieter Abbeel and Sergey Levine. I worked very closely with Sergey when I was starting off because he was a postdoc, and I learned a ton from working with him.

Would you say that your path was a continuous one?

I think it was a continuous process. There are certainly milestones. When I started my PhD, I had some papers that were very well received and some ideas that kind of worked out really well. Then also, of course, when I applied to faculty positions. But it's definitely kind of a journey; you don't ever arrive at any point.

Do you ever think, well, I should have done something else instead of this?

I'm very happy with where I'm at. I don't have any regrets. I mean, there are certainly things that I'm excited about and things I'm excited about that I'm not doing, and so there's always a balance of trying. You don't have the time to do everything. But overall, I'm very happy with where I'm at, and I don't think I would have done something different.

Can you tell me what is the biggest sacrifice you have to do in order to get yourself to this point?

[*thinks...*] I don't think that there's anything significant that comes to mind. There have been times in which I could have spent more time with friends or family or something. At the beginning of my PhD, there were times, such as paper deadlines I would lose sleep and not exercise and kind of lose... my scheduling and everything because I wasn't very good at managing and figuring out how much work it is to submit a paper. I think that over time, I figured out how to manage things such that I didn't lose a

significant amount of sleep, didn't stop exercising, and didn't stop doing stuff just for a paper deadline. So that's something, but I don't think that is necessarily a huge sacrifice. The other thing is that doing a PhD is financially somewhat of a sacrifice. I think you can make more money doing other things. It ended up working out nicely for me in the end. I was kind of fortunate enough not to have loans after my undergraduate because I got really excellent financial aid during undergrad. I think those are a couple of sacrifices. I don't think there's a big one.

Tell us one message that you would like to give the community.

I think it's important to work on what you're excited about and what you think are the most important problems. And to help kind of move the field forward the most. Always keep on trying to grow and learn, which is true for everyone. There's so much that we don't know in the world of research!

More than 100 inspiring interviews with successful Women in Computer Vision!

ROAD-R: THE AUTONOMOUS DRIVING DATASET WITH LOGICAL REQUIREMENTS

Eleonora Giunchiglia has just completed her PhD at the University of Oxford and started her postdoc at TU Vienna. Her paper seeking to create a platform to make deep learning models safer won the Best Student Paper Prize at the 2nd International Joint Conference on Learning & Reasoning (IJCLR) in the UK in September. She speaks to us about this fascinating work.

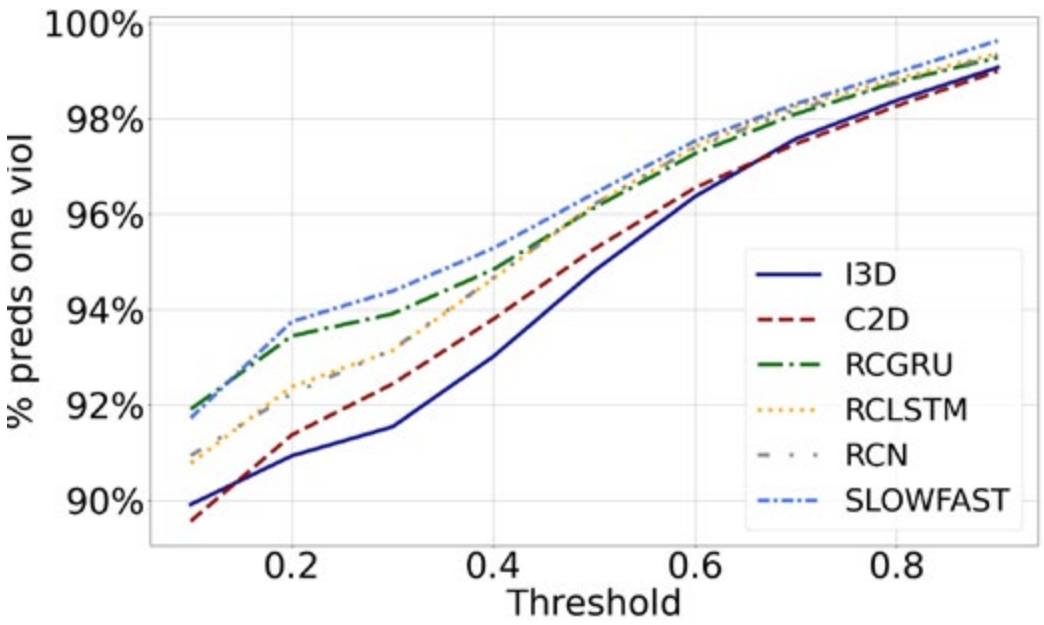
How can we make deep learning models safer? That is the question Eleonora poses in her award-winning paper. The answer, she says, is to start writing requirements for them as we do for standard software.

"*In standard software, you have a phase where you write the requirements, and then you write software compliant with the requirements,*" she explains. "*I would like to do the same for deep learning, but you need a dataset to create deep learning models with requirements. Therefore, we propose* ***the first dataset for autonomous driving with requirements***."

Eleonora and team annotated the dataset with requirements expressed as logical constraints but found that by taking a purely data-driven approach, irrespective of the thresholds used, **89% of the predictions violated the requirements**.

"*We annotated in the hundreds, so I was expecting some violations of the requirements, but maybe 15-20% –* ***when I saw 89%, I was in shock!***" she recalls. "*I remember checking the results for a week and rerunning the entire pipeline multiple times because I couldn't believe it. When working with logical constraints, most people annotate just one or two, and it's normal for a neural network to violate these requirements once or twice. With hundreds of requirements, the percentage increases naturally, but 89% was a surprise!*"

To solve this, Eleonora proposed some basic approaches **to incorporate these constraints into the training or post-processing phases** to make the deep learning models compliant with the requirements without dropping their performance.



Percentage of predictions that violate at least one requirement when varying the decision threshold. For each label, neural networks return a value in [0,1], and a label is predicted if the corresponding output is greater that the decision threshold (often such decision threshold is set equal to 0.5)

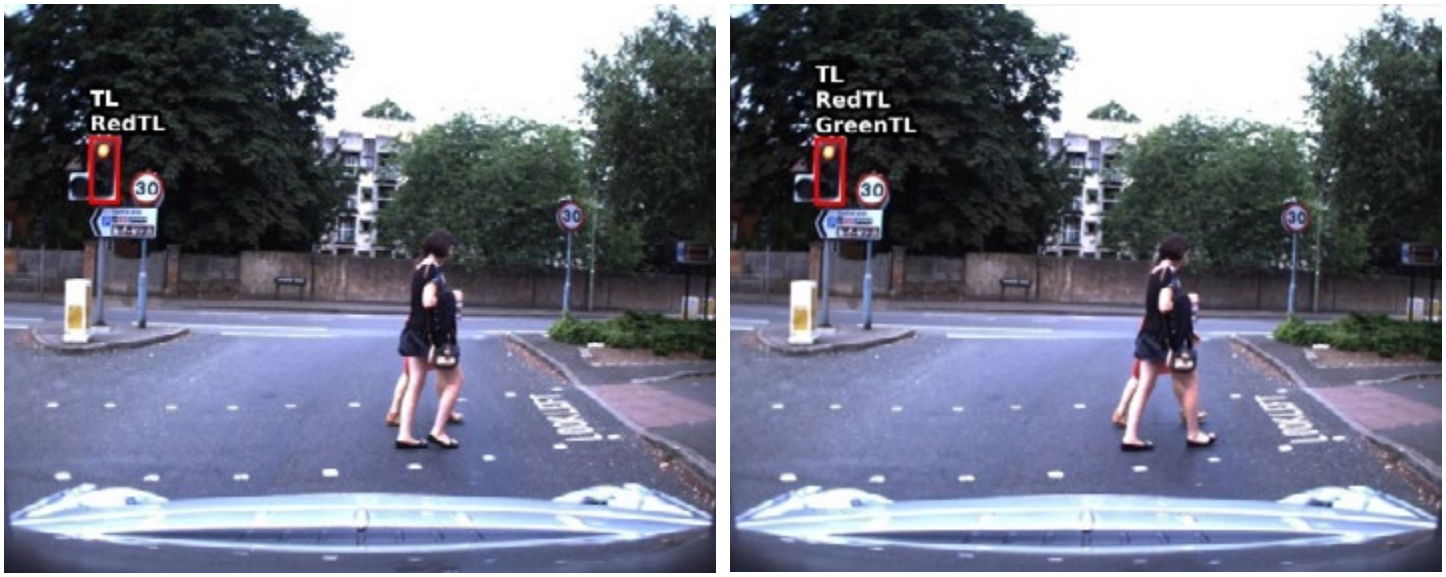
At first, she anticipated that applying the constraints in a way that minimally changed the neural network prediction would not negatively impact the performance. However, it turned out that **adding these requirements naively can hurt performance, whereas learning to use them smartly helps teach the neural network and can improve its performance**.

"Until a few years ago, it was felt that it was enough to have very high accuracy to avoid these problems," she points out. *"At conferences, people told me that if we have 99.9% accuracy, then neural networks won't make these stupid mistakes. But then* ***neural networks started to be applied to more complex problems and safety-critical scenarios where even a single error can make a big difference****."*

The basic idea behind this work grew from Eleonora's master's thesis when she was working on healthcare data and predicting survival curves. She achieved high accuracy on every performance metric, but the neural network would sometimes tell her the survival probability at time T+1 was much higher than at time T, which did not make sense from a human perspective. **She needed a way to tell the neural network it was doing something she knew was wrong**.

"The main action detection pipeline is very much rooted in computer vision," Eleonora explains. *"We use the* ***3D-RetinaNet model****. I still believe we need a data-driven, computer vision-based approach to do the basic detections. Then you can bring in some reasoning. We included the requirements in the loss function, or we used the* ***MaxSat solver*** *to correct the predictions at post-processing time. It's very much a combination of computer vision and reasoning."*

Does all this mean that the papers we were celebrating two or three years ago were built on an inaccurate basis? Not



Predictions made by the 3D-RetinaNet model for the same traffic light in two consecutive frames.

necessarily, says Eleonora, but they were built on a different assumption. Looking back at the history of AI, in the '80s and '90s, **people thought reasoning was enough**. Then neural networks came along, and **people thought data was enough**. Now, it has reached the phase where **people realize we need reasoning and learning capabilities**.

"*There is this whole neuro-symbolic wave happening,*" she continues. "*I hope my tiny contribution will be to say we need the reasoning not only to improve performance or to learn from fewer data, but we also need it for safety reasons. We can learn lessons from standard software engineering and logic and apply them in neural networks and deep learning. We don't need to reinvent the wheel!*"

Despite her modesty, this work is potentially game-changing. Does Eleonora think that is why **the jury at IJCLR 2022 recognized it with a top award**?

"*I think what was interesting for them was that we had a clear application domain,*" she responds. "*These neuro-symbolic methods are powerful, but they're still applied to toy datasets, while our work bridged the gap between theory and practice. That's all it did.* ***We said let's port this neuro-symbolic world into the real world and see how it performs on a real-world, safety-critical task.***"

Eleonora continues to work on this task in Vienna. She says the problem of incorporating hard logical constraints or requirements is understudied, and there is still much to do.

"*In the future, I hope requirement solicitation and specification for deep learning will be as normal as it is for software development,*" she tells us. "*If this message gets through, we'll have much safer models, and nobody will lose anything.*"

Gül Varol is a Research Faculty at École des Ponts ParisTech.

Gül, what is your work about?

Computer vision, first of all, but mainly videos, and particularly, humans in videos. This could be sign language applications, human motion synthesis, action recognition from video, text-to-video retrieval, and a lot of things that nowadays contain vision and language.

Do you consider that you are doing something very specific, or are you covering a wide area?

I'm covering multiple specific things, which becomes a wide area in the end! [*she laughs*]

Did you choose them?

Some happened organically, some I chose, and some are a continuation from my postdoc. Human motion synthesis is one that I determined.

Is this a passion of yours?

Yes, I found a space which is really unexplored. Of course, once you start exploring it, other people do too, so it's not as underexplored anymore. The text-to-motion synthesis was something I came up with as I felt it hadn't been done before. That's where you type some text and want to generate a 3D avatar that does the instruction, like raising your arm, for example.

For the benefit of a non-scientist like me, why does the real world need something like this?

Well, there could be many answers to that! A lot of applications are in entertainment for things like gaming and the film industry. People do motion capture in studios with expensive equipment, and collecting that is very difficult, whereas automatically generating it is cheap and free. That is the main motivation. Beyond entertainment, it could be used for any human-computer interaction, even health. We'll see where that goes.

You are dedicating your best years to this work. That can't be just so that people can play better games. Do you have another motivation?

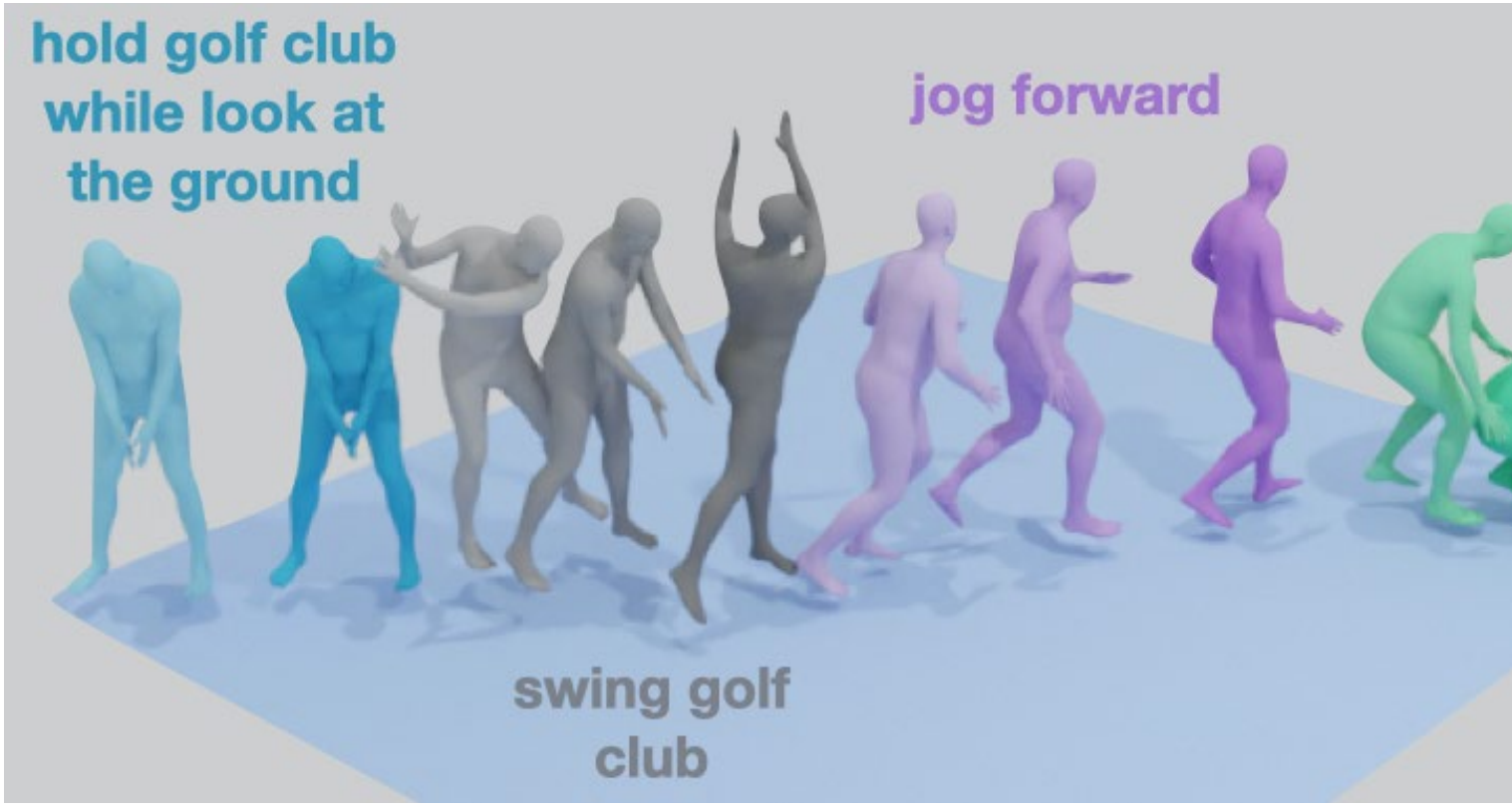
The real reason is to generate automatic training data, but that's more of a research than a real-world application.

You are a researcher and a teacher. How do you find juggling both?

My position in France is a bit non-standard. It's equivalent to Assistant Professor, but my main job is research. I do a little less teaching than most other professors, but I like it a lot. I had a three-hour lecture last night, quite late in the evening, and I was really happy about it.

I know you have students working and progressing with you. How big is your team?

I work with 7-8 people, but they're all scattered, so not necessarily in Paris. It's always co-supervised with multiple people. Basically, I'm not doing the work; they're doing the work, and I'm doing the talking! [*Gül laughs*]



What other field would you have chosen if you could not do this?

I was also thinking of becoming a math teacher, but I'm in a field related to math, and in a way, I'm a teacher, so that's not too far. At one point, I wanted to become a film director, and that does seem far now!

Considering everything you work on, how much came along organically, and how much was chosen by you?

Sign language is the thing that happened as part of my postdoc because the funding was coming from that. I hadn't determined it, but I liked it after getting into it. Human motion is from me. Video retrieval, video representations, and everything about video in general is what I'm passionate about, mainly because it's very difficult. There are so many open questions.

What is something you have learned that we may not know?

People don't know how difficult sign language is and about the computer vision problems involved with it. It's not a simple gesture recognition problem. It's a language problem. It's like any language, with all the grammar and vocabulary rules, but it also has the visual aspects, so we have to deal with video representation learning and language learning. It's considered a low-resource language for machine translation because you don't have more than 1 million sentences. Before getting into this field, I didn't realize it required know-how from so many adjacent fields. That made the problem super interesting. I would encourage people to work on it.

How long will the sign language problem take to be solved?

It might be a 50% chance in 10 years. It's

a difficult one.

We have spoken a bit about the past and the present. Where is Gül going in the future?

I'm not going anywhere! [*she laughs*] Obviously, things change quite fast, so it's difficult to estimate what will happen next year. With the human motion project, I ask myself sometimes: If my research is only applicable to entertainment, is it useful enough? For example, I want to go a bit more into the medical applications of human motion. A personal issue I'm having is foot pain while walking. Many people have this, apparently, and it's a big unknown because the human body is very complex. You move one point in the body, and it destroys the whole posture and goes all the way back to the feet usually. I'm in contact with the Institute of Podology. They're training the doctors who make better shoes for people. I'm in contact with them to define their problems and see whether we can help because they have the data, and people in dynamic motion is my expertise.

Gül's PhD defense. From left to right: Iasonas Kokkinos, Marc Pollefeys, Ivan Laptev, Francis Bach, Andrew Zisserman, and Cordelia Schmid.

I want to enter that field, but at the same time, I feel like I don't know enough, so I need to read and learn.

How come something so problematic in the real world is not answered by science?

Well, that's what annoys me. I've had multiple MRIs, and everybody says nothing is wrong with it. They don't see anything abnormal, but the pain is there. It's something about more personalized medicine, maybe. Like knowing your own body.

How confident are you that you will be able to solve the human motion problems you have told us about today?

Oh, 100%.

That is perfect. How long will it take you?

That's the main question, right? [*Gül laughs*] At this rate, maybe five years.

Okay, so we will come back in five years.

Yes, let me just get another PhD student, and I'll be done!

What other skills will you need to employ to achieve it?

Just management. That's my main role.

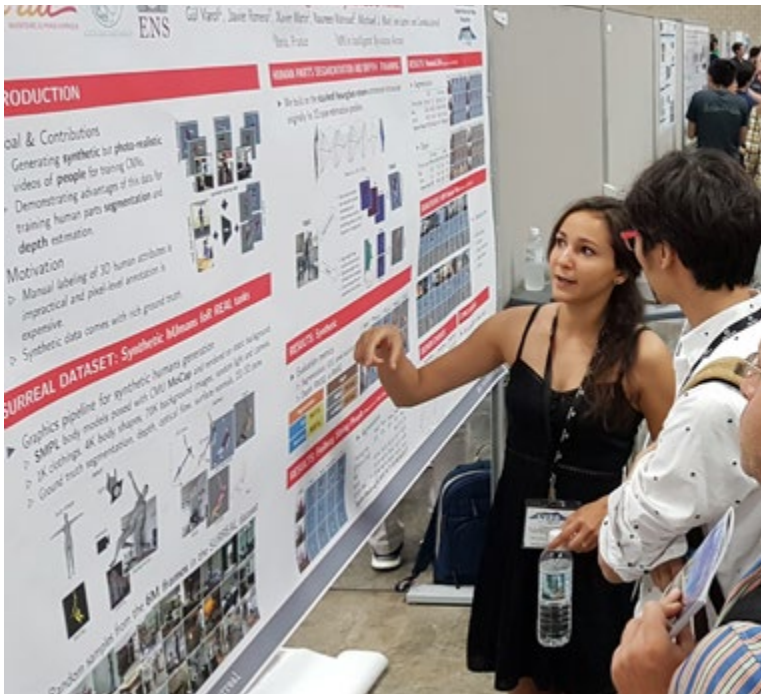
Time management or content management?

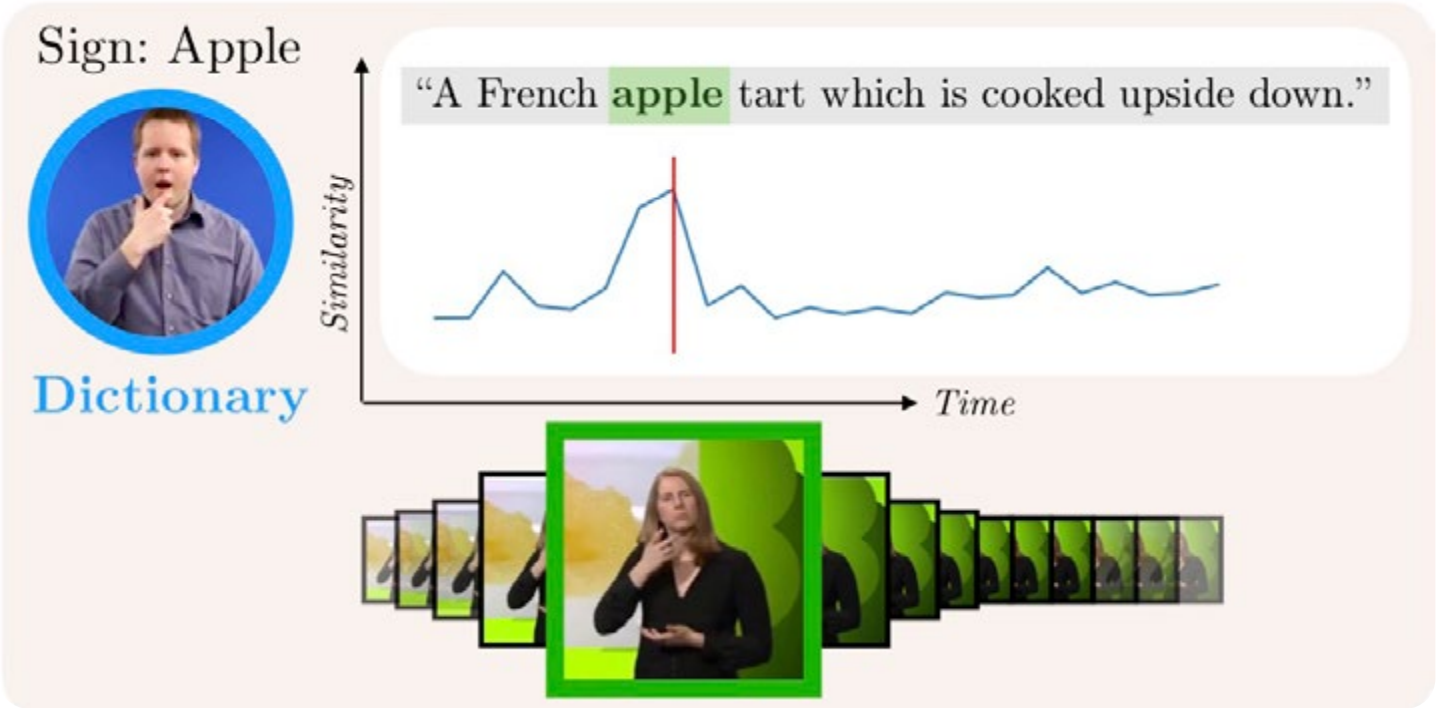
No, people management.

Once I asked Yoshua Bengio: Why are you so successful? He told me I am not smarter than the others, but I am more focused. What do you think of that?

Yes, I sometimes feel I'm not as useful as I could be in terms of contributing to the development of things, but in terms of having an insight into what might work, what might not work, being able to detect failures in advance, helping people focus, checking in regularly to make sure they're not lost, that's what I call the management part.

Can you pick one thing in the community that you would like to improve?





Work-life balance. It's a global community, so you can't determine the dynamics on your own because it depends on other countries and cultures. Here, it's relatively good in terms of work-life balance, but it has to be the same in other parts of the world for it to be stable because there's an imbalance if people publish three times more somewhere else or work at night and the weekend. Of course, people are free to do whatever they want, but when you're in the same community and the dynamics are different in other places, it creates pressure on the other side. I don't feel that stress anymore, but I understand that for PhD students, it's increasing, and I find it difficult to isolate them from that.

Finally, you must have learned much about yourself and the interaction between researcher and research in the last few years. Is there one precious thing you think our readers should know?

That's hard to pick, but probably the value of social interaction. I realized the job itself is quite social, from communicating the research and the papers to interacting and collaborating with people and meeting them at conferences. That alone is not enough, but it's probably one of the most important things.

Congrats, Doctor!

Fabio Pizzati has recently completed his PhD at the ASTRA team of Inria Paris. Under the supervision of Raoul de Charette, he developed a thesis on generative models, with many applications for autonomous driving in complex conditions, such as adverse weather and rare scenarios. Currently, Fabio is committed to pursuing an academic career and has recently accepted a postdoc position in Oxford with Phil Torr. Congrats, Doctor Fabio!

Have you ever wondered if you could do more with generative models?

The past few years have seen an exponential increase of papers related to **image generation**. However, it became clear that those systems require plenty of data to be trained efficiently, while not offering controllable outputs that may be necessary for practical applications.

In his PhD thesis, Fabio deals with these issues, by proposing improvements to the capabilities of **image translation networks**. First, he tackles training with **data scarcity**, focusing on low-shot systems that can be trained by just looking at a few examples. To achieve this goal, he proposes to bring back humans in the loop, by exploiting domain

Fig. 1

priors easy to define thanks to human abstraction capabilities. In particular, in his latest paper *(F. Pizzati et al., ManiFest, ECCV 2022)* he shows results of **few-shot or even one-shot generation of complex road scenarios** such as nighttime or adverse weather, by exploiting a semantic consistency mechanism learned on additional domains. But the applications are not limited to autonomous driving: have a look at Fig. 1, that showcases erupting volcanoes and auroras, generated by only using four images for training!

His thesis offers a second line of research, this time focused on **physics-informed learning**. Image-to-image translation networks, indeed, fail to be accurately controllable when applied to physics-based generation, including time of day modifications, or rendering of weather-related phenomena. With CoMoGAN *(F.Pizzati et al., CoMoGAN, CVPR 2021)*, the results of which are visible in Fig. 2, it is possible to process images acquired at daytime, and generate realistic timelapses in which the angle of the sun is explicitly controllable. The training is guided by naive physical models that roughly describe the appearance of daytime changes on images for different sun elevation values. In a previous ECCV paper *(F. Pizzati et al., Model-based Disentanglement, ECCV 2020)* it was also proposed to **combine realistic physical models and generative networks in a disentangled manner**. In fact, when generating images, it may be convenient to rely on the availability of realistic rendering for well-known traits, and generate the rest of the scene with **neural networks**. This makes the output images way more realistic and variable.

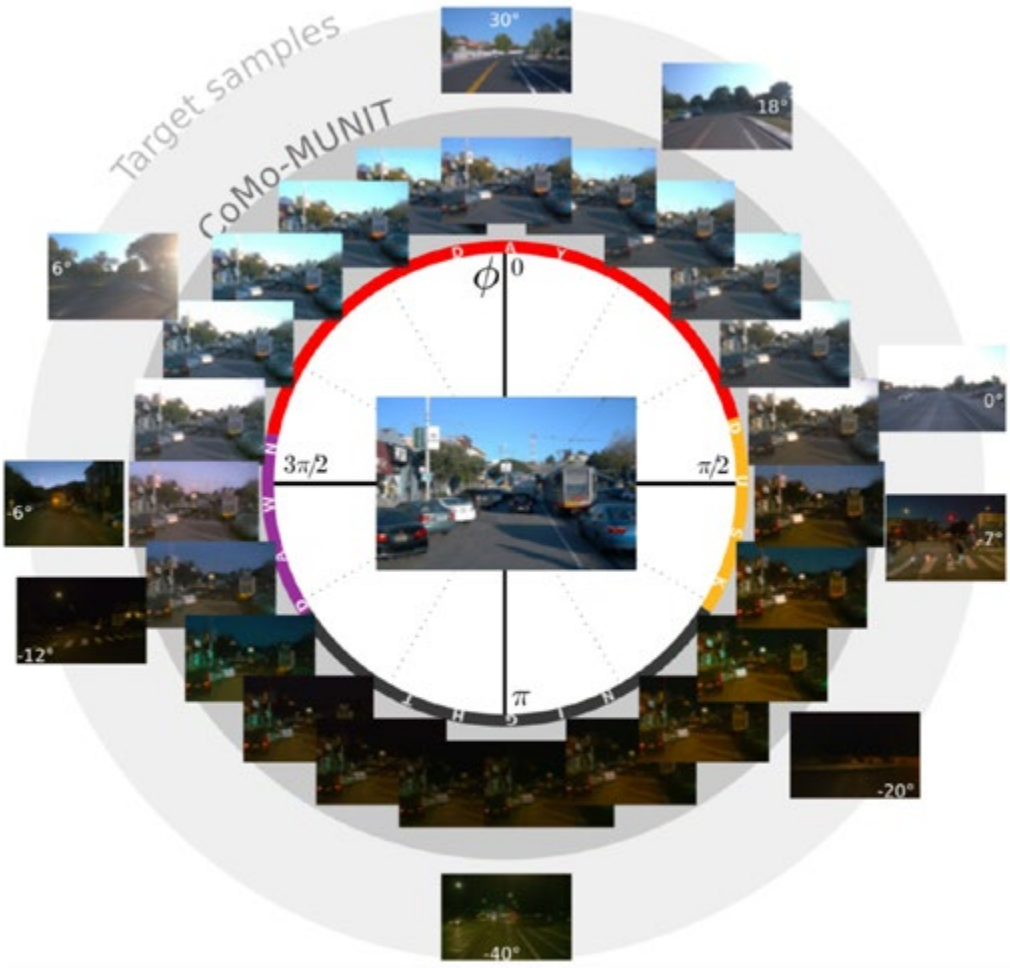


Fig. 2

In the future, Fabio plans to extend his research to **text-driven diffusion models**, while also exploring exciting new directions such as **continual learning and adversarial robustness**. Good luck!

Computer Vision News has found great new stories, written somewhere else by somebody else. We share them with you, adding a short comment. **Enjoy!**

Amazon's Just Walk Out and Amazon One - Beyond the Future of Shopping

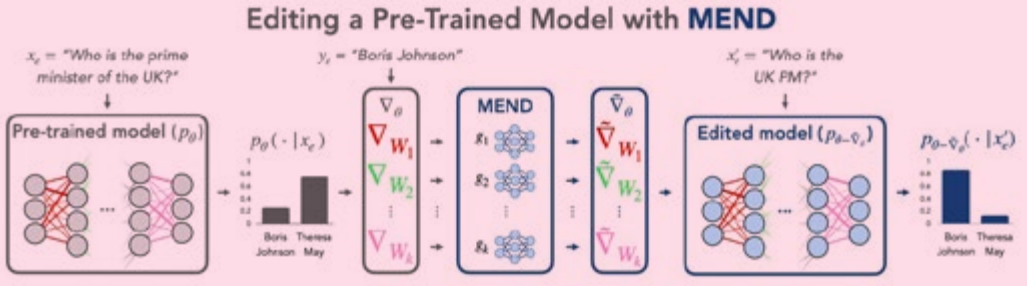
WIRED Brand Lab wrote a very nice article for **Amazon**, describing how their **Just Walk Out** technology and **Amazon One** are transforming the way we interact with the world, from retail to entertainment and business. Of course, this is scripted text, but it is interesting to see what Amazon sees in the future of shopping. This piece includes many quotes by **Gérard Medioni** who, besides being a big friend of our magazine and a pillar of our community, is also Vice President and Distinguished Scientist at Amazon. He has a lot to say about "**what machine learning can do to create magical experiences for consumers!**" Read More

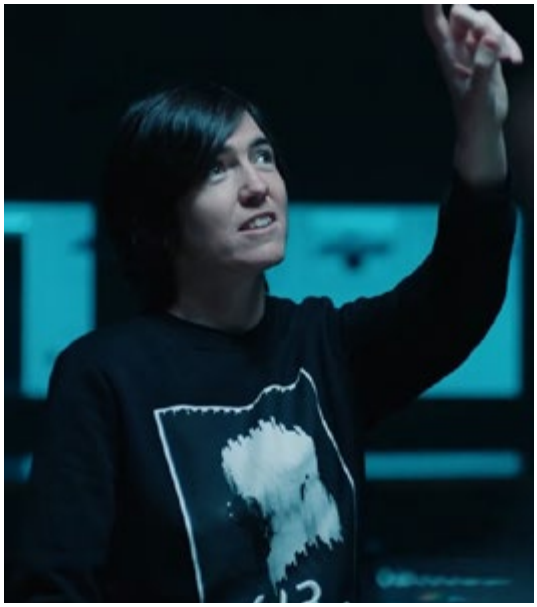
Researchers at Stanford Have Developed an AI Approach for Fast Model Editing at Scale

Large models have improved performance on a wide range of modern computer vision problems, in particular in **NLP (Natural Language Processing)**. However, issuing patches to adjust model behavior after deployment is a major challenge in deploying and maintaining such models! To enable easy post-hoc editing at scale, Eric Mitchell and his colleagues at **Stanford** propose **Model Editor Networks with Gradient Decomposition (MEND)**, a collection of small auxiliary editing networks that use a single desired input-output pair to make fast, local edits to a pre-trained model's behavior. Read More

Automatic Forest Fire Detection System with AI Enables Early and Efficient Fire Fighting

Here's a brilliant way by which AI helps contain one of the issues generated by climate change: **forest fires**. Deep learning algorithms teach an image processing system to see, recognize and verify **the occurrence of smoke**. Furthermore, AI enables a corresponding image processing system to draw conclusions from what it learns. It's a French company called **Paratronic** that has developed this automatic forest fire detection system called **ADELIE (Alert Detection Localization of Forest Fires)**: four industrial cameras observe a forest area within a radius of up to 20 kilometers and monitor it at 360 degrees in a couple of minutes. Read More





Raquel Urtasun's startup Waabi have unveiled their core product: Waabi Driver, an autonomous trucking solution built for safety, scalability and flexibility.

If you want to learn more about Waabi, read our latest interview with awesome Raquel:

We asked for more info about the Waabi Driver and here it is:

the **Waabi Driver** combines their revolutionary **AI-first autonomy stack as software with sensors and compute as hardware**. Together, they form a complete solution designed for factory-level OEM integration, commercialization, and safe deployment. What makes Waabi Driver so unique is that it is end-to-end trainable, interpretable, and has superior generalization capabilities, unlocking new autonomous lanes with unprecedented speed. Check the video:

The hardware is designed with production intent from day one and is purposely built for factory-level OEM integration. It is plug-and-play, lightweight, simple to maintain, and aerodynamic to maximize fuel savings.

The Waabi Driver is ushering a new era of trucking: the future of trucking is autonomous; the future of trucking is here!

Good luck Raquel and Waabi!

COMPUTER VISION EVENTS

NeurIPS
New Orleans, LA and virtual
28 Nov. - 9 Dec.

TechEx / AI & Big Data Expo Global
London, UK
1-2 December

CVMP Conference on Visual Media Production
London, UK
1-2 December

Computational Modeling, Simulation and Data Analysis
Zhuhai, China
2-4 December

IEEE Robotic Computing
Naples, Italy
5-7 December

SIGGRAPH Asia
Daegu, South Korea
6-9 December

The AI Summit New York
New York, NY
7-8 December

IEEE AIVR AI & Virtual Reality
Virtual
12-14 December

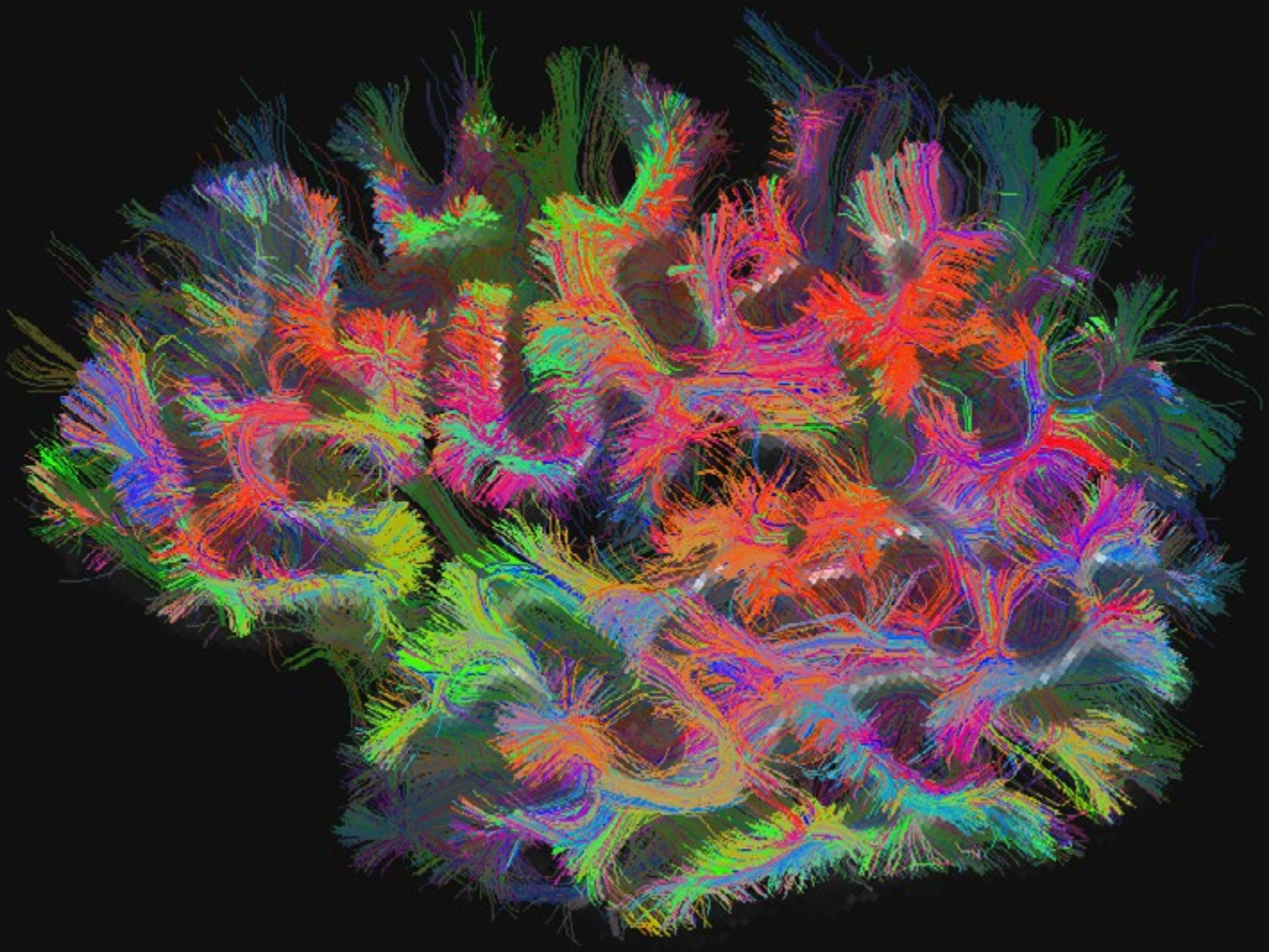
VCIP Visual Communications and Image Processing
Suzhou, China
13-16 December

Computational Science & Computational Intelligence
Las Vegas, NV
14-16 December

Due to the pandemic situation, most shows are considering going virtual or to be held at another date. Please check the latest information on their website before making any plans!

MEDICAL IMAGING NEWS

DECEMBER 2022



AI-BASED FEATURES TO ASSIST IN SINGLE-PORT ROBOTIC SURGERY

The use by the surgical community of single-port as a surgical approach has become increasingly frequent. In this context, **a surgical port is the pathway by which tools and cameras are inserted into the body via an incision**. That means that, instead of having multi-port and multi-arm as we are used to seeing in **Da Vinci and other medical robots**, some vendors are proposing a single-port robot, meaning that all the tools and cameras that are used during the surgery come from a single-port.

The advantages of single-port approach are multiple: from easier and shorter convalescence to less discomfort for the patients and less scars on their body. The tradeoff is between these benefits and the challenge of operating through one only angle: if the surgeons are trained to access the interested body region through different angles, they will need to adapt and work from one only pathway. This is why a single-port robot needs to make sure that its arms can twist and bend in all needed directions, so that the surgeon is not limited in the surgery by the single-port constraint. The goal is to let the surgeons conduct the intervention as if they were working from multiple ports, assuring them the degrees of freedom they need to operate safely and successfully. Of course, the incision place which will be used as a single port must be chosen with extreme attention.

Artificial Intelligence (AI) already assists multi-port robotic surgeries in many ways. It is often found that the same assistance can be given to single-port robotic surgeries as well, in addition to specific assistance it can give to solve the challenges of single-port. For instance, the field of view allowed by cameras in a single port is more limited than in the multi-port approach, where multiple access angles are provided. A unique port constrains cameras to be

smaller in size. The main solution that AI can give to these challenges is to **keep track of tools also when they are not found in the limited field of view available**: when a tool goes out of the frame, it is precious to know exactly where it is. AI keeps tracking the tools also when they are out of view, allowing full information to the surgical team about where are things in the space in which they operate.

AI can also **identify and automatically label each anatomical structure**, even those that are outside the field of view. This allows the surgeon a much larger and more precise view than allowed by the camera(s). AI algorithms can exploit a pre-operation map, register it to the real-time image and efficiently map the full surgical area.

This requires AI to perform **precise calibration and real-time tracking, segmentation, and labeling**. Tools, organs, and tissues move during the procedure and AI must give the surgeon precise information at any time, regardless of what has changed in the interested area.

Are you interested in adding AI features to single-port robots? Contact RSIP Vision and we will be happy to assist.

VISUALIZING DATA WITH MAYAVI

By Marica Muffoletto (twitter)

Welcome to the last tool review of the year! To top it off, we want to focus on a very “pretty” topic: a slightly different way **to display scientific data and meshes on Python**. Whether you just need to visualise your experiments for better understanding, or you need to show your results during a meeting, **this might be exactly the library you were looking for!**

Through the mayavi.mlab module (mlab), you can visualise datasets through a Python script or with an interactive prompt. Mayavi uses the TVTK module which is a Python wrapper of the VTK objects. This great flexibility allows the user to grab vtk modules and filters without having to deal with C++ or another software, to read several file formats, create good visualizations and use pipeline architectures like in VTK.

Installation

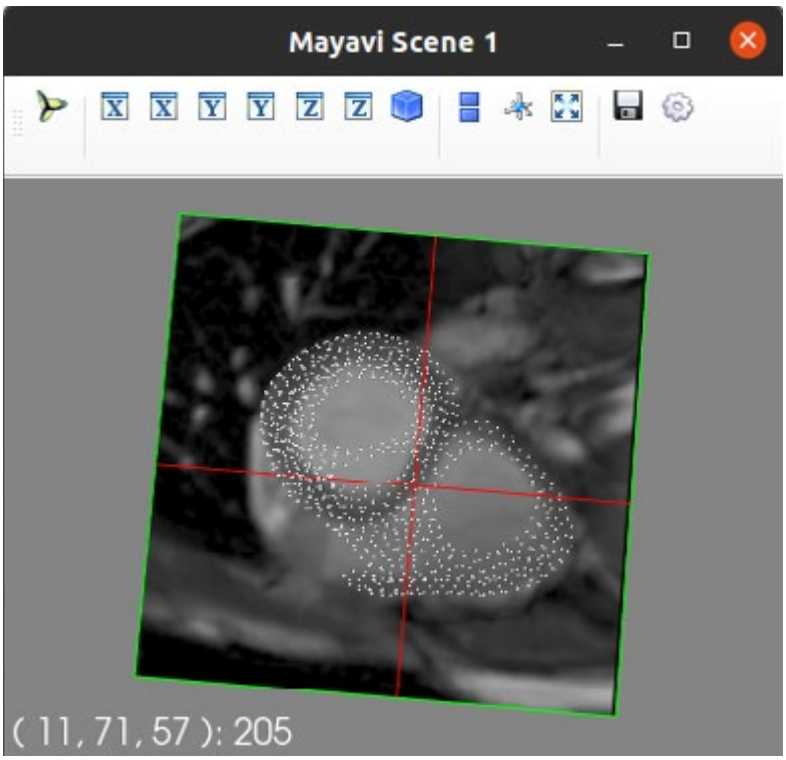
The installation of Mayavi might be a bit troublesome because it requires a compatible version of VTK. If you are lucky, installing Mayavi and a GUI toolkit like PyQt5 through pip will be enough. On my Linux machine, a successful combo environment contains Python 3.7, VTK 9, Mayavi 4.8, PyQt5 and Traits 6.2.

Image & Mesh

Before starting, we downloaded a dataset including a heart cine image and a mesh from the Multimodality STRAUS dataset.

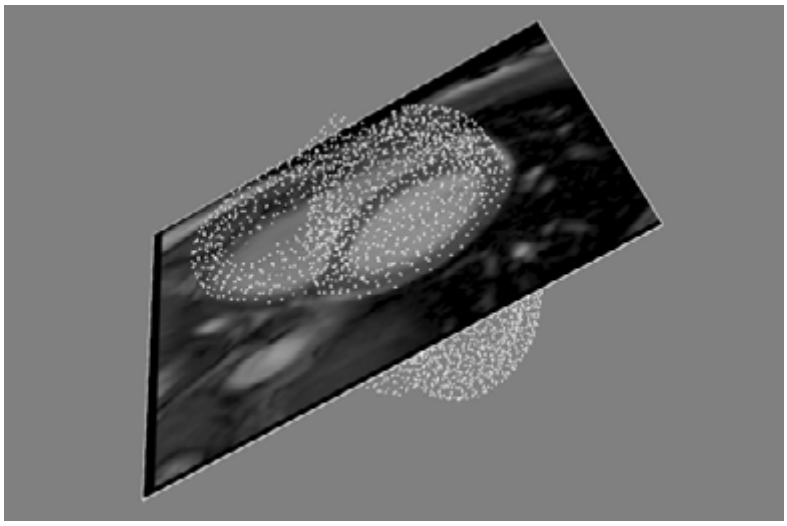
In this tutorial, we will read and visualise the image, and overlay a mesh of the ventricles as 3D points.

```
# Import libraries
from mayavi import mlab
import SimpleITK as sitk
import numpy as np
import vtk
from vtk.util import numpy_support

# Reads the image using SimpleITK, and returns origin and spacing to plot
the image in world coordinates
def load_itk(path):
    itkimage = sitk.ReadImage(path)
    scan = sitk.GetArrayFromImage(itkimage)
    origin = np.array(list(reversed(itkimage.GetOrigin())))
    spacing = np.array(list(reversed(itkimage.GetSpacing())))
    return scan, origin, spacing

# Define image and mesh paths
image_path = "HealthP/image/cinefrm02.mhd"
mesh_path = image_path.replace('image','mesh').replace('frm','mesh').
replace('.mhd','.vtk')

# Read scan
scan, origin, spacing = load_itk(image_path)

# Create Mayavi figure
mlab.figure()

# Add pipeline to visualise the image oriented on the x-axis
src = mlab.pipeline.scalar_field(scan)
src.origin = origin
src.spacing = spacing
plane = mlab.pipeline.image_plane_widget(src,plane_orientation='x_axes',
colormap='black-white')

# Read mesh
reader = vtk.vtkDataSetReader()
reader.SetFileName(mesh_path)
reader.ReadAllScalarsOn()
reader.Update()

header = reader.GetHeader()
mesh=reader.GetOutput()

# Read coordinates as x,y,z points
point_coordinates = reader.GetOutput().GetPoints().GetData()
numpy_coordinates = numpy_support.vtk_to_numpy(point_coordinates)
x = numpy_coordinates[:,0]
y = numpy_coordinates[:,1]
z = numpy_coordinates[:,2]

# Visualise 3D coordinates
mlab.points3d(z,y,x, scale_factor=1)_
mlab.show()
```



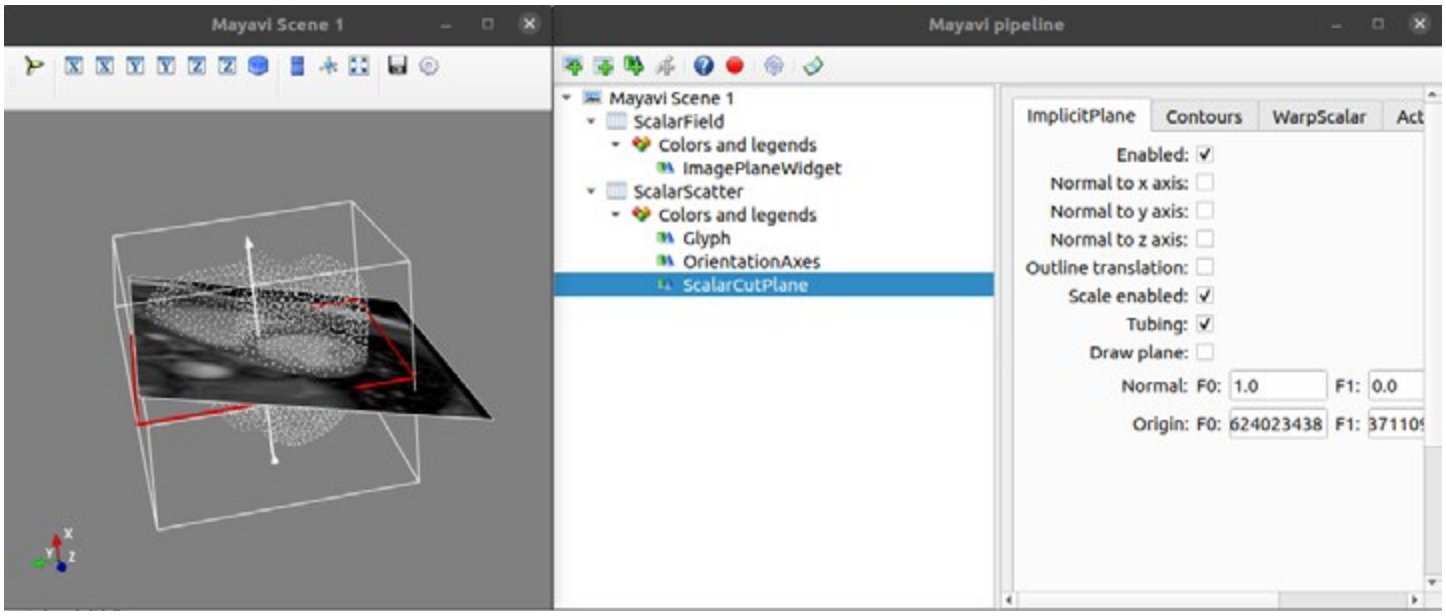
After running this code, Mayavi will open a window containing a Scene. Here, you can select a point and visualise its coordinates, you can change orientation and scroll through the slices. When you are happy with the view, you can press on the save icon.

Mayavi will prompt a menu to select the folder for saving the image in the desired orientation.



<--- Check out my result on the left

If you want to use the interactive prompt, it's even easier. You can click on the Mayavi icon on the left and when the pipeline window opens, add filters or modules. Below we just use this feature, to display the orientation axes and the cut plane.

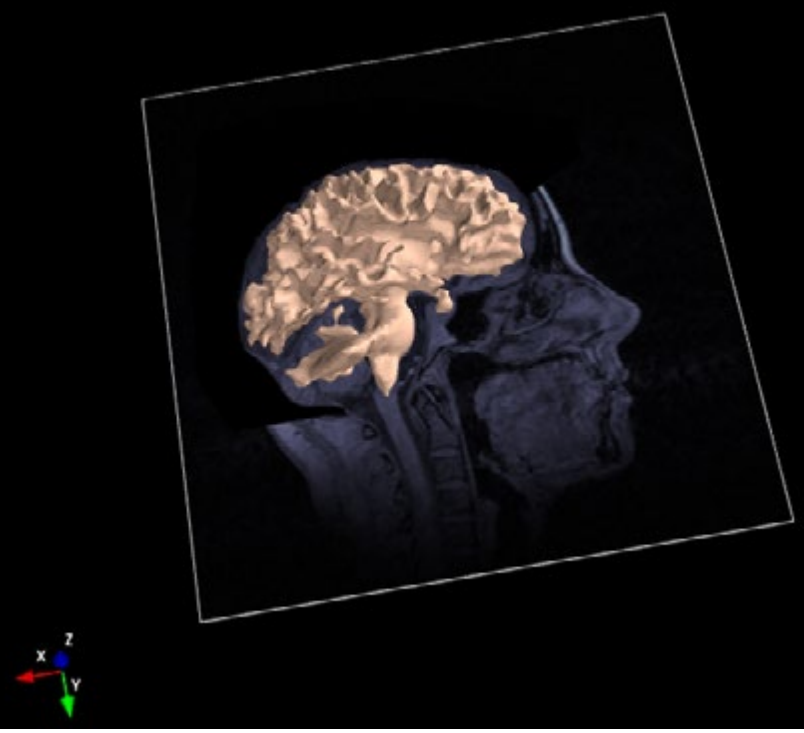


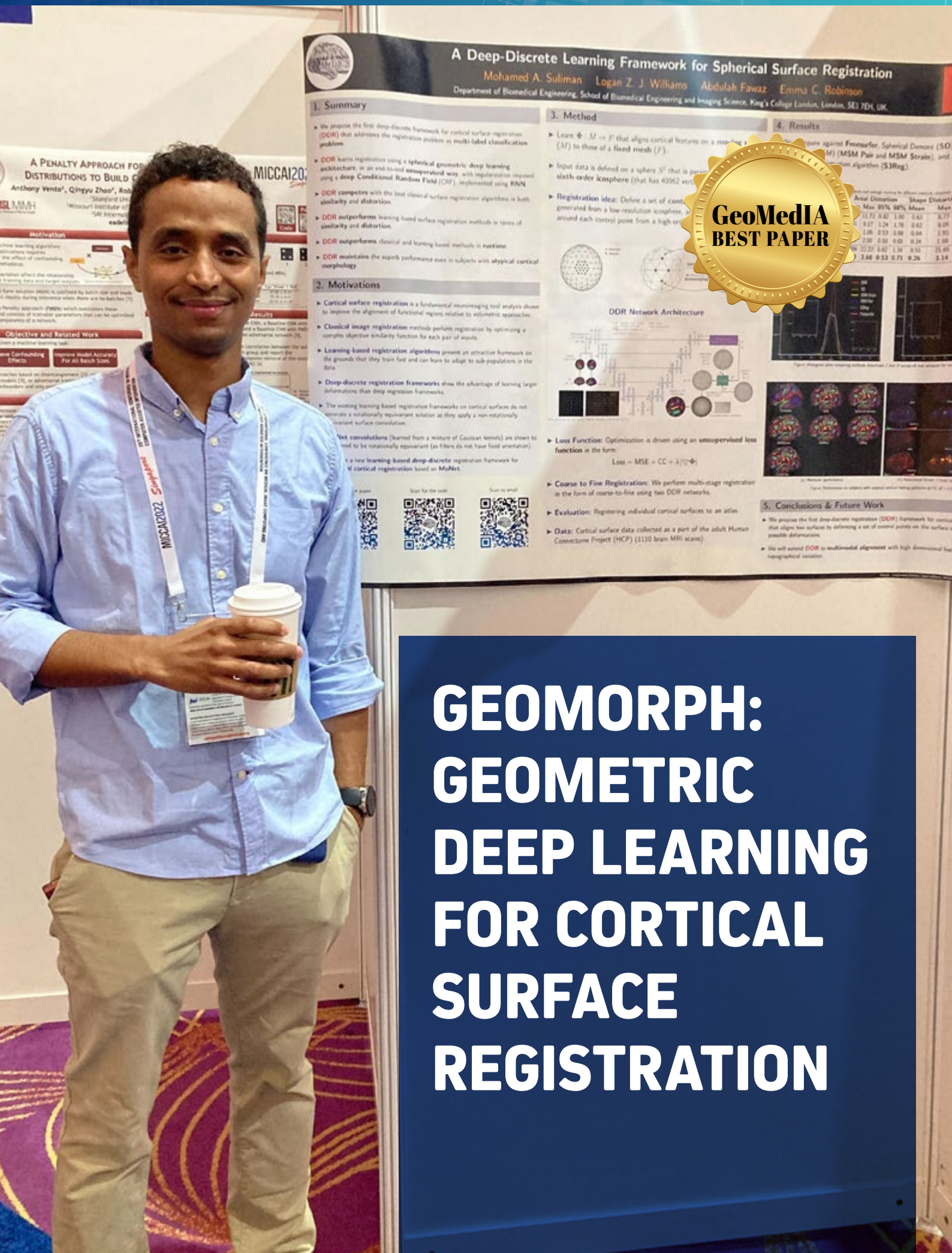
Extra

Since we previously mentioned the importance of the TVTK module, below we report an example of how to use filters to threshold and segment an image. This is an example reproduced from the Mayavi documentation on a brain scan. You can use this code, apply it on your image and personalise it with suitable values for the thresholding.

```
from tvtk.api import tvtk

# Apply image-based filters to clean up noise
thresh_filter = tvtk.ImageThreshold()
thresh_filter.threshold_between(lower_thr, upper_thr)
thresh = mlab.pipeline.user_defined(src, filter=thresh_filter)

median_filter = tvtk.ImageMedian3D()
median_filter.set_kernel_size(3, 3, 3)
median = mlab.pipeline.user_defined(thresh, filter=median_filter)

diffuse_filter = tvtk.ImageAnisotropicDiffusion3D(
                                    diffusion_factor=1.0,
                                    diffusion_threshold=100.0,
                                    number_of_iterations=5, )

diffuse = mlab.pipeline.user_defined(median, filter=diffuse_filter)

# Extract brain surface
contour = mlab.pipeline.contour(diffuse, )
contour.filter.contours = [brain_thr, ]
```

This was a very short overview of Mayavi, but we won't stop here and there is plenty of other visualization toolkits for Python to explore. Keep following us for more in 2023!



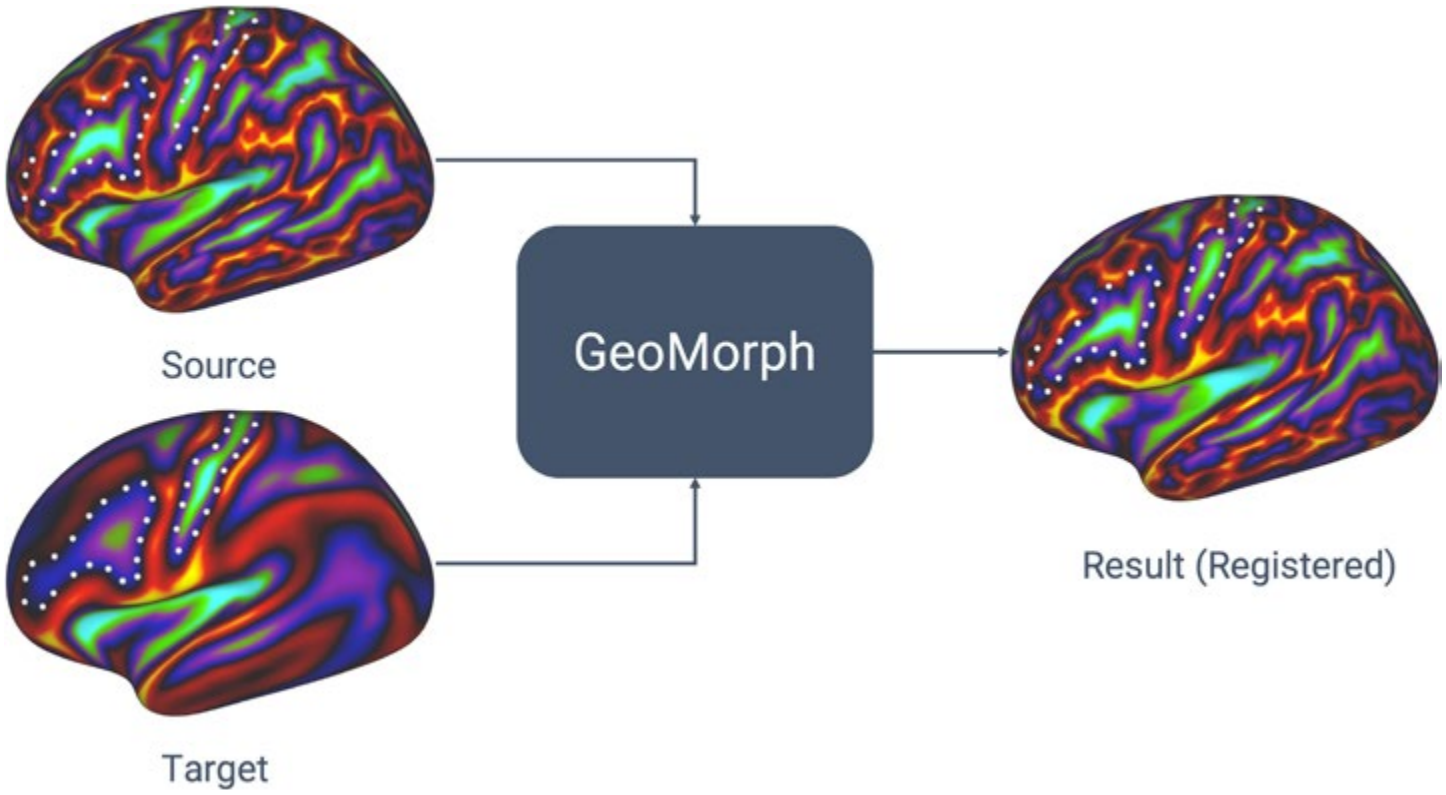


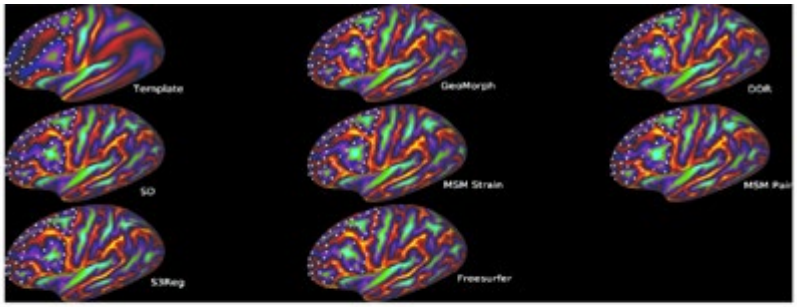
GEOMORPH: GEOMETRIC DEEP LEARNING FOR CORTICAL SURFACE REGISTRATION

Mohamed Abdalla E. Suliman is a research associate at KCL, working on developing artificial intelligence (AI) methods for analyzing medical imaging data. His focus now is on developing geometric deep-learning-based frameworks for cortical surface registration. Mohamed holds a Ph.D. from Imperial College London.

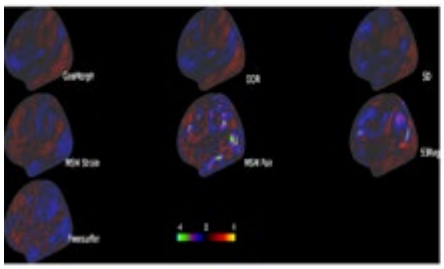
The human cerebral cortex is the brain's outer surface responsible for the cognitive functions that distinguish us from primates. The structural and functional organization of the cortex is highly variable across individuals, which makes comparing this structure across individuals a challenging task. One example might be investigating the brain regions responsible for understanding spoken language across a large number of subjects. Nonetheless, thanks to advances in cortical surface registration methods, **we can place individual cortical surfaces into a common coordinate system** that allows us to compare different features across subjects accurately.

Most commonly, **medical image registration aims to align cortical folds across subjects**. The issue with this approach is that folding is not a very good measure of brain function and therefore does not allow us to probe

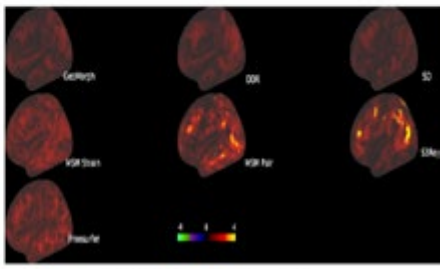




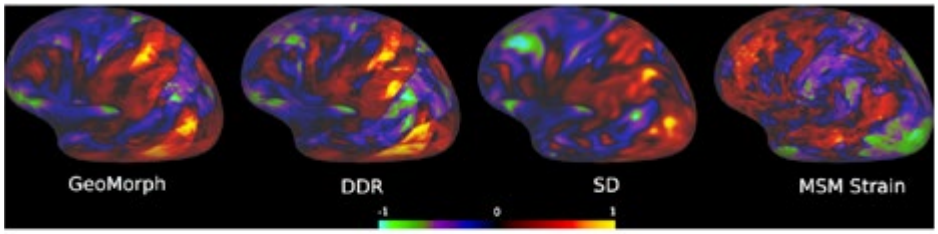
(a) Alignment Quality.



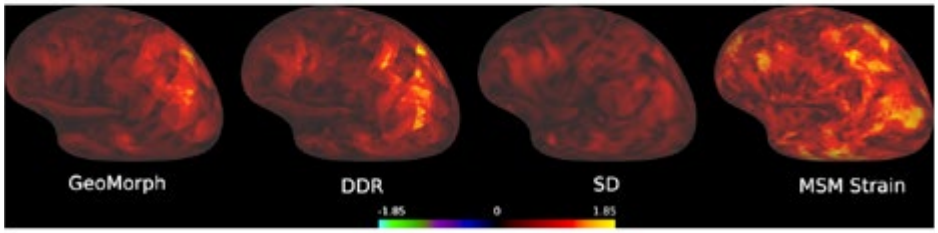
(b) Areal Distortion.



(c) Shape Distortion.



(a) Areal Distortion.



(b) Shape Distortion.

brain organization with the precision we would like. In recent years, classical cortical surface registration algorithms have been extended to allow **multiple cortical features** that better reflect brain organization to drive registration to a common space. However, even these approaches are limited in their ability to fully align all subjects to a common space. **Learning-based registration algorithms** emerged as alternative frameworks on the grounds that they are better at learning the space of variation to tackle topographical variation and are more efficient in learning population-specific templates.

In this work, we propose **GeoMorph**, the **first deep-discrete learning framework that performs registration on the cortical surface**. GeoMorph takes two cortical surfaces (individual subject and template) as inputs and aims to learn a smooth displacement field on the

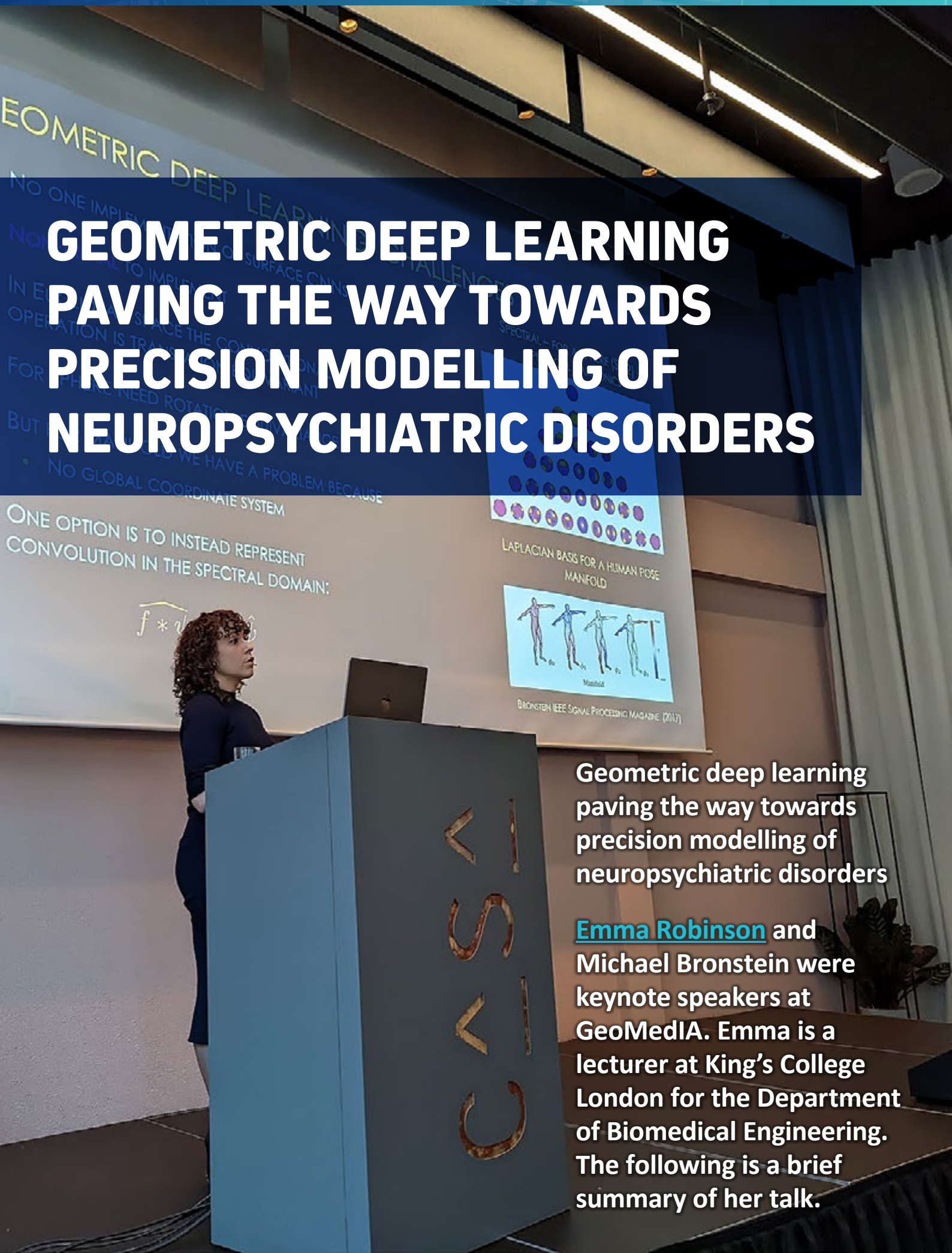
individual subject surface in a way that improves the similarity to the template. The displacements are learned in a discrete manner, taking inspiration from recent deep-discrete registration frameworks, which show capability in **learning larger deformations than continuous-based ones**.

To learn the deformation field on the cortical surface, we convert the registration problem to a **multi-label classification problem**, where each point in a low-resolution control grid on the cortex deforms to one of fixed, finite number of endpoints on the cortex. This deformation is learned using a **spherical geometric deep learning architecture**, in an end-to-end unsupervised way, with regularization imposed using a **deep Conditional Random Field (CRF)**.

Our results show that GeoMorph performs competitively in terms of alignment and distortion qualities relative to the most famous classical surface registration algorithms **while superbly outperforming them in running time**. What is more promising is that GeoMorph is capable of generating smoother deformations than all existing learning-based surface registration methods, even in subjects with atypical cortical morphology, i.e., **generating neurobiologically plausible distortions**.

Besides telling you of the Best Paper and the Keynote Speech at the GeoMedIA Workshop 2022, we want to show you some of the lovely images that we received from co-organizer Jelmer Wolterink (University of Twente). Enjoy!


GeoMedIA
AMSTERDAM



GEOMETRIC DEEP LEARNING PAVING THE WAY TOWARDS PRECISION MODELLING OF NEUROPSYCHIATRIC DISORDERS

Geometric deep learning paving the way towards precision modelling of neuropsychiatric disorders

Emma Robinson and Michael Bronstein were keynote speakers at GeoMedIA. Emma is a lecturer at King's College London for the Department of Biomedical Engineering. The following is a brief summary of her talk.

To my mind, **cortical imaging** represents a most fertile ground for the development of **geometric deep learning technologies**. For many years now, we have known that features of cortical function, microstructure and organisation are best studied on the surface. At the same time, it is becoming abundantly clear that traditional pipelines, based on image registration are insufficient to detect subtle features of pathology and cognition from individual scans. The degree of variability across individuals simply breaks the assumptions of classical registration approaches. **Geometric deep learning offers incredible opportunity for registration-free, diagnosis, prognosis, and generative/casual modelling from cortical imaging data**. However, deep learning on non-Euclidean domains represents unique challenges, which need to be addressed before this nascent technology can reach its full potential.

One feature that makes Euclidean CNNs so powerful is their equivariance to translations, meaning that as the view across an image translates the output of any convolution will translate in the same way. This allows networks to detect, or segment, objects wherever they are in a scene. The equivalent operation for surfaces would be equivariance to rotation but this is highly difficult to achieve since manifolds have no global coordinate system. This means that **it is not possible to transform filters over a surface in a transform invariant way**, since the orientation of a filter will depend on the path it has taken to that point.

For these reasons, early graph networks adopted an approach of **fitting convolutional filters in the generalized Fourier domain**, where it can be shown that the Fourier transform of the convolution of a filter and an image is equal to the product of their Fourier transforms. An equivalent operation can be defined for any domain: for spheres we use the Spherical Harmonics, for 3D rotations we use the Wigner D matrices, and for general graphs we use the eigenspectrum of the graph Laplacian.

The problem with these implementations is that: filters defined in this way are not spatially localised, and these operations tend to be highly computationally demanding. This limits their practical deployment in complex, deep networks. For these reasons a range of different flavours of approximation have been implemented from polynomial graph convolutions, to rotationally equivariant, localised, mixture of Gaussian filters (MoNet, Monti 2017). Cortical imaging also has its own variant: Spherical UNet (Zhao MICCAI 2019), which fits localised hexagonal filters to a sphere by leveraging the regular tessellation of an icosahedral grid.

Each of these methods trade off computational complexity, rotational equivariance and filter expressivity. Research from my lab (Fawaz A 2022) robustly benchmarked the relative importance of these properties for phenotype regression and cortical segmentation and determined that feature expressivity was key, but that registration independence cannot be obtained without some degree of transformation equivariance.

Since then we built from this understanding to develop a range of novel deep learning applications for cortical surface modelling from **propagating the HCP multimodal cortical segmentation** (Glasser Nature 2016) to **UK Biobank data** (Williams MLCN 2021), to **robust and generalisable frameworks for surface registration** (Suliman MA MICCAI & GeoMedIA 2022 - see previous pages), **translation of vision transforms to the surface** (Dahan MIDL 2022) and **novel generative models** that can simulate healthy ageing of individual preterm cortices and use deviation from true scans to predict cognitive outcomes (Fawaz A, MIUA 2022, GeoMedIA 2022). These show geometric deep learning offers enormous potential to tackle the biggest challenges in cortical imaging. At the same time, we continue look to ongoing development of group equivariant and efficient geometric networks to further improve the precision of these models and power of what they can represent.

A New Definition of Interpretability

Mara Graziani is a Postdoc Researcher at IBM Research Europe and the University of Applied Sciences Western Switzerland, where Vincent Andrearczyk is a Senior Researcher. They speak to us about their work with global stakeholders to converge on an overarching definition for the interpretability of AI systems.

In April 2021, a virtual roundtable public meeting was held on '**A Global Taxonomy for Interpretable AI,**' where a group of researchers from various backgrounds came together to work on **a global definition of interpretability**.

The meeting was endorsed by AI4media, a European Union Horizon 2020 project on which Mara and Vincent both work. Its goal is to build **a network of excellence for AI in the media sector and society**. The multidisciplinary consortium has a wide range of partners, including lawyers, sociologists, ethicists, and human rights defenders.

"*We realized we were all using a different terminology to address the goals of the project,*" Mara tells us. "*One of these goals is to develop **explainable and robust AI tools for the media sector**. We decided to call a bunch of experts and invite them to participate in a global event to express their opinion on the meaning of interpretability.*"

After the meeting, Mara and Vincent went away and performed an extensive literature review to find all the proposed taxonomies of interpretable AI. Even papers within the same technical domain could be contradictory.

"*Some researchers in the technical sciences like us define interpretability in one way, and others define it in another way,*" Mara continues. "*This isn't helping us get anywhere, especially when discussing interpretability with people from different backgrounds. For example, a lawyer writing the AI Act for the EU uses words and sentences that don't interface very well with what we're doing and vice versa.*"

They worked for months with philosophers, social scientists, ethicists, cognitive scientists, and colleagues from the technical domain, looking at many papers and definitions to find agreement. They submitted their work to the **Artificial Intelligence Review journal** and received helpful feedback, allowing them to formalize the definition further.

Computer Vision News is happy to circulate this newly agreed definition among the community:

An AI system is interpretable if it is possible to translate its working principles and outcomes in human-understandable language without affecting the validity of the system.

Having arrived at such a clear and sensible result, we have to ask: What took so long?

"***Our world is very structured*** *– we use definitions, we have a formulation of the problem,* ***we write down maths, and maths is universal!***" Mara explains. "*But we're talking to sociologists, ethicists, philosophers. In the paper, we're talking about religion and God and the interpretation of the Bible. How many years have people spent on that?*" she asks, laughing.

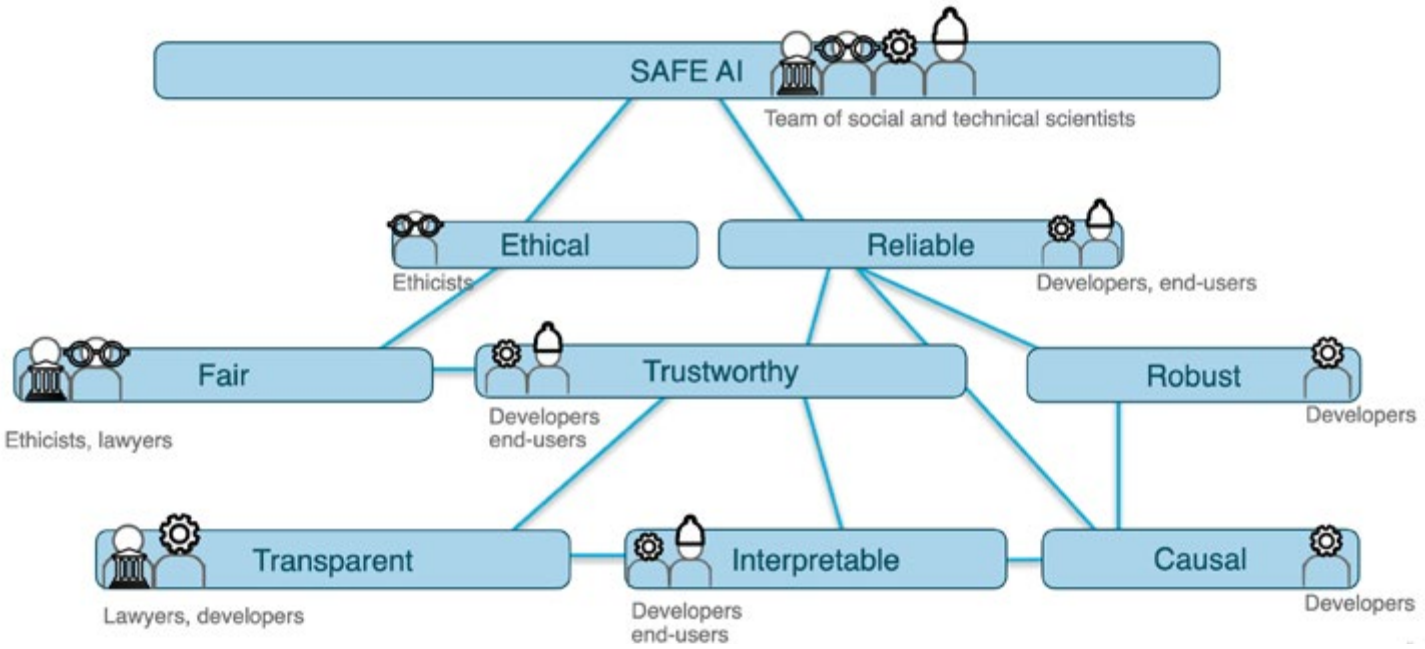
Vincent adds:

"Also, the review took a long time because there were many different definitions. We were trying to understand the viewpoint of each domain and why they employed these terms. Then we had a big group of experts from different domains, and ***communication was more difficult than it is when we are all tech people from the same background.***"

One of the critical areas for discussion was **whether the definition of interpretability had to be linked directly to human understanding**. Multi-agent AI systems, operating cooperatively, require interpretability in the exchange and decision planning at a level. It was a non-trivial question that took at least a month of back-and-forth debate.

With an agreed definition that Mara, Vincent, and others are all happy with, is the exercise now complete?

"Not at all – there are still many things to work on," Mara responds. *"For example, the definition of interpretability in law is still very fuzzy. They see it as privacy. We need to*



agree on a definition shared in the technical sciences as we design the systems that will deliver interpretability and accountability. But it's not like we've set the Rosetta Stone. It can be changed."

Vincent continues:

"We need a common basis to discuss the methods we need to develop to get feedback and to have a loop with all the experts. We may even need to give up on certain collaborations. If there is no understanding, there is no start of a collaboration. Then we want to start fostering new collaborations. We had a group of experts, and we tried to include as many experts as possible, but it can still be challenged and adapted in the future."

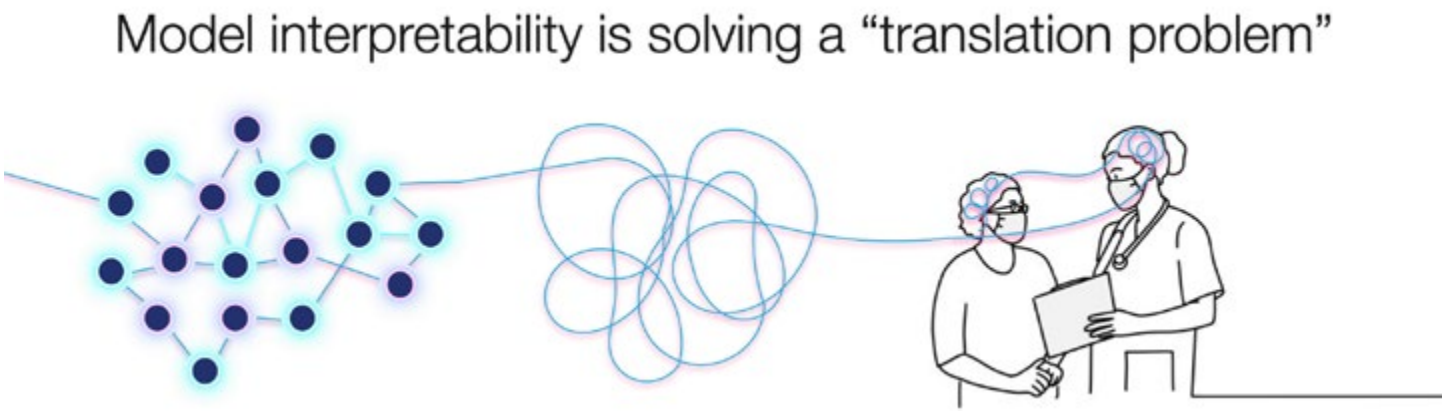
In the medical domain, also, there are barriers to overcome. Mara says that clinicians always ask: Are you developing something that will replace me in the long run? That is entirely not what they are doing, she counters. Instead, they are trying **to build systems that can interact with domain experts** because without that, systems can only communicate with tech people that know machine learning, and that's not the point.

In her day job, Mara is currently working on projects for AI4media and **IBM Research Europe**.

"At AI4media, I'm developing new tools for interpretability and looking into the unsupervised discovery of concepts in the latent space and the generation of causal explanations," she tells us. *"I'm also working with IBM Research Europe on a slightly different project related to colorectal cancer patients. The idea is to look into the images of their tissues, or histopathology inputs, and molecular profiles."*

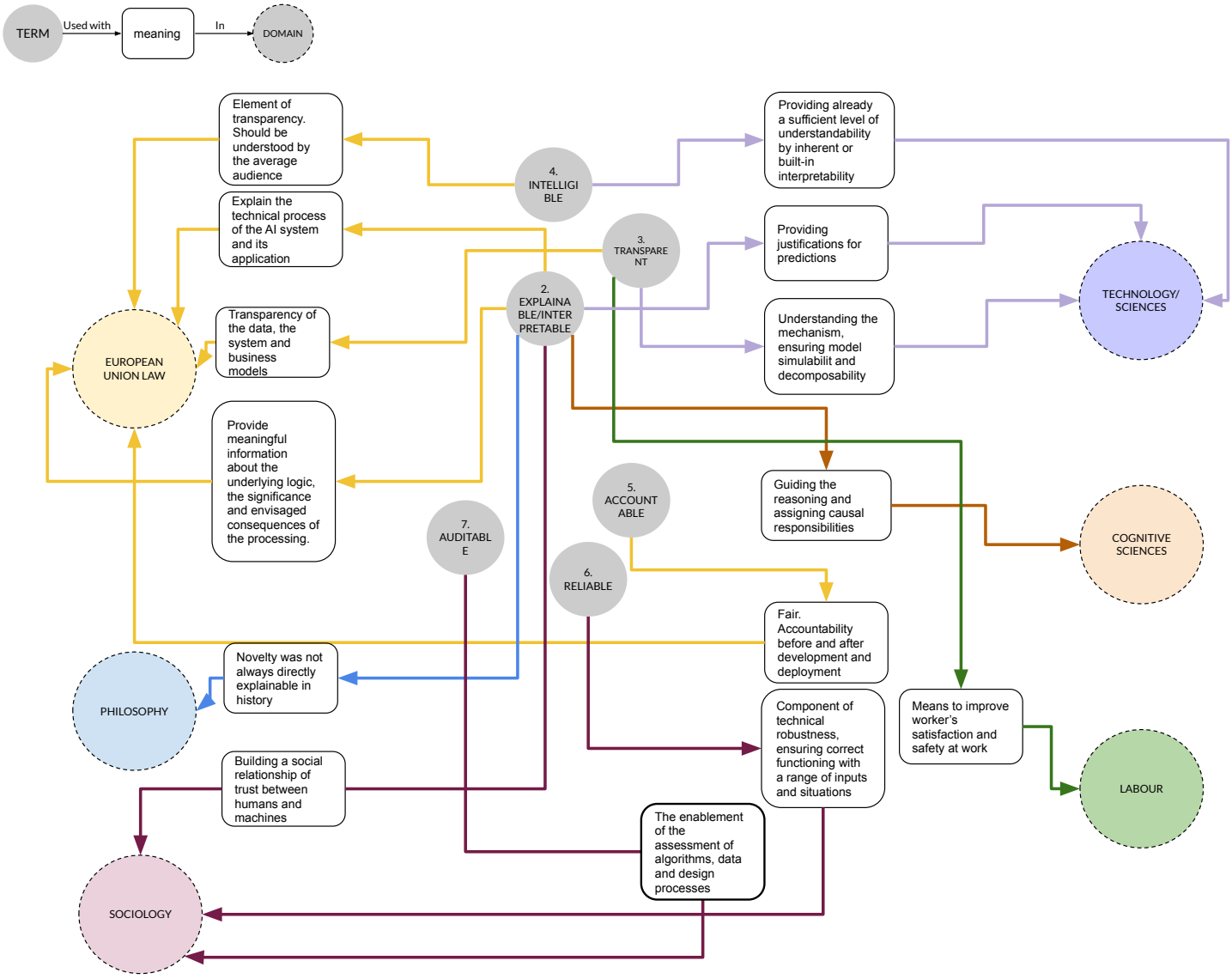
Vincent is currently working on various projects with deep learning for medical imaging.

"I'm starting a new project to analyze brain metastases," he reveals. *"We have*



Interpretable AI terminology

Main terms and domains



a unique database from the University of Lausanne with MRIs at different time points for evaluating the segmentation of brain metastases, the evolution of the metastases, the appearance of new metastases, and patient outcomes. I'm working on the technical side, and we have experts at the hospital helping us."

In terms of next steps, Mara and Vincent plan further development **to integrate the human in the loop** and say that running user tests is on the to-do list for everyone developing interpretable models.

"The point of this work was not only the definition but also restructuring the way we develop systems and adding social scientists, who study human understanding and how we process information, in the loop," Mara points out. *"I've already seen other papers published after ours do this, so I think it's a shared view, and it's one we should push forward!"*

Congrats, Doctor!

Anita Rau recently finished her PhD with the Surgical Robot Vision group at University College London. Her research aimed to improve navigation during endoscopic procedures by estimating 3D structures from monocular video. She now serves as a Postdoctoral Scholar at Stanford University, where she will continue her research on 3D scene understanding. Congrats, Doctor Anita!!!

Endoscopes have changed how we diagnose diseases. They allow us to examine the inside of the body without requiring large and painful incisions and the resulting recovery. But endoscopies require extensive training because navigation within the body is challenging for both humans and machines. Animal tissue has different properties than outdoor or urban environments. It is deformable, reflective, and lacks robust features. There is also no ambient light, which means the appearance of any given surface changes whenever the camera and its attached light source move. So, given all these challenges, how can computer-aided systems help clinicians or patients during endoscopy?

During my PhD, I helped develop a system that promises to one day provide clinicians with a real-time 3D map of the human colon, so they can better navigate within it. Such a system will allow its operators to plan their next movements or evaluate past performance.

Thanks to learning algorithms, researchers can easily predict local 3D structures in urban environments. But due to the extreme deformation within the body (bowel movement, breathing, changes in body pose etc.), it is impossible to obtain ground truth

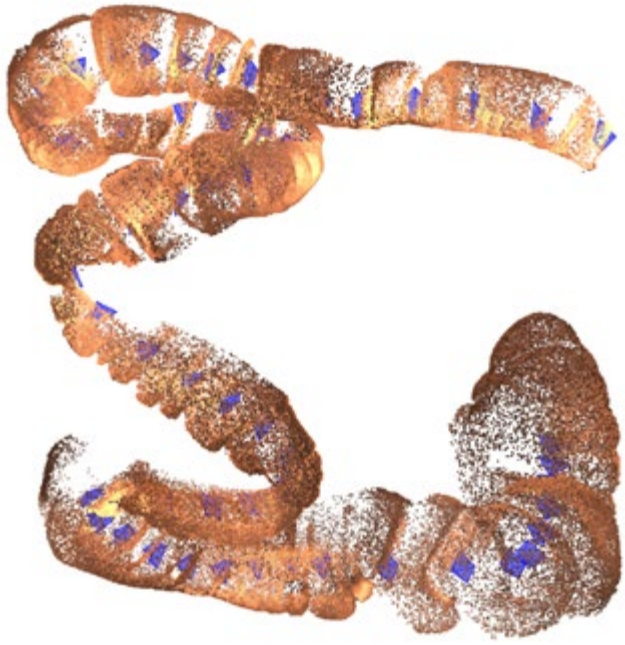


Figure 1: Our synthetic data provides 3D information and camera poses.

3D information from colonoscopic images. But ground truth is essential for learning. Therefore, one central question that I sought to answer was how to predict depth during medical procedures without the need for real ground truth.

I found synthetic data (Figure 1) to be a helpful alternative to real data because we could simply generate all necessary labels. I thus created one of the first public, synthetic datasets for colonoscopy **[1]**. While synthetic data provides ground truth and can be used to train networks, its appearance is distinguishable from real data. So, the focus of my research shifted from the question of how to predict depth to how we can integrate two different domains (synthetic and real) into a mutual framework. Previous works use two networks: one for domain adaptation and another for the task. But we found that combining both tasks into a mutual framework leads to more resilient depth maps (Figure 2) **[2]**.

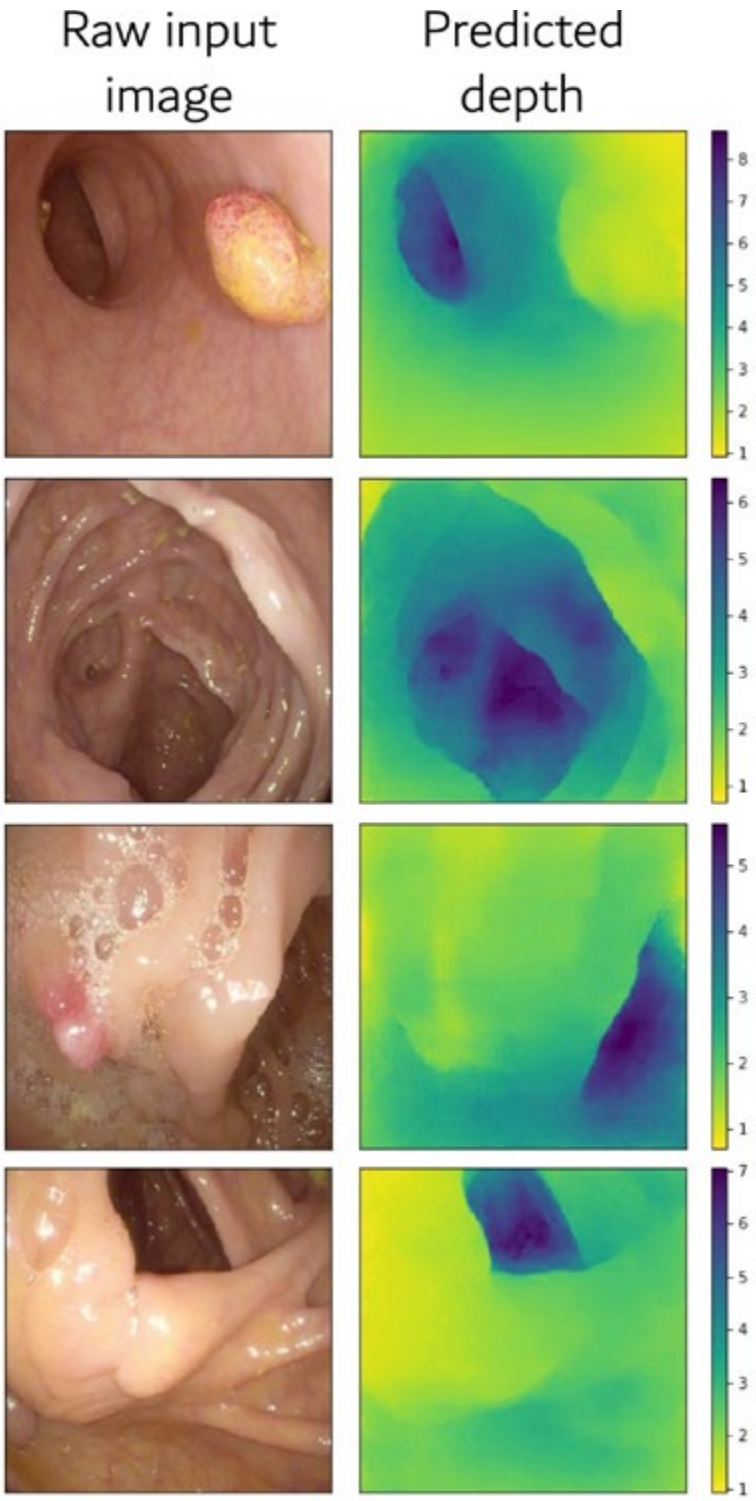


Figure 2: Trained with synthetic ground truth, our model can predict depth from real images.

Today, depth networks for colonoscopy can predict local 3D shapes fairly well but understanding the geometric relationship between two images remains challenging. My collaborators and I found that box embeddings—a concept derived from natural language processing—can be applied to images and help predict the relationship directly and in a human-interpretable way **[3]**. But integrating local structures into a global map of a colon has yet to be solved. Therefore, one of my final tasks during my PhD was to improve our public dataset **[4]** and help organize a challenge that will hopefully help other researchers tackle the remaining challenges in colonoscopic 3D reconstruction **[5]**!

PI-CAI (PROSTATE IMAGING: CANCER AI) GRAND CHALLENGE

Anindo Saha and Joeran Bosma are PhD candidates with the Diagnostic Image Analysis Group at Radboud University Medical Centre, working on prostate cancer detection in MRI under the supervision of Henkjan Huisman. Jasper Twilt is a PhD candidate at Radboudumc, with a background in technical medicine, under the supervision of Henkjan and MD-PhDs Jurgen Fütterer and Maarten de Rooij. Anindo, Joeran, and Jasper are co-organizers of the PI-CAI (Prostate Imaging: Cancer AI) Grand Challenge, and they speak to us as it enters its testing phase.

The number of **prostate MRI** acquisitions to be collected is expected to rise by quite a large margin in the coming years, meaning there will be more and more cases to be evaluated. Expert radiologists perform this evaluation, but there is a global shortage of these experts. **Prostate cancer detection in MRI with AI** aims to alleviate this workflow and make these radiologists quicker. Then if there are sufficient radiologists, its goal is to make it **more accurate**.

The **PI-CAI Challenge** targets clinically significant prostate cancer detection and diagnosis in MRI. One other challenge in this space, in 2016, was the PROSTATEx Challenge, also from Radboudumc. The PI-CAI team wanted to **scale up with more data, updated evaluation strategies, and a multi-center, multi-vendor dataset and cohort**.

"This time around, we wanted to partner with different centers and experts worldwide to get a unified perspective on how we should tackle this challenge," Anindo tells us. *"The study design was reviewed by 16 multidisciplinary experts from the fields of **radiology, urology, and AI related to prostate**. We wanted to make a bigger impact with the Grand Challenge and take it that one step further."*

Annually, one million men are diagnosed, and 300,000 die from clinically significant prostate cancer. When someone has a prostate MRI, it first identifies if there is an aggressive lesion. The AI models are expected to give a case-level diagnosis, with each patient receiving a percentage likelihood of cancer being present.

"There are many research topics within this field that specifically investigate the added

value of AI," Jasper tells us. "*In a concurrent reading setup, we've seen that AI improves the workflow period, so how long it takes to do your reporting. It can also assist in annotating suspicious areas and providing suspicion scores. We've also seen in these studies that agreement among radiologists improves, with less experienced readers getting better and finding they are more likely to agree with the experts.*"

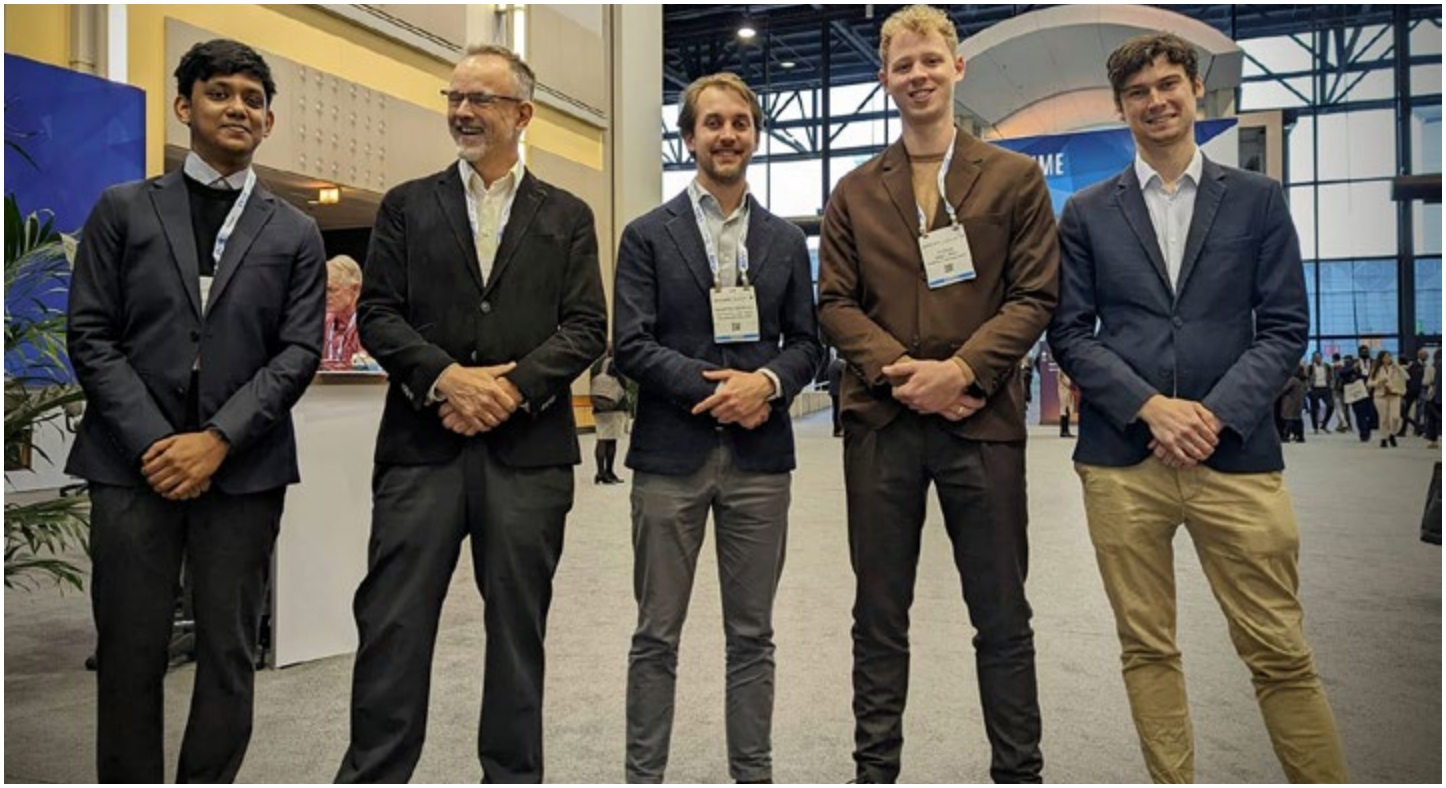
Does the team expect the winning model to perform better than doctors? Anindo reveals that at the start, he was skeptical, based on prior studies that they have compared against expert radiologists, but his view has shifted.

"*We have an estimate of how many cases are needed to train such a model and how well experts rank,*" he says. "*We're doing a Reader Study in conjunction with PI-CAI, and it's not just a handful of experts. We have* ***79 radiologists enlisted worldwide****, and to the best of our abilities, they include the entire spectrum of radiologists. Now, we have started to see that* ***AI can make a big difference****.*"

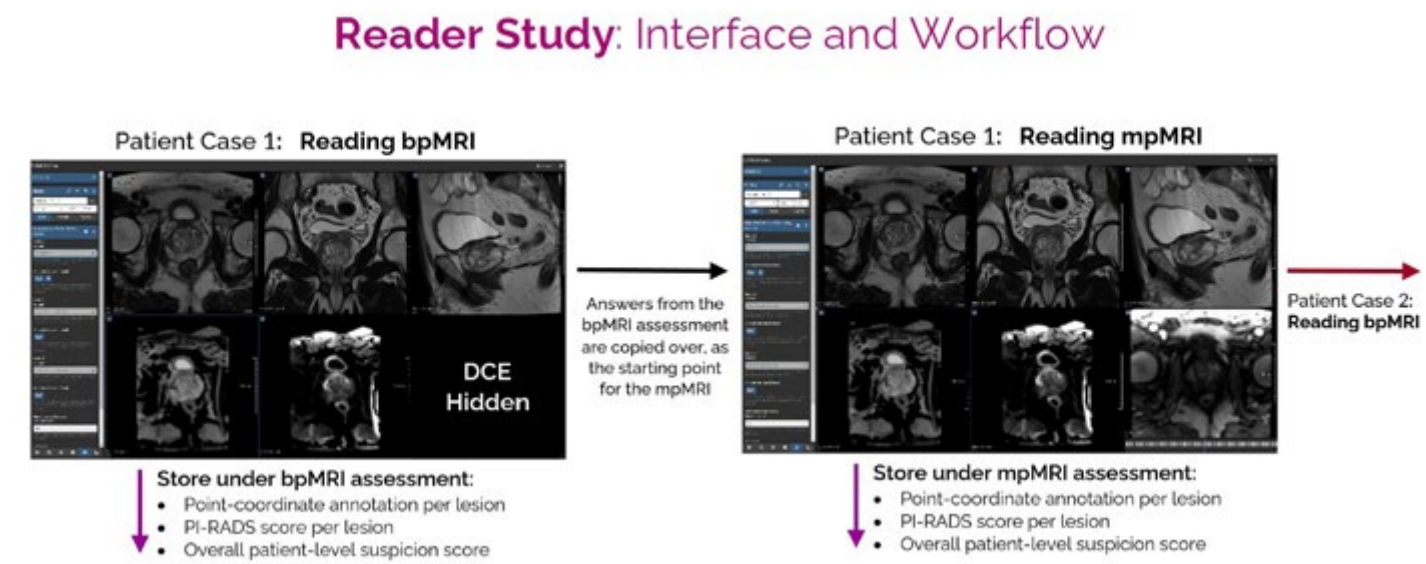
Joeran adds: "*For me, the goal is to have* ***an AI detection system that performs at least as well as expert-level radiologists!***"

The team will not give too much away but tell us that preliminary results are promising. Another goal is to unify how this task is performed so that everyone agrees on what they want from AI, how to evaluate it, what the datasets are, and how clinical and technical experts work together. The team also hopes it inspires others in the community to contribute to PI-CAI in the future, including proposing new datasets and tasks.

Although it is too late to give anyone advice for submissions this year as the challenge is already in the testing phase, what do they hope participants have taken into account?



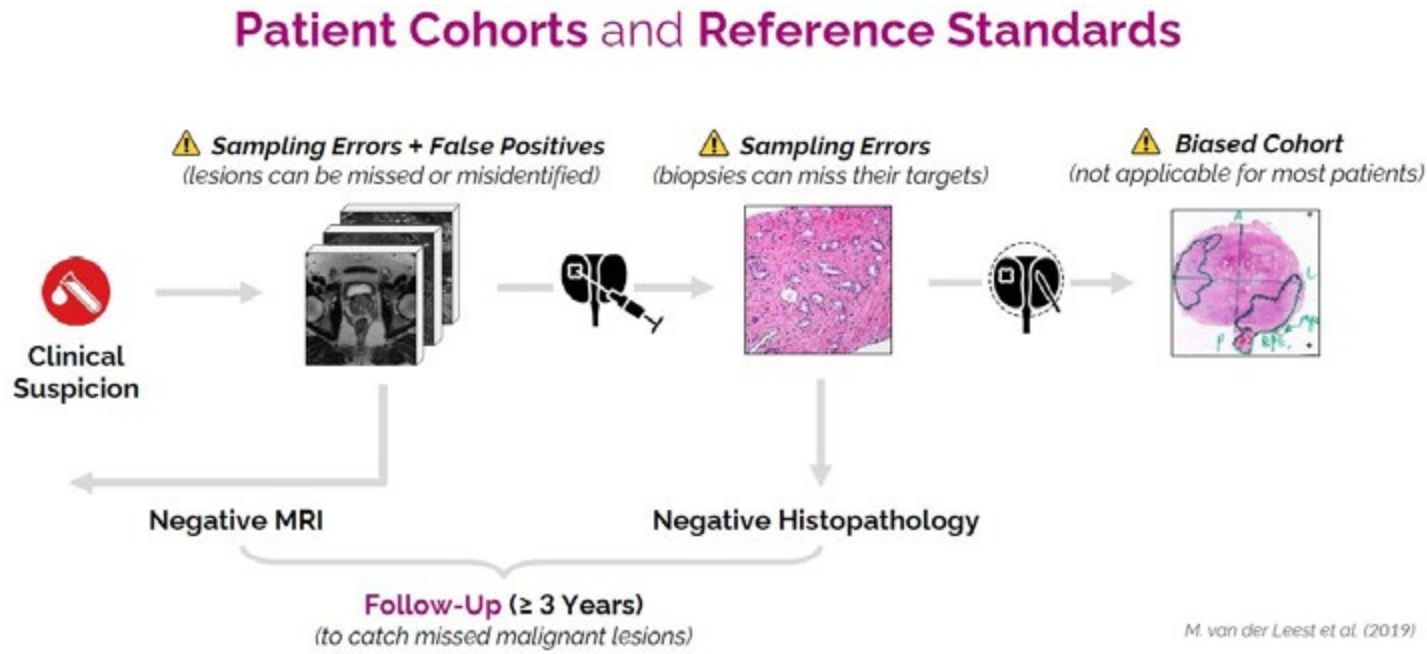
From left to right: co-organizers Anindo Saha, Henkjan Huisman, Maarten de Rooij, Jasper Twilt, and Joeran Bosma this week at RSNA.



"One big thing that would benefit any team is to really understand the clinical problem and the sort of data they're dealing with," Anindo points out. *"What is prostate MRI good or bad at? What are things in the prostate* ***that look very similar to aggressive cancer****? What are things that are not similar at all but could also be cancer? What are the clinical variables? We've included, for example, patient age and a PSA value in blood, which are connected with cancer screening usually. How do these play in, and how can they use this information to* ***reduce false positives or negatives?****"*

PI-CAI not only focuses on developing and estimating the best AI performance possible today but does it in conjunction with a **Reader Study** to know the best performance of radiologists overall in the field because one doesn't exist without the other. With his clinical background, Jasper is more involved in this aspect of the work. The team aims to benchmark the state-of-the-art AI algorithms developed in the challenge against prostate radiologists at the end of the study.

The challenge also goes one step further on the technical side. Very often nowadays,



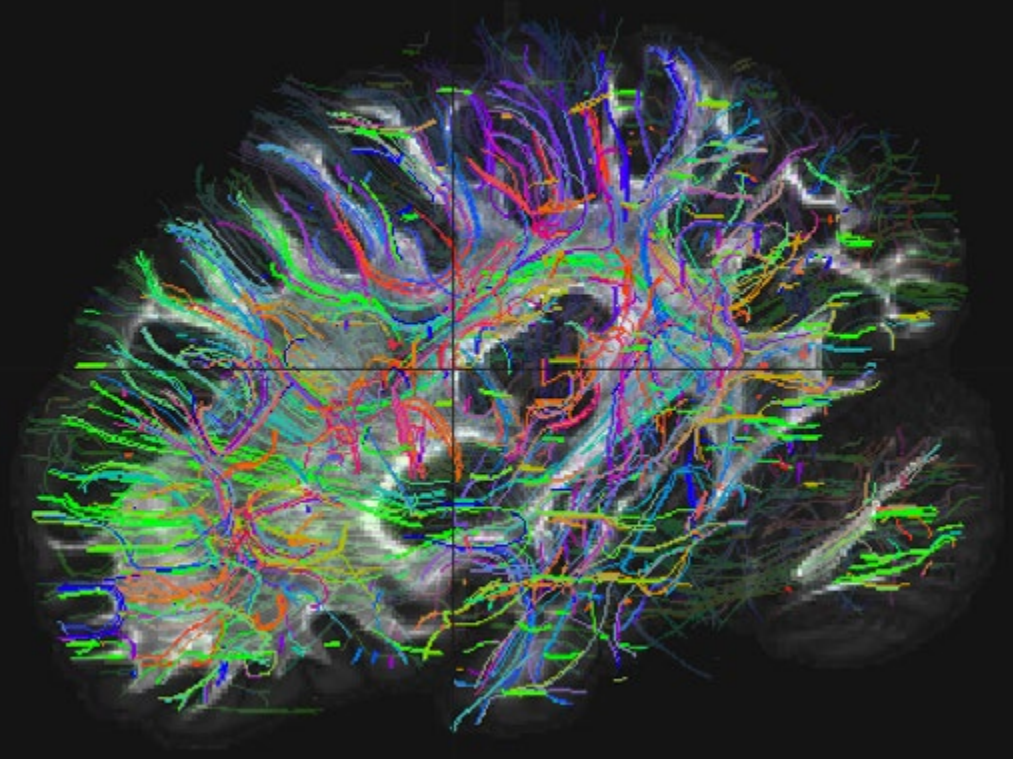
a challenge has a hidden test set on which you submit your algorithm, but for the final phase of PI-CAI, the top teams will be invited to provide their **training algorithms** - not the trained algorithm, the one that's finished, but **the scripts that made it based on the input dataset**.

"*We'll train their algorithms on an extended dataset of 9,000 cases, of which not everything was publicly available,*" Joeran explains. "*This will show us how well their algorithms scale up to a new dataset. We'll be able to leverage all data that is feasible. Also, because we're collaborating with Amazon, we'll train each of the top five teams' algorithms 10 times, so 50 runs in total, to estimate the variance between the training runs, and we'll do a proper statistical analysis on that.* ***Participants are evaluated on how good their training algorithms are rather than on something that's come out of their pipeline****. Then these training algorithms can be used in future work as well.*"

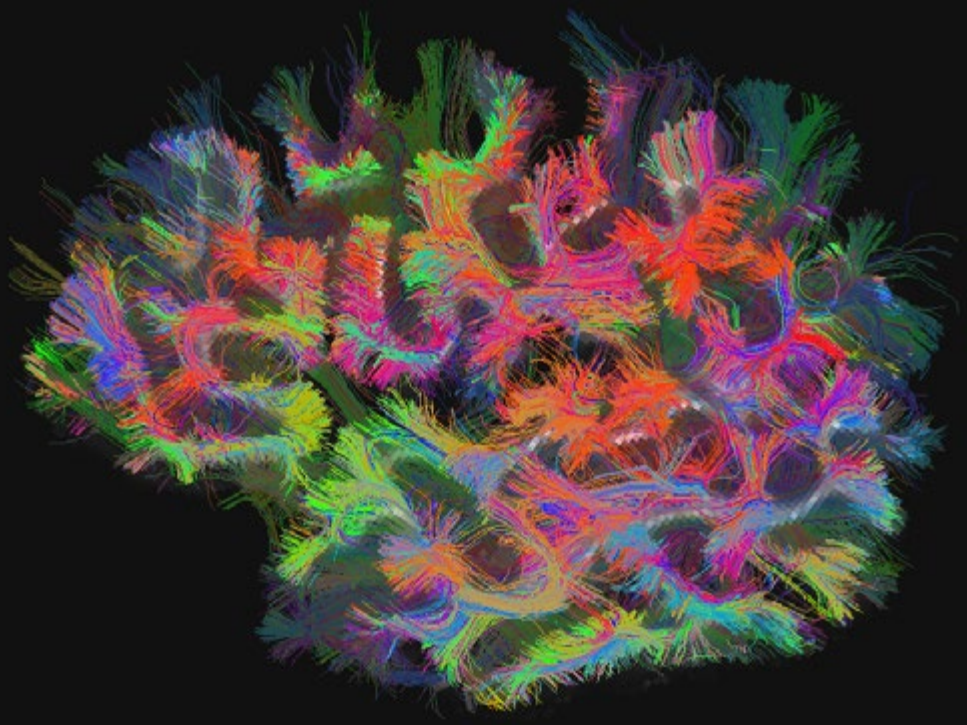
As we wrap up, Anindo offers us a wonderful insight into the process of preparing for this challenge, which harks back to the old saying: if at first you don't succeed, try, try again.

"*Before I started as a PhD here, I was a master's student, and there was another PhD working on this topic who was using part of the dataset we use now,*" he recalls. "*He found a lot of issues with that dataset that we fixed. Then whilst working, I found many more issues that were never spotted, and we fixed those. I was thinking, okay, fine, we now have perfect data. Then Joeran started his PhD, and he found even more issues, and I thought, oh my, this never ends! Then we fixed those. Great. Once we publicly released the dataset, the community found even more issues we had to fix! But that's just how it is. Our expertise individually is limited, but when you release it to the public, you see more things, and you can collectively make a nice, strong, clean dataset****. The work with data never ends. It's just forever data, data, data!***"

This is possibly the first ever tractogram in the universe!



Created entirely in **Julia language**! A tractogram is a set of streamlines representing the network of fiber bundles that connect different brain regions.

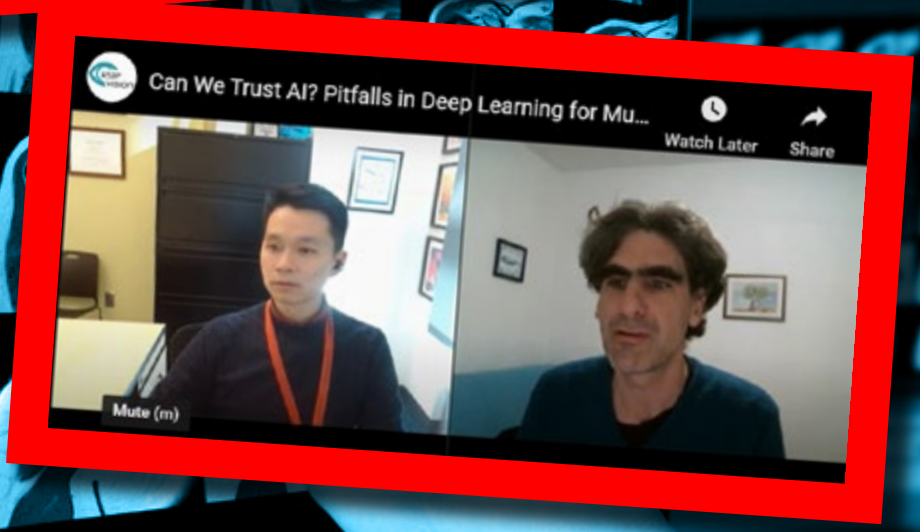


And the following one is probably the nicest ever!

Tractogram from an ex vivo human hemisphere that was scanned with 0.75 mm resolution diffusion MRI and analyzed with FreeSurfer.jl, a new Julia package developed at the Martinos Center for Biomedical Imaging by **Anastasia Yendiki**, Associate Professor of Radiology at **Harvard Medical School**.

RSIP VISION WEBINAR

If you missed the webinar, don't miss the video!



Can We Trust AI?

Pitfalls in Deep Learning for Musculoskeletal Imaging and Opportunities for Improvement

HOSTED BY

Moshe Safran

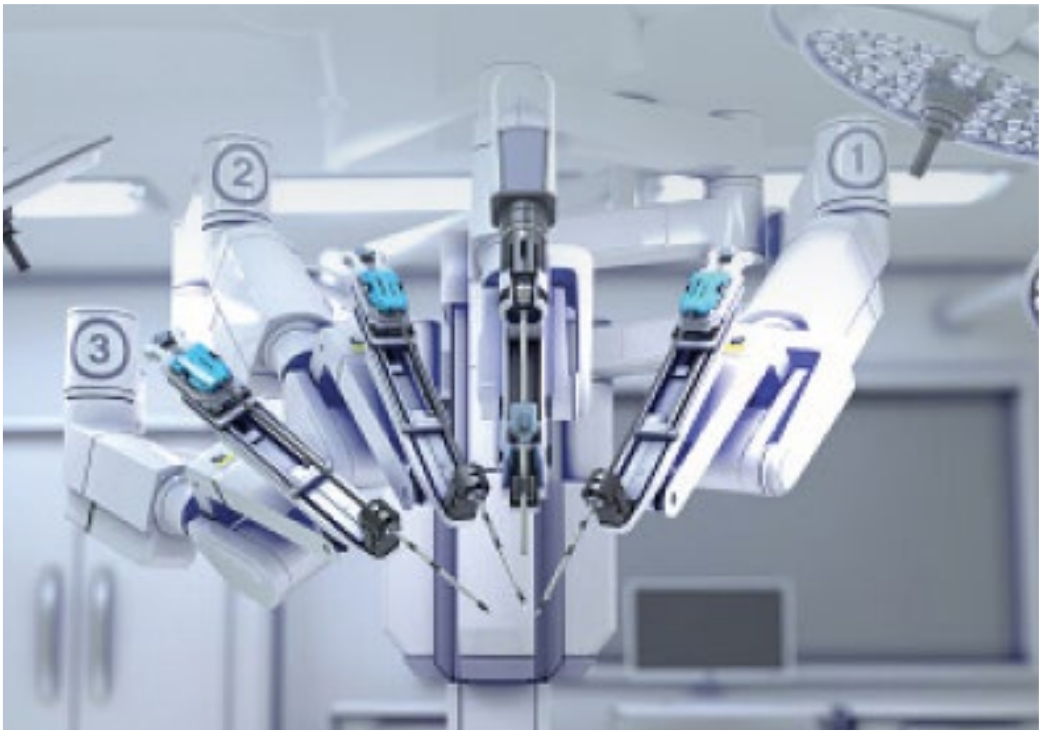
CEO of RSIP Vision USA.

Defining & developing innovative Medical Visual Intelligence solutions, in partnership with MedTech industry leaders.

GUEST SPEAKER

Paul Yi, MD

Director, University of Maryland Medical Intelligent Imaging (UM2II) Center. Assistant Professor, Diagnostic Radiology and Nuclear Medicine.



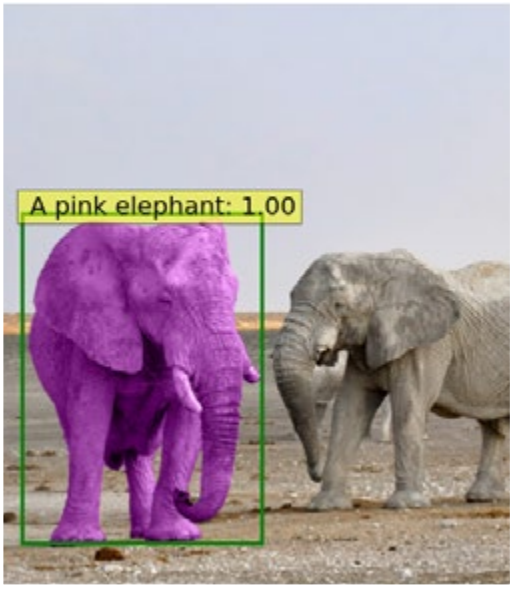
IMPROVE YOUR VISION WITH

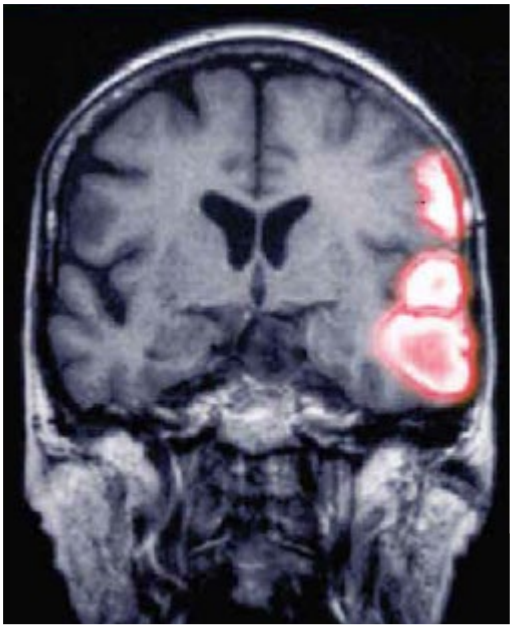
Computer Vision News

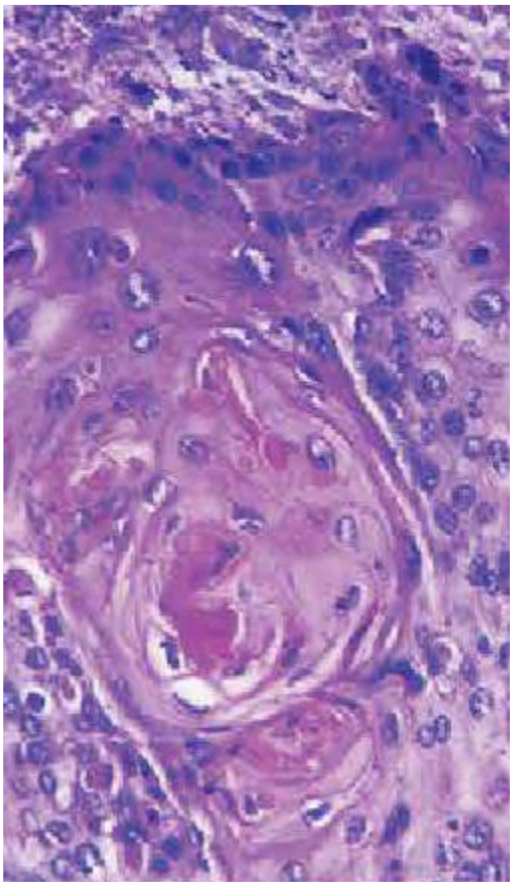
SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!







RSIP VISION