NOVEMBER 2022

Computer Vision News
& Medical Imaging News

The Magazine of the Algorithm Community

RSIP VISION
Visual Intelligence for MedTech

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

Dear reader,

Last week saw **ECCV 2022** take Tel Aviv by storm, with more than **5,000 participants**, mostly in person, coming together to discover the latest and most innovative new research that the computer vision community has to offer. It was, of course, a roaring success and a testament to the organizers' hard work. **General co-Chair Lihi Zelnik-Manor** told us: "*It was very important to me to give our guests an excellent experience, to learn how Israelis are great hosts.*"

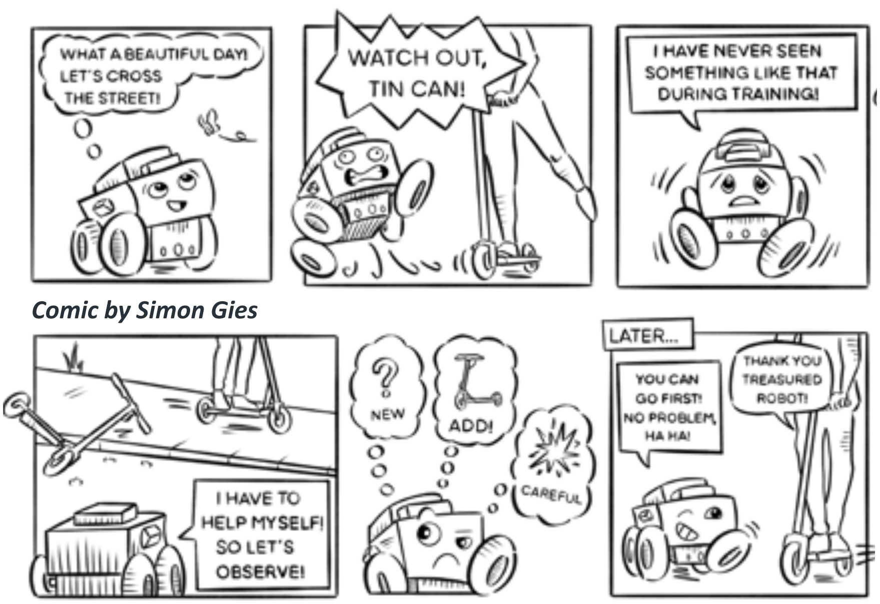
After five days of great technology, presentations, and friendship, you can see the **BEST OF ECCV** in our two-part feature this month. The first part showcases the **three Best Paper award winners** and their exceptional works on page 4. The second part focuses on **medical computer vision workshops**, starting on page 42.

On page 34, Ioannis Valasakis introduces scientific code and two retinal image datasets for you to experiment with, related to screening for the widespread ocular disease glaucoma, which can lead to permanent loss of vision if not diagnosed and treated early.

Also, this month, we speak to Bruno Occhipinti, CCO at **Qritive**, a start-up with a bold mission to solve the inefficiencies and subjectivities in cancer diagnosis by harnessing the power of AI. Find out more on page 60.

We hope you enjoy this November issue of **Computer Vision News**. Don't forget to tell your friends about us and **subscribe for free!**

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

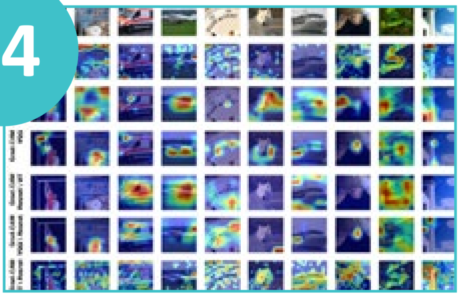


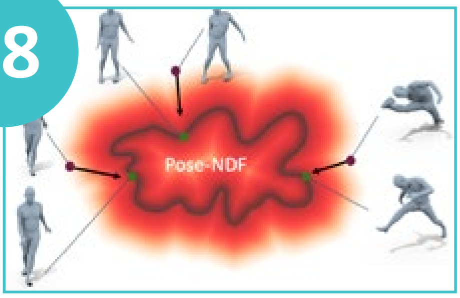
Comic by Simon Gies

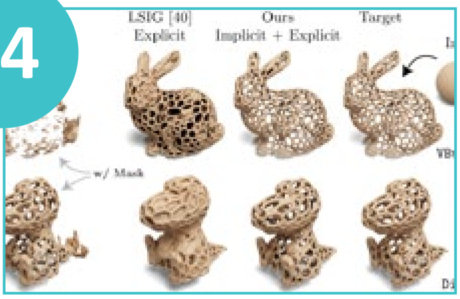
Follow Us

Do you want to learn more? Read about this groundbreaking work on page 28!

Computer Vision News

04


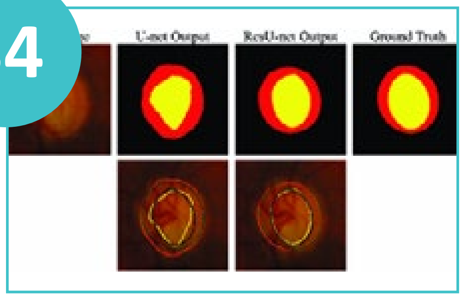
08


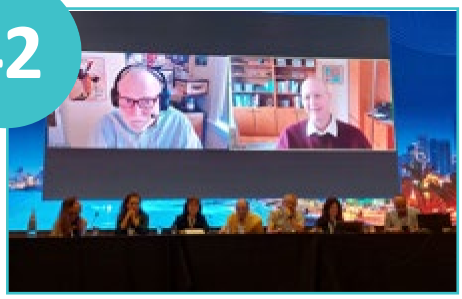
14


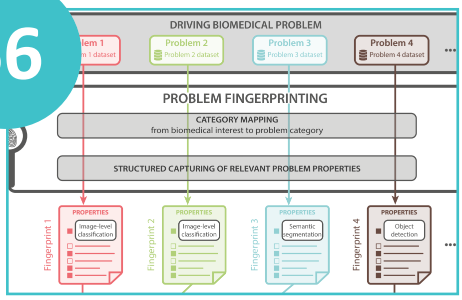
18


Medical Imaging News

32


34


42


56


ON THE VERSATILE USES OF PARTIAL DISTANCE CORRELATION IN DEEP LEARNING

Xingjian Zhen is a PhD student at the University of Wisconsin-Madison, advised by Vikas Singh. Zihang Meng is a Research Scientist at Meta AI in New York. He recently graduated from UW-Madison with the same advisor. Rudrasis (Rudra) Chakraborty is a Senior Research Scientist at Butlr, a start-up in California. He graduated with a PhD in Computer Science in 2018. Together, they just scooped the Best Paper Award at ECCV 2022 for their work, introducing the traditional statistical domain of distance correlation into deep learning.
They are here to tell us all about it.

With so many neural networks available, choosing which one to use for a given task can be difficult. There may be several different options, which poses an important question: **How do you know if Network A is better than Network B**?

The community often addresses such a simple question in an unnecessarily complex way. In approaching this work, Xingjian, Zihang, and Rudra believed that no one had looked for a simple answer.

"We're conditioning one network on another network," Xingjian tells us. *"Most of the time, when people are comparing networks, they care more about the performance or accuracy, but questions regarding the information remaining in the network have not been well studied. We've borrowed the partial distance correlation method from the statistical domain to remove the information from one network off another pre-trained network. We compare the remaining information to see if it's still meaningful regarding our data. Then we can say that one network contains more information than another."*

Rudra adds:

*"**Correlation is very simple.** Everyone can understand the relationship between two things. That's why we borrowed something from the statistics textbook, which is well understood and well known, to answer this question."*

Once the team had figured out the mathematical side, incorporating that into the engineering part was more challenging due to the large gap between theory and practice.

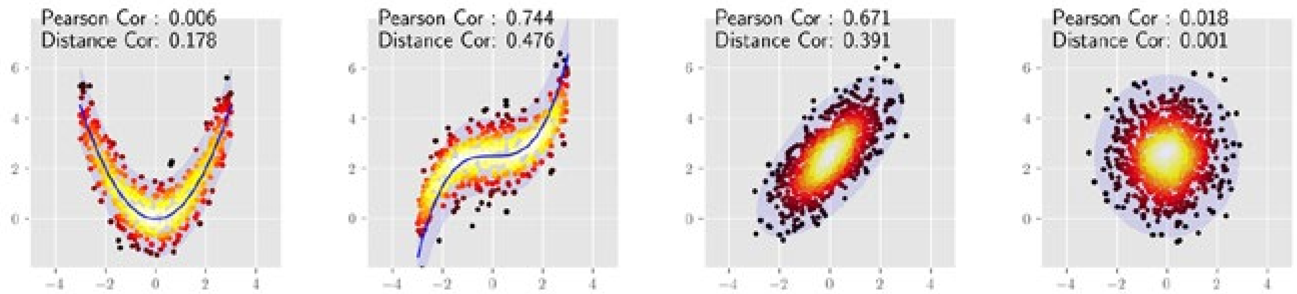


Fig. 1: Examples of Pearson Correlation and Distance Correlation in different settings. (a): $y = 0.5x^2 + 0.75n, n \sim \mathcal{N}(0,1)$; (b): $y = 0.15x^3 + 0.75n + 2.5, n \sim \mathcal{N}(0,1)$; (c): $\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 2.5 \end{bmatrix}, \begin{bmatrix} 1 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}\right)$; (d): $\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 2.5 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1.25 \end{bmatrix}\right)$

"We all know as vision and machine learning scientists that not all theory holds in practice, even if it's promising," Rudra points out. *"The challenging part is how to make it applicable in practice. Then comes the complexity, such as memory usage, how much GPU we have, and all those things."*

Understanding networks is a hot topic for researchers. While this work demonstrates how effective something as simple as correlation can be to help, the team is keen to point out that this is just one way to compare networks, and there are many other statistical tools that people could use, some simple and some more complex, that have not currently been explored in computer vision or machine learning. The problem is not yet fully solved, and they hope this work opens a path for people in academia and beyond to dig deeper.

"There are many applications we don't believe can be done in academia," Rudra continues. *"These are things only a company or industry can do, like self-driving, for example. They train huge networks, and we think our method could reduce their training and computational time to make them more effective. I hope this opens more research in that direction."*

Xingjian adds: *"As a university researcher, it's more like you're on the demo side. We show things work, and then to make it profitable or applicable at a business level,*

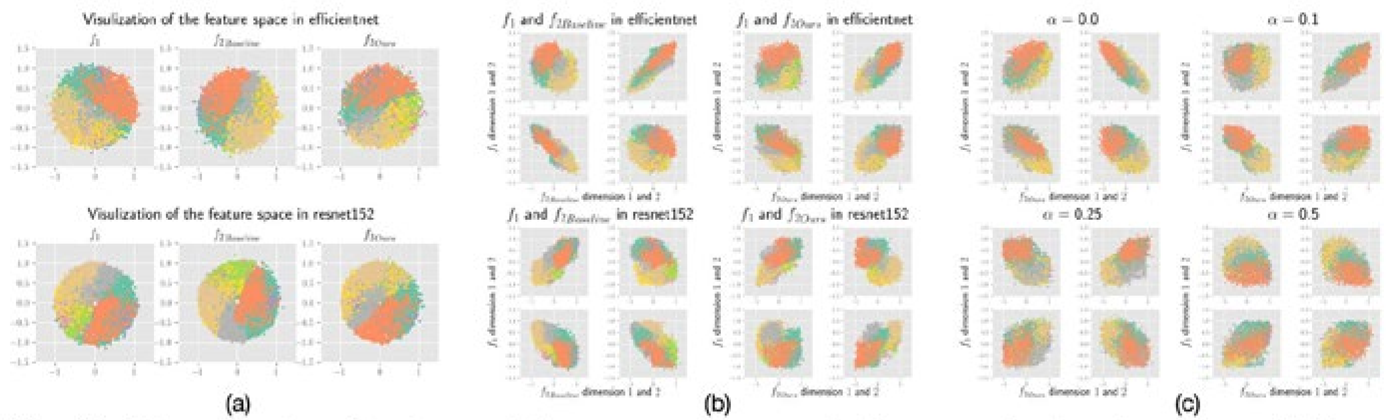


Fig. 2: Picasso visualization of features space and the correlation between different models. **(a)** Feature space distribution. **(b)** Cross-correlation between the feature space of f_1 and f_2 trained with/without DC. We get better independence. **(c)** By increasing the balance parameter α of DC loss, Mobilenet is more independent to f_1.

Generated images pertaining to the different ages for the same individual. While the results are qualitative, perceptually the generated results appear meaningful.

it's another story. For now, we're too early to think about that."

Winning **Best Paper award** for this work at a top-tier conference like ECCV is a fantastic achievement and not to be underestimated. The team would like to thank everyone who made it possible, including their research funders – the National Institute on Aging (NIA/NIH) and the National Science Foundation – their reviewers, and the program committee. They even have a word of thanks for this magazine and our readers.

"In the morning on the day we found out we won, I was in bed, and I checked my email and saw 'Your paper has been selected for the Best Paper Award,' and it was an oh-my-god moment for me!" Xingjian recalls.

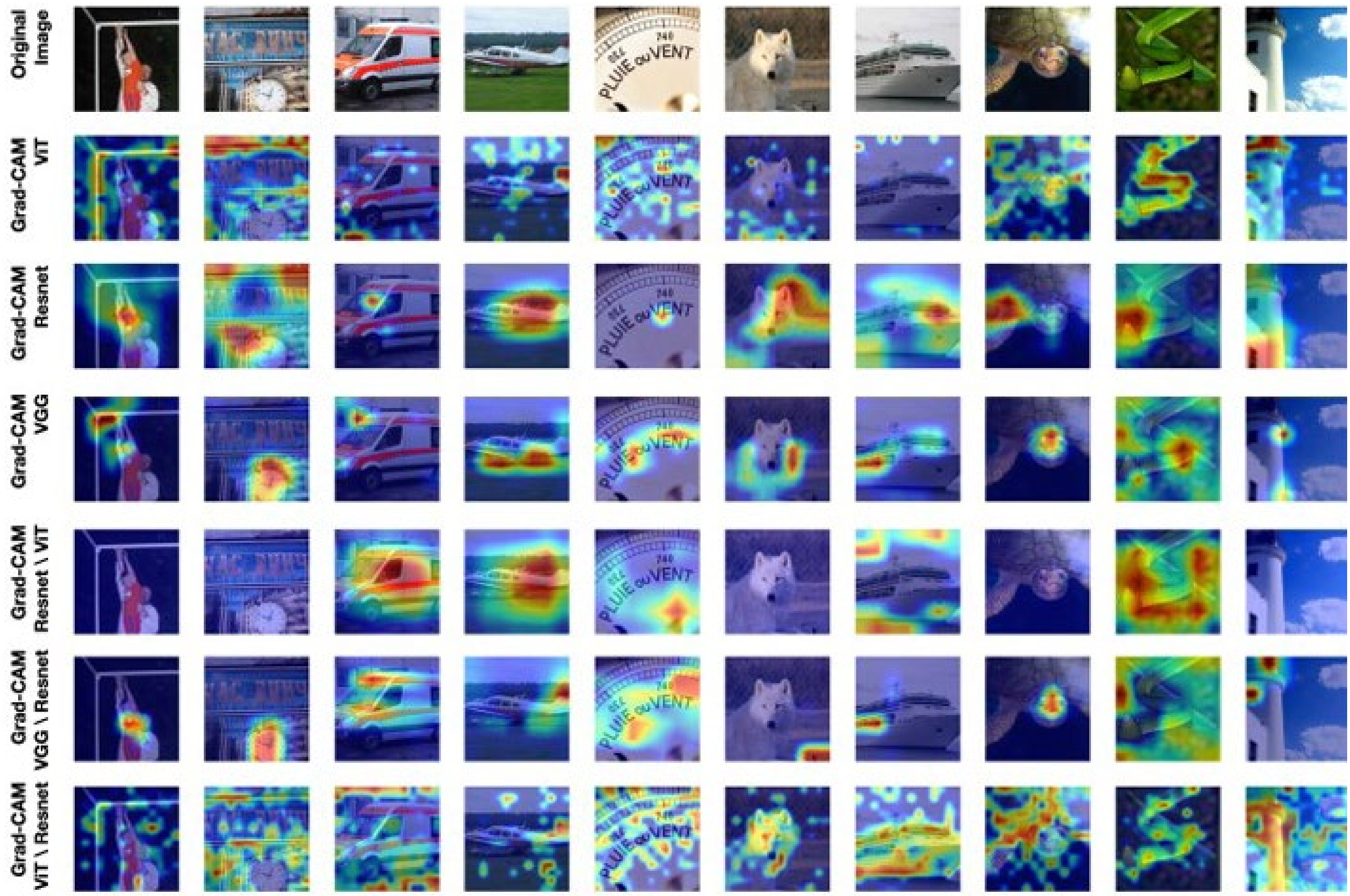
Rudra found out similarly: *"I was half asleep, and at first, I thought it was a dream. A very good dream!"*

Witnessing Xingjian deliver his acceptance speech live in Tel Aviv was particularly touching and moving for the community in the room, who have spent the last few years apart. Can the team tell us what they think swayed the judges to award their paper the top prize?

*"It's new and has opened the door for **many potential future directions**, either by following the idea along the distance correlation side or by introducing benefits from other well-studied statistical tools into this new domain of deep learning,"* Xingjian replies. *"I also think my presentation and visualization may have added some points. I hope so!"*

Rudra believes the judges saw the merit in **simplicity**.

"Most of the time, we overlook simple things, but if you look at well-addressed



problems in any field, the best solutions are generally the simple ones," he points out. *"The beautiful part is borrowing something from textbook statistics on machine learning and seeing how applicable it is in computer vision literature. We're all glad that the reviewers and committee members saw what we saw in this work."*

Zihang believes that researchers in the deep learning community have always been interested in **how to qualitatively or quantitatively understand what a network has learned**.

"Five or six years back, people did some beautiful visualizations of the neurons within deep neural networks to see what they had learned," he recalls. *"More recently, researchers tried to use* ***canonical correlation analysis (CCA)*** *to analyze the neurons and compare two different networks. Here, we're offering a new method that hasn't been used in the community before to efficiently compare two networks and offer some ability to condition on a network, which previous methods did not provide.* ***It's a simple and elegant way to solve a popular problem!"***

As we wrap up the interview, we uncover what may have been fueling the team's success when we ask them to tell us something about UW-Madison that we do not know.

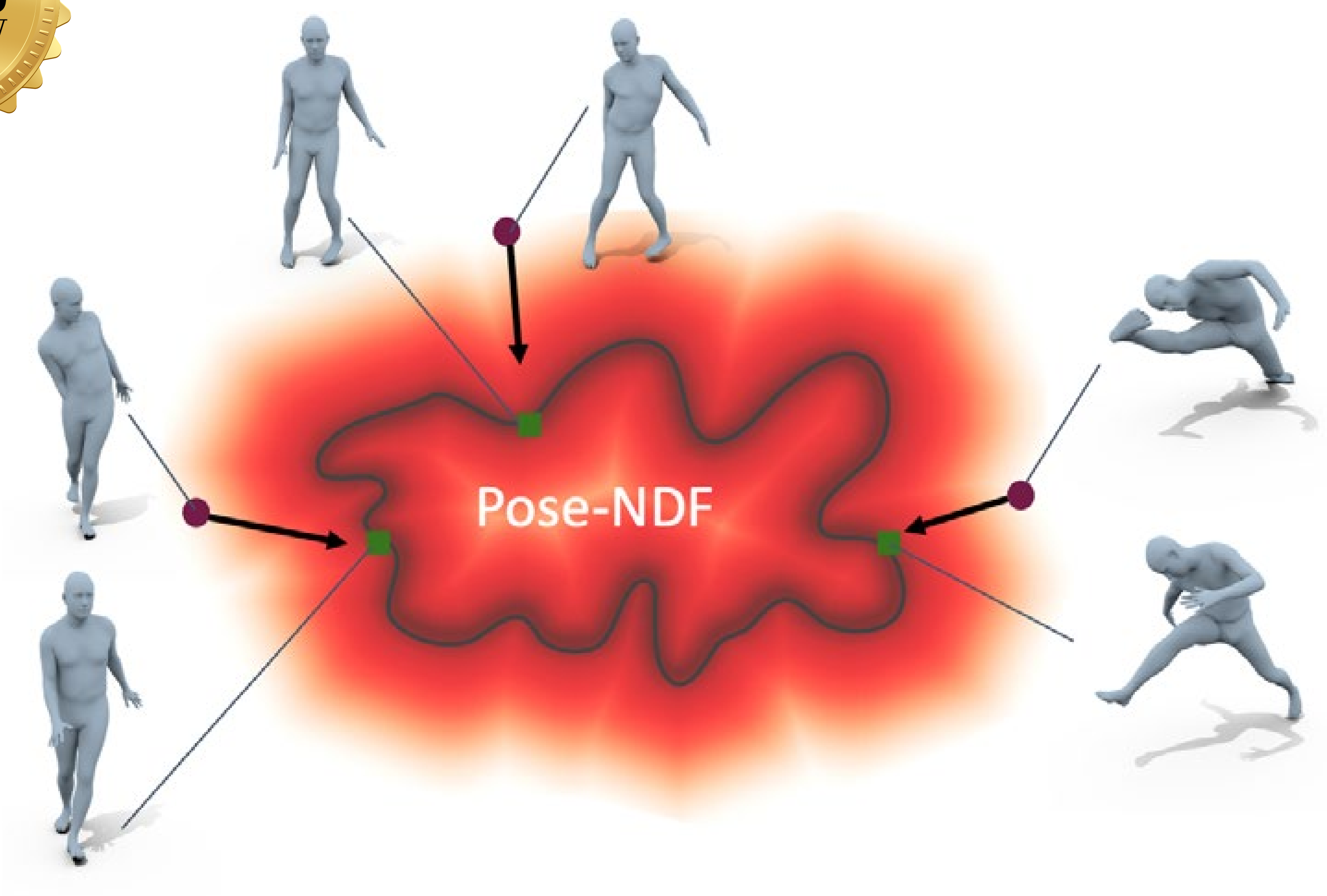
"It has the best ice cream of all universities!" Zihang laughs.

"That's what I wanted to say," Xingjian adds. *"I think Zihang is still missing The Daily Scoop."*

"He's right," Zihang confirms. *"I'm in New York, but I still miss the ice cream at Wisconsin-Madison!"*

POSE-NDF: MODELING HUMAN POSE MANIFOLDS WITH NEURAL DISTANCE FIELDS

Garvita Tiwari is a third-year PhD student at the University of Tübingen and the Max Planck Institute for Informatics in Germany under the supervision of Gerard Pons-Moll. She is also undertaking an internship at Meta in London. Her paper presents a continuous model for plausible human poses based on neural distance fields and has just won a Best Paper Honorable Mention award at ECCV 2022. She speaks to us about her work.



3D human modeling is about modeling how a human behaves, moves, and looks in the real world, which is closely related to seeing a person in a virtual or augmented reality environment. In this work, Garvita explores how they move. It introduces a human pose prior model, which is essential for modeling realistic human poses and avoiding results such as the unrealistic rotation of joints.

In contrast to previous work on human pose priors, which tackled it from a very specific approach, this work merges two different fields.

"***Neural implicit fields*** *are a recent advancement and important for modeling 3D shapes,*" Garvita explains. "*We extend the same idea to a very high dimension and use the same formulation* ***for a complicated data distribution and a complicated space****. We learn a prior using that approach.*"

The work has several real-world applications, including in the entertainment sphere. In animation, for example, animating a character dancing to a song would usually require expensive animators working long hours. However, this research direction and these models make it possible to achieve the same result much faster and semi-automatically.

"*This is the first time anyone has done this,*" Garvita tells us. "*I guess that was appreciated by the program and area* *chairs, which is why we won an award."*

Gerard Pons-Moll had the original idea to use neural fields, which the research group ran with, and then Garvita, co-author **Jan Eric Lenssen**, fellow PhD student Julian Chibane, and others started working informally on the initial idea for this work. Ultimately, Garvita decided to take it forward in a more concrete direction. Later, **Tony Tung** and **Nikolaos Sarafianos** joined the project.

*"Tony and Nikolaos are at **Meta Reality Labs Research**,"* she tells us. *"We used to think that by collaborating with them, our research would be slightly limited, but that's not been the case. Whatever research you want to work on, Tony is always supportive. It's super nice to work with those guys."*

Garvita says that being accepted for an oral and poster at ECCV this year has been a massive help in thinking about this work's future direction.

"During my sessions, people asked a bunch of questions, and they were all good future directions," she recalls. *"It's so nice that people are already thinking about the future of this work. It's the kind of idea that opens up a new field. Once you read the paper, you immediately think about follow-ups. The model is unconditional, so it doesn't have many*

Garvita and team celebrating their Best Paper Honorable Mention award. At their sides, ECCV 2022 Program Chairs Shai Avidan, Gabriel Brostow (left), Tal Hassner and Giovanni Maria Farinella (right).

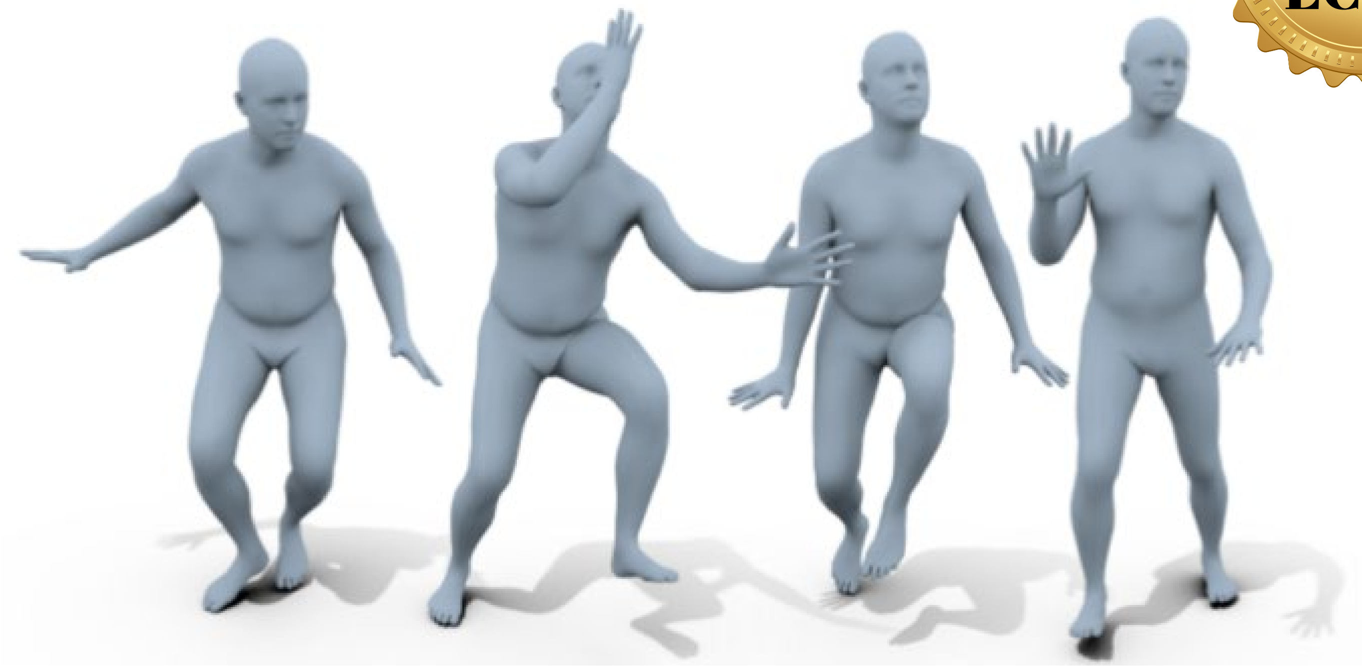
applications right now, but if you want to make it more applicable, you can make it conditional. There are many possible follow-ups because it's an initial idea in this direction, and you can make it work for specific applications."

If Garvita could add one more feature to the model, what would it be?

"As I mentioned, ***we are working in a very high-dimensional space****, and there is no way to see what is happening in that high dimension,"* she points out. *"If it were possible, I'd like to see what's happening there because there's no direct way to do that."*

Garvita came to Germany from India and has always wanted to explore different places and work in a field where she got to read a lot and really dig into research.

"I got lucky that I did my Master's with Gerard and really enjoyed it," she says. *"It's a great combination of practical*



and theoretical research and keeps me motivated to continue in this direction."

Gerard Pons-Moll is a friend of our magazine. Can Garvita give us an insight into what it is like to work with him?

"It's fascinating working with Gerard and the whole group," she responds. *"With Gerard, the research feels very light and informal. We hang out a lot as a group. We can have cool discussions over lunch and coffee sessions. Sometimes on weekends, we go for drinks, and those discussions are nice. Like every supervisor, Gerard has high expectations for his students, but that's good because we can see we're getting better day by day."*

What does she plan to do next?

"For all PhD students, that's a difficult question!" she laughs. *"I'd like to learn more about theoretical machine learning or deep learning in the future because I don't have much exposure to that. I want to work for a company after my PhD."*

Tübingen is a traditional university city in central Baden-Württemberg, Germany. Now in her third year there, what can Garvita tell us about it?

"It's a really lovely city," she smiles. *"It feels like you're in a bubble. If you go just outside of Tübingen, people have very different political views and lifestyles, but in Tübingen, everyone is so aware of the environment, social issues, the economy, and politics, and they are super nice. The old town has a great history. I know so many historical facts about places in Tübingen because I like to know the history of the place I live in!"*

We point out that if anyone chooses to visit now, they will know whom to call for a tour guide.

"Yes, I will show you around!"

My FIRST ECCV

Neerja Thakkar is a 3rd year PhD student at UC Berkeley under the supervision of Jitendra Malik.
She presented her work "Studying Bias in GANs Through the Lens of Race".

Neerja with fellow UC Berkeley PhD students and co-authors Tim Brooks (advised by Alyosha Efros), Vongani Maluleke (advised by Angjoo Kanazawa and Jitendra Malik) and Devin Guillory (advised by Trevor Darrell).

BEST OF ECCV

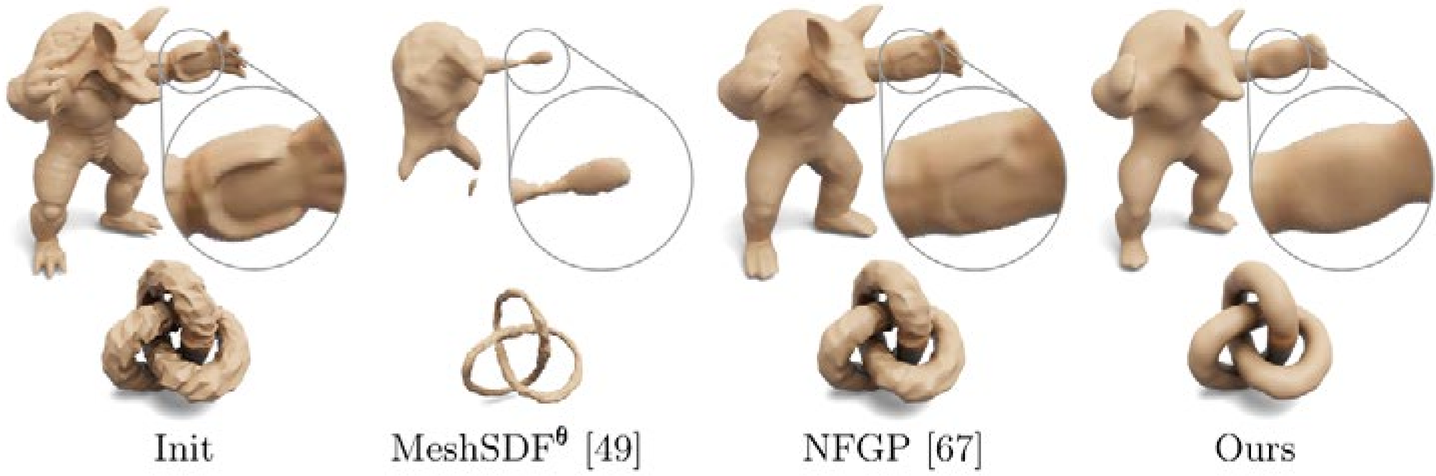
A LEVEL SET THEORY FOR NEURAL IMPLICIT EVOLUTION UNDER EXPLICIT FLOWS

Ishit Mehta is a PhD student at the University of California San Diego under the supervision of Manmohan Chandraker and Ravi Ramamoorthi. His paper proposing a theoretical framework to analyze explicit and implicit surface representations in shape optimization has just won a Best Paper Honorable Mention award at ECCV 2022.
He speaks to us about his work.

Shape optimization is everywhere. It is a critical component in manufacturing, engineering, and many other fields and is used to model aircraft, cars, and satellites. One approach to optimizing these shapes is by using explicit surface representations, which means performing shape optimization with triangle meshes and point clouds, for example.

"The downside of using explicit representations in geometry optimization is the problem of controlling the topology of the surfaces. ***Topology is what the connectivity of the shape looks like,****"* Ishit tells us. *"The topology of a sphere and a donut are different because a donut has a hole, whereas a sphere doesn't. That's the classic example. If you use explicit surface representations, going from a sphere to a donut is difficult because it's a discontinuous operation. You need to create a hole. That's hard to do with standard optimization."*

Recovering or optimizing geometry is a very common process. It is called an inverse problem, where you know the



external constraints and want to recover what the shape looks like.

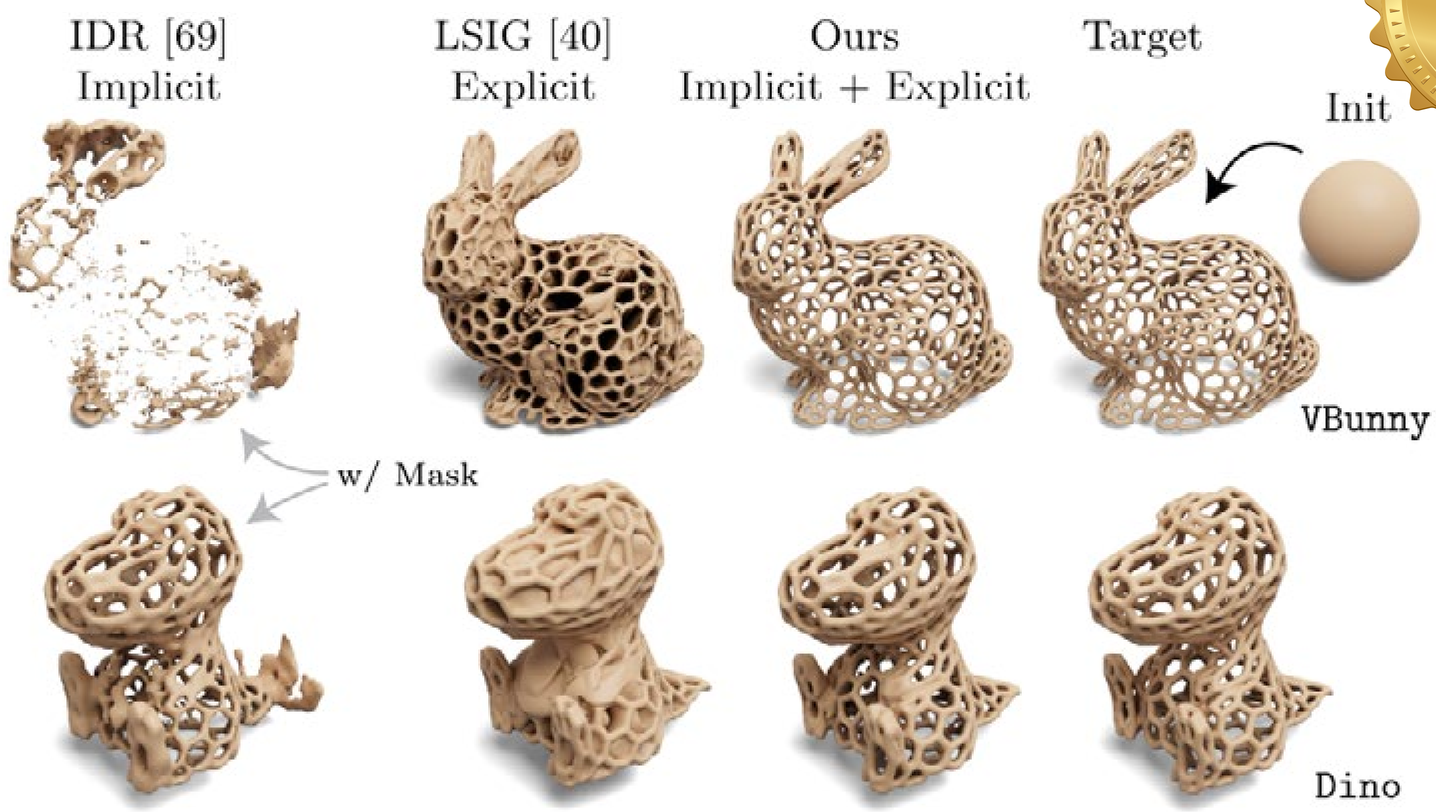
Say you are designing an aircraft, and you know what the plane is going to fly, what the air is going to look like, and what speed it's going to fly at, and you want to design it so there is minimal drag and it runs efficiently with the least amount of fuel consumption. You may want to design a bridge to have maximum strength using the least amount of material, and you know how many people will walk on it, the force field and pressures of the river, and the type of soil involved. Maybe, using **computational fluid dynamics**, you are designing efficient aerodynamic components for high-speed Formula One race cars. In all these kinds of settings, geometry optimization is essential.

Did Ishit encounter any challenges in the course of this work?

"One of the biggest problems working in this space is there is just too much happening," he points out. *"This is a problem faced by most PhD students working in this area. So many papers are coming out every week, and trying to keep up with that literature while writing papers and doing relevant work is a challenge. You might have solved something, or you might have an interesting idea, but if you are six months late, then the relevancy of that idea goes down."*

The work brings ideas from many areas, including **geometry processing**, **computer vision**, and **computer graphics**. These areas all intersect but operate in silos. People working in these specific areas do not always read the literature from other areas. The challenge here is how to write things that are relevant to all three communities simultaneously and are interesting and useful.

"The inception of the idea behind the paper comes from the ***inverse rendering community****, where you're given a set of images in a constrained environment or an unconstrained environment, and you want to recover all the properties of the scene where those images are captured,"* Ishit explains. *"For instance, you have a glass,*



take 10 images of it, and want to recover its geometry, what it looks like, the material aspect of it, and the environment in which it was captured."

It is an age-old problem. In the 1960s, people from **cognitive science and psychology** were trying to figure out how humans could do this. We can walk in these environments naturally, we're able to grab things easily, and we know what the shape of the glass looks like. We do not have a 3D model in our head of that glass, we are visualizing it using two images, but we can still recover that.

"*It's been looked at from so many directions, and I think it's* ***one of the most interesting problems in computer vision***," he tells us. "*Recovering the material aspect of the glass, which is almost translucent, and the geometry together is hard. If I visualize it, I see there are interesting and complex visual effects that you observe. None of*

the methods right now can model all of this complexity to recover all of these things together. It's like skin. If I want to recover the geometry of my hand using images, it will be hard to do because of how the light interacts with our skin. It's not a simple thing."

There are many complex materials around us with complex geometry, each needing a solution. These challenges are far from being solved. There are a lot of open problems here. Can we expect to see Ishit receiving another award at CVPR next year for a paper bringing forward these kinds of technologies and further solving these kinds of problems?

"I doubt it!" he laughs. *"Getting a **Best Paper Award** is extremely rare in your career, and it's almost unfair to award two or three papers. It's a very subjective thing. There are 1,500 papers at ECCV, which are all equally good. I'm lucky to be working on the right problem at the right time."*

We can tell the experience humbles Ishit. What does he think the judges saw in his paper that was particularly special?

*"**I think it's the simplicity of it**,"* he answers. *"We tried to draw ideas from three disparate communities. The paper appeals to most of these audiences. Our theoretical developments are easy to parse through, and **anyone with a high school math education can probably understand what's going on**. We tried to focus on writing the paper well so that it reaches a bigger audience and doesn't obfuscate the details. I think that more than the scientific contribution won it for us."*

Ishit is keen to point out that the work draws on the collective knowledge built by the community. It does not claim to be the latest and greatest method out there; it is just trying to have a conversation about what is a good geometry representation for shape optimization.

"We're just trying to provide a lens to analyze existing techniques," he adds. *"That's all we're trying to do here. Our goal is not to be the best method out there. My only hope is that people get more interested in this fascinating area of **inverse shape optimization**. If this paper even manages to bring a few more people into this space, I think that will be a job well done!"*

BEST OF ECCV

"JUST KEEP TRYING, EVEN IF YOU'RE TERRIFIED INSIDE!"

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 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.

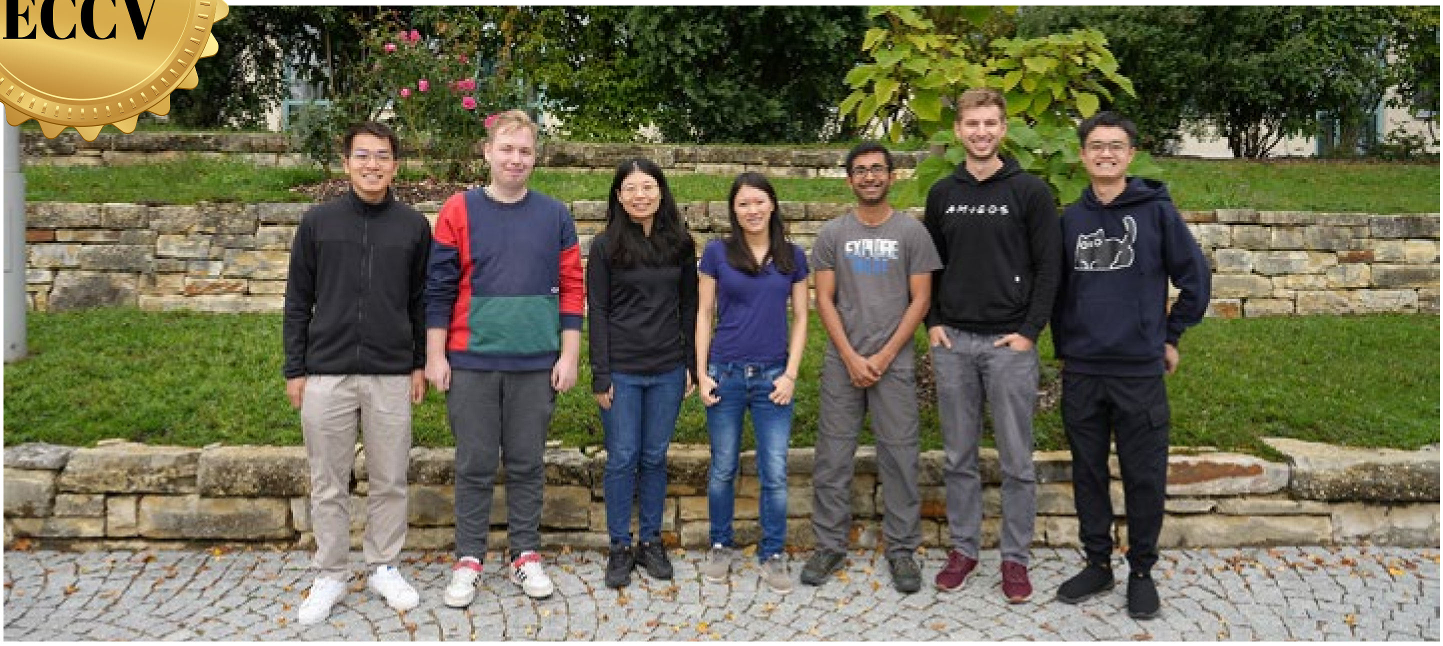
What was the experience like to leave your native country and go to Europe?

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

BEST OF ECCV



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 Filder 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!

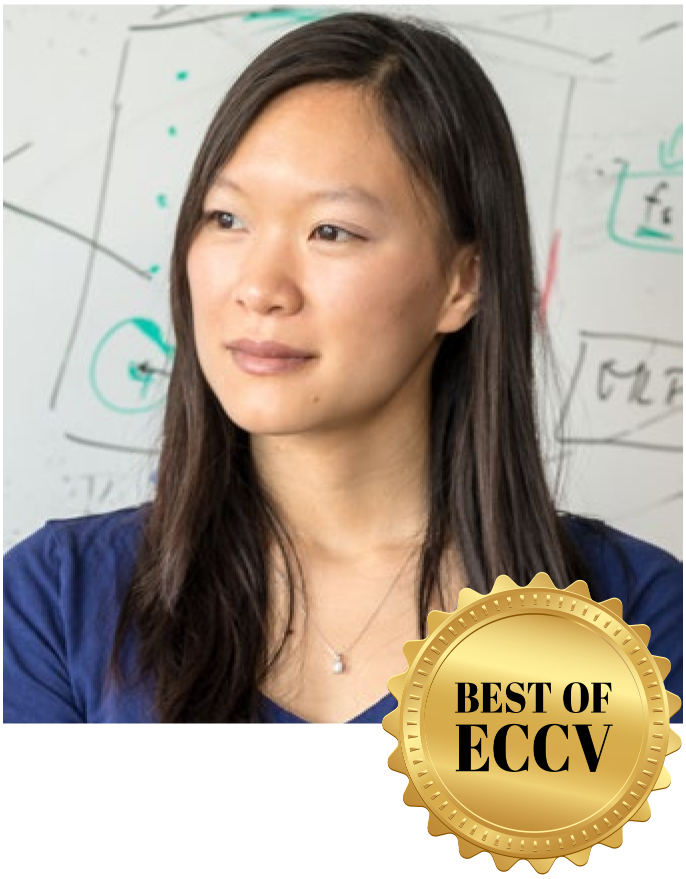
No broken bones yet! [*laughs*]

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?

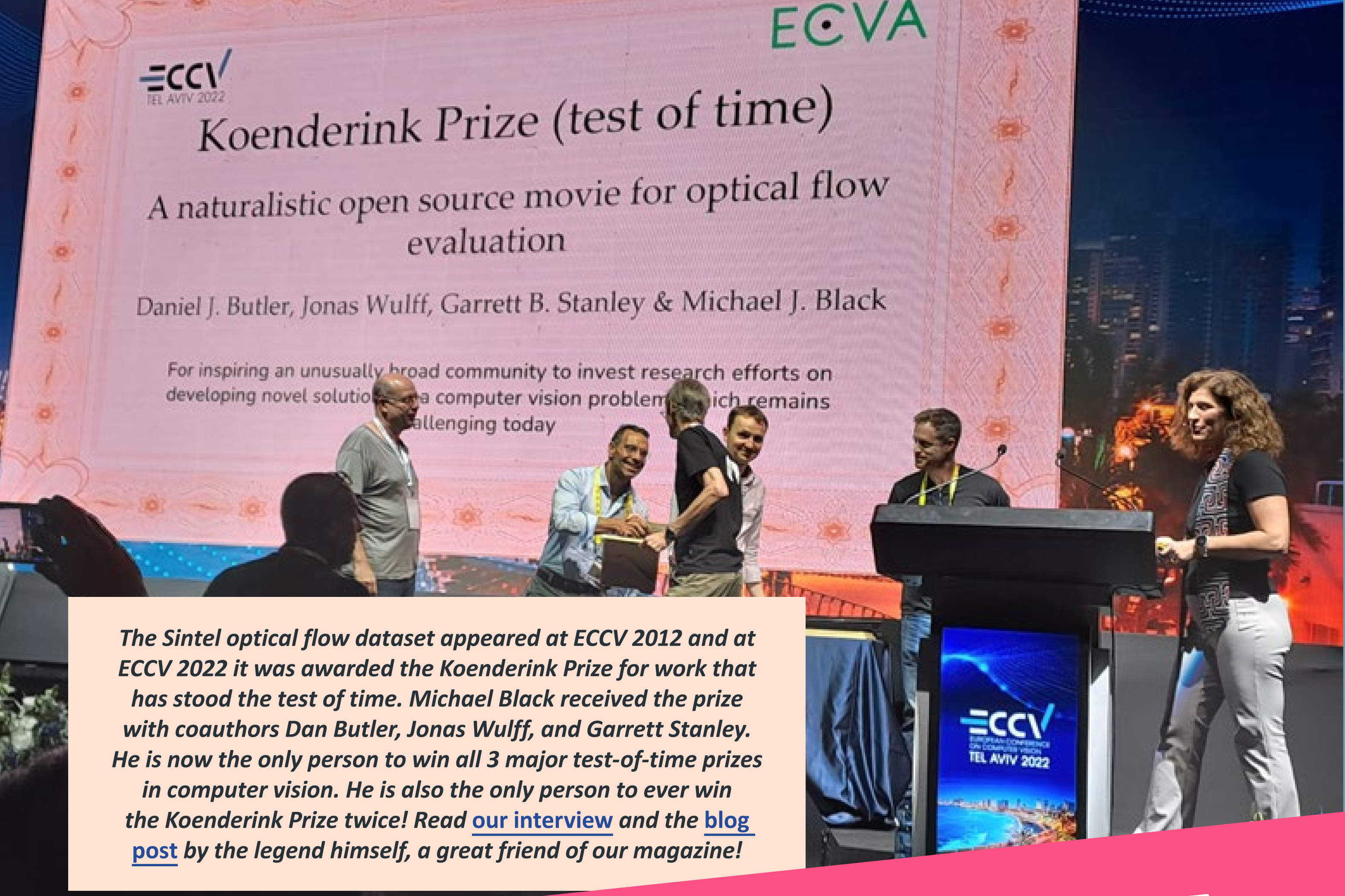
The work that people have done leveraging diffusion models in order to provide supervision for 3D is pretty cool. So, for instance, DreamFusion, I think it’s called, actually showed that you can get some notion of 3D structures out of image-based models. I don't think it will solve everything, but I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

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

Ralph celebrating the Ukrainian presence at ECCV with Tetiana Martyniuk. Tetianka presented her poster FEAR: Fast, Efficient, Accurate and Robust Visual Tracker.

FEAR / NO FEAR

BEST OF ECCV



The Sintel optical flow dataset appeared at ECCV 2012 and at ECCV 2022 it was awarded the Koenderink Prize for work that has stood the test of time. Michael Black received the prize with coauthors Dan Butler, Jonas Wulff, and Garrett Stanley. He is now the only person to win all 3 major test-of-time prizes in computer vision. He is also the only person to ever win the Koenderink Prize twice! Read our interview and the blog post by the legend himself, a great friend of our magazine!

STANDING THE TEST OF TIME

Andrea Burns is a fifth year PhD Candidate at Boston University advised by Kate Saenko and Bryan A. Plummer. She presents her ECCV paper on a new dataset for vision language navigation with unknown command feasibility, the first to study an interactive instruction following task in which the request may not be possible in the visual environment.

BEST OF ECCV

SEEN AT ECCV

Qunjie Zhou is a 4th year PhD student at TUM at the Dynamic Vision Learning group, led by Laura Leal-Taixé. She is presenting her poster at ECCV 2022 about geometry-online camera pose estimation.

Hermann Blum recently completed his PhD in the Autonomous Systems Lab of ETH Zürich. His research focuses on enabling robots to semantically understand their environments in complex open-world settings. Probably best known for the fishyscapes.com benchmark, he first worked on anomaly detection in semantic segmentation and later on self-improving perception systems. He recently joined the Computer Vision and Geometry Lab of ETH Zürich. Congrats, Doctor Hermann!

In construction, delivery, transportation, care, or household work there are many applications where robots could take over dangerous tasks, help to reduce energy and resource use, or enable novel design principles. However, they require robots to operate in our everyday environments instead of factory floors, which often requires semantic scene understanding. Data-driven algorithms, especially deep learning, have greatly improved the capabilities of machines to detect and identify objects. Yet, these methods fail under domain shift and in the presence of unseen object types.

Hermann's thesis investigates the problem of robotic scene understanding in open-world, everyday environments. The proposed systems are able to identify unknown parts of a scene, and even adapt and improve their perception capabilities in these environments fully autonomously.

In a first step, his thesis introduces a benchmark (fishyscapes.com) to measure how well robotic perception methods can identify outliers. It focuses on anomalies for semantic segmentation in urban driving, where unknown categories should be correctly segmented from images. As part of a larger collaboration a second benchmark was later added (segmentmeifyoucan.com) that focuses on real-world images and detection of any obstacle on the road. The analysis of existing solutions then revealed that good anomaly segmentation methods often had lower segmentation accuracy on the known classes. To counter this, Hermann worked on SynBoost, a wrapper method that is still the SOTA for detecting anomalies from given segmentation networks without messing with their training.

The second part of the thesis follows the question how robots can improve their semantic segmentation autonomously by combining recent advancements in self-supervision and continual learning. Through self-supervision, perception systems can be trained without human input. That makes it possible to adapt these systems to the robot's environment during autonomous deployment. On the other hand, continual learning methods make it possible to integrate all the knowledge a robot gathers when deployed to different environments or executing different tasks sequentially. By combining the two, the thesis studies a robot that from experience alone builds up a perception system that works well in all explored environments. The advantage of continual learning shows up in those results where a robot that has already seen some other environments will adapt better to a target than an inexperienced robot.

It all comes together in a final system that can incrementally learn those parts of a scene that were identified as unknown by an anomaly detector.

On hermannblum.net you can find many videos of real-world robotic experiments related to these works.

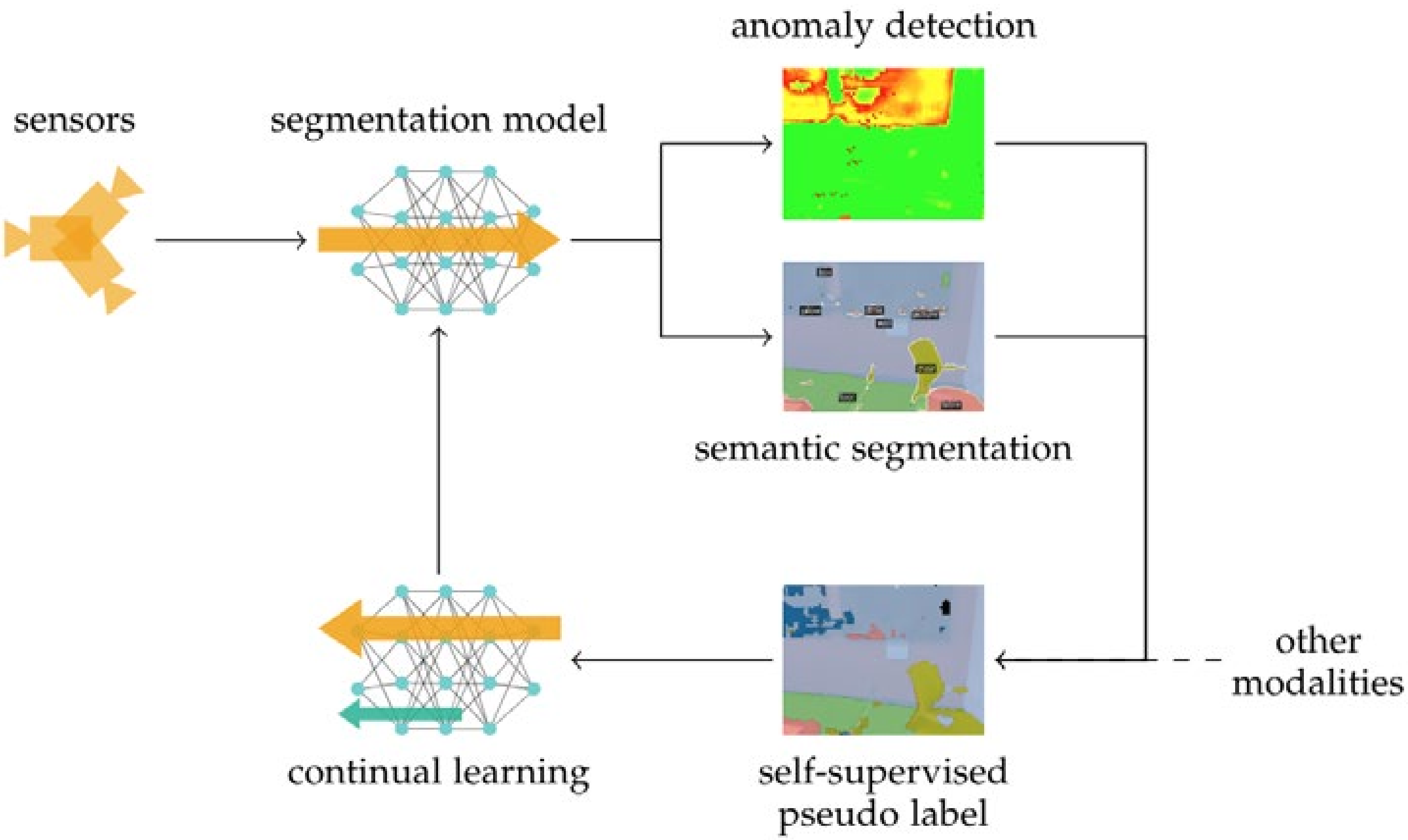


Figure 1: Overview of Hermann's approach to self-improving robotic perception. In this example, the segmentation model has never seen TVs before. Consequently, they are detected as anomalies and a self-supervised pseudo label is created that adds TVs as a new category.

COMPUTER VISION EVENTS

Deep Learning Summit Toronto, Canada 9-10 November	SIPAIM Valparaiso, Chile 9-11 November	Innovation Summit San Francisco, CA 14-16 November
GeoMedIA Workshop Amsterdam, the Netherlands 18 November	MICAD University of Leicester, UK 20-21 November	NeurIPS New Orleans, LA and virtual 28 Nov. - 9 Dec.
TechEx / AI & Big Data Expo Global London, UK 1-2 December	SIGGRAPH Asia Daegu, South Korea 6-9 December	
The AI Summit New York New York, NY 7-8 December		
VCIP Visual Communications and Image Processing Suzhou, China 13-16 December		

FREE SUBSCRIPTION
(click here, its free)

Did you enjoy reading Computer Vision News? Would you like to receive it every month?

Fill the Subscription Form
it takes less than 1 minute!

SUBSCRIBE!

Join thousands of AI professionals who receive Computer Vision News as soon as we publish it. You can also visit our archive to find new and old issues as well.

We hate SPAM and promise to keep your email address safe, always!

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

NOVEMBER 2022

2D-TO-3D KNEE BONES RECONSTRUCTION - SUCCESSFUL PRELIMINARY RESULTS FROM CLINICAL STUDY

Total knee arthroplasty (TKA) is a common orthopedic procedure, conducted when the knee joint malfunctions and cannot be repaired. Classic open surgery resulted in large incisions, and success depended on the surgeon's experience. New technological improvements brought assistant robots into the operation room (OR) and democratized TKA. Most **robotic assisted surgeries (RAS)** include a planning stage, where a 3D model of the knee is acquired and used for prosthesis assessment approach planning. This model is produced using dedicated algorithms from a knee CT scan, as it provides a high-resolution 3D scan, resulting in an accurate model.

RSIP Vision has recently begun developing a solution for **reconstructing an accurate 3D model of the knee from two X-ray images, anterior-posterior (AP) and lateral**. Creating such a solution obviates the need for a CT scan, benefitting the procedure from two aspects - clinically and operationally. Standard X-ray imaging results in significantly less harmful radiation exposure to the patient compared to a CT scan, ultimately reducing procedural risk. Additionally, X-ray imaging has better accessibility than CT, it is more widely reimbursed in the U.S. healthcare system, and it is often lower in cost. However, X-ray lacks the spatial information available in CT, therefore, it is challenging to reconstruct a 3D model with comparable accuracy.

To address the challenge of 3D reconstruction from standard X-ray images, both the unique AI technology, including dedicated deep neural networks,

and a custom tailored, low-cost calibration device developed by RSIP Vision were employed. After acquisition of AP and lateral X-rays, **the software automatically creates a detailed 3D model of the knee joint**, specifically the tibia and fibula, which can then be used for surgical planning, providing access to precision planning for a much wider patient population.

This solution is currently assessed in **Assuta Medical Center, Tel Aviv**, a leading medical center in Israel, as a part of a clinical trial. The accuracy of the solution is proven by comparing the resulting 3D models to the ground truth patient anatomy given in a corresponding CT scan. The similarity between the X-ray based model and the ground truth suggests that this solution can be used for surgical planning during TKA.

RSIP Vision plans to expand this clinical study into a medical center in the USA, and ultimately receive FDA clearance. In parallel, RSIP Vision is working to extend this solution to additional anatomical regions such as the hip and shoulder joints.

We hope that in the near future TKAs will be planned and executed using this solution. Implant suitability and positioning will be tested on the 3D knee model, and the surgical plan will be designed. Throughout the procedure navigation will be conducted based on the surgical plan and the 3D model, and all this without the need for a CT scan, making this procedure more available. RSIP Vision continues working day and night to expedite the marketing of this solution for better orthopedic healthcare.

GLAUCOMA AND SEGMENTATION

IOANNIS VALASAKIS, KING'S COLLEGE LONDON

@WIZOFE

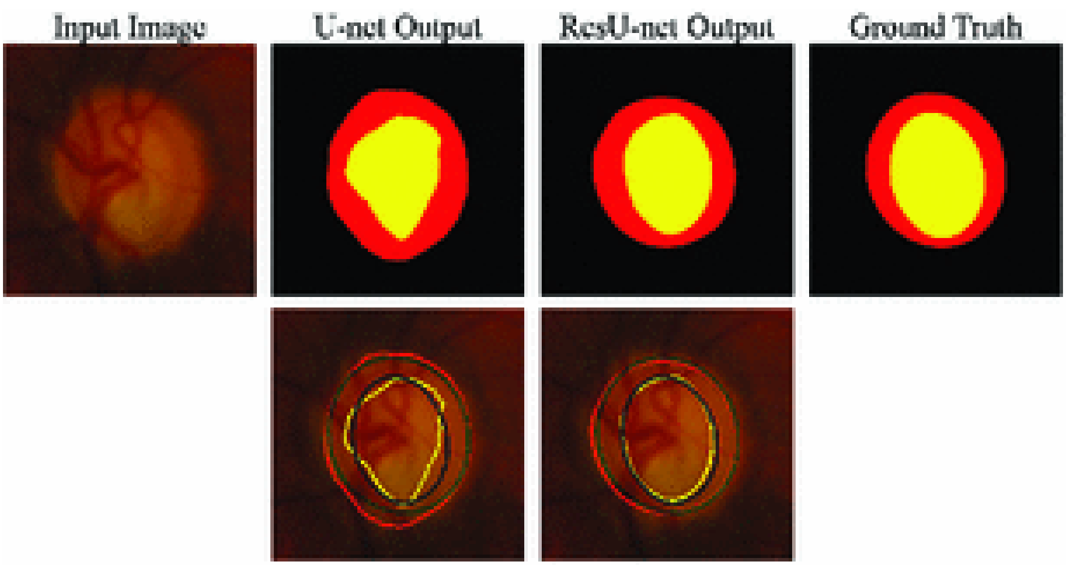
Hi again! This month's issue is really special for me as it includes some state-of-the-art techniques, so you can experience scientific code in the works! The code is tested but as some parts were removed for brevity, feel free to reach out and I'll try to solve any problems arising! Of course, do the same if you have more ideas, examples, and requests. Thank you kindly for always sending nice words and suggestions for future issues.

Let's always be kind to people around us, educate and be patient! Keep up the amazing work you are doing in your life, professional and academic world, and most of all: enjoy it ☺

Glaucoma and segmentation

Glaucoma is a highly threatening and widespread ocular disease which may lead to a permanent loss of vision. One of the important parameters used for Glaucoma screening is the cup-to-disc ratio (CDR), which requires accurate segmentation of the optic cup and disc.

Glaucoma is a disorder which may lead to blindness and currently affects a large number of people worldwide. The assessment of the Optic Disc (OD) and optic cup is important for diagnosis and it's done using 2D color fundus images. **Cupping** signifies an enlargement of the cup due to the loss of optic nerve fibers. The enlargement of the cup concerning OD, measured as a vertical cup-to-disc ratio (CDR) is one of the most important indicators of the disease. This necessitates **accurate segmentation** of the optic cup and disc.



Here are some examples taken from Sharath M's paper about Joint Optic disc and cup segmentation

Coding area!

The dataset used in that example is the RIM-ONE and DRISHTI which you can find online (either on GitHub or various Google Drive Sources). A small introduction to the code is that it trains a UNet network, similarly as we've seen before and then performs a cup segmentation using two approaches.

I am not including the UNet creation but you can refer to our older code articles or of course look at many great resources online if you want to modify it and create your own!

Without much further ado, let's reach the code and see for ourselves!

```
train=True
Epochs=30
preload=False
hough=False
rams=False
dynAug=False
depth = 6

print(depth,chan,ir,rand_sv,ds,zone)

6 64 1.1 42 MIX CUP
```

Imports

This is a part where you need to do all your imports.

Data preparation

The data from RIM-ONE and DRISHTI are prepared and modified to be ready for the model injection. Notice we are using the H5 file format for the data.

```
Xori= np.asarray(h5f2['DRISHTI-GS/orig/images'])
disc_locations = np.asarray( h5f2['DRISHTI-GS/512 px/disc_locations'])
FC = np.asarray(h5f2['DRISHTI-GS/512 px/file_codes'] )
indDRI = np.arange(0,Xori.shape[0])
bsqside=np.maximum((disc_locations[:,3]-disc_locations[:,1]),(disc_locations[:,2]-disc_
locations[:,0]))
isize=Xori.shape[1]

  Yf = np.asarray(h5f2['DRISHTI-GS/512 px/cup'])

  Xc = [Xori[i][index512_resize(disc_locations[i][0],isize):index512_resize(disc_
locations[i][0]+bsqside[i],isize), index512_resize(disc_locations[i][1],isize):index512_
resize(disc_locations[i][1]+bsqside[i],isize)]
                   for i in range(len(Xori))]

  Yc=[Yf[i][disc_locations[i][0]:disc_locations[i][0]+bsqside[i], disc_locations[i]
[1]:disc_locations[i][1]+bsqside[i]]
                   for i in range(len(Xori))]

X=[cv2.resize(img, (128, 128),interpolation=cv2.INTER_NEAREST) for img in Xc]
```

```
X=np.asarray(X)
am=np.amax(X)
X=X.astype(np.float32)/am
print(np.amax(X),np.amin(X))

Y=[cv2.resize(img, (128, 128),interpolation=cv2.INTER_NEAREST)[..., None] for img in Yc]
Y=np.asarray(Y)
ym=np.amax(Y)
Yf=Y/ym
#get the binary mask
Yb=(Y>0.5).astype(np.float32)

print(np.amax(Yf),np.amin(Yf))
print(X.shape)
```

```
1.0 0.0
1.0 0.0
(50, 128, 128, 3)
```

Similar to the RIM-ONE dataset.

```
Xvori= h5f1['RIM-ONE v3/orig/images']
disc_locationsv = h5f1['RIM-ONE v3/512 px/disc_locations']
Glauv=np.asarray(h5f1['RIM-ONE v3/512 px/is_ill'] )
FCv=np.asarray(h5f1['RIM-ONE v3/512 px/file_codes'])
indRIM=np.arange(0,Xvori.shape[0])
bsqsidev=np.maximum((disc_locationsv[:,3]-disc_locationsv[:,1]),(disc_locationsv[:,2]-
disc_locationsv[:,0]))
low_cont_n=0;
ivsize=Xvori.shape[1]

  Yvf = h5f1['RIM-ONE v3/512 px/cup']

  Xvc = [Xvori[i][index512_resize(disc_locationsv[i][0],ivsize):index512_
resize(disc_locationsv[i][0]+bsqsidev[i],ivsize), index512_resize(disc_locationsv[i]
[1],ivsize):index512_resize(disc_locationsv[i][1]+bsqsidev[i],ivsize)]
                     for i in range(len(Xvori))]

  Yvc=[Yvf[i][disc_locationsv[i][0]:disc_locationsv[i][0]+bsqsidev[i], disc_locationsv[i]
[1]:disc_locationsv[i][1]+bsqsidev[i]]
                     for i in range(len(Xvori))]


Xvn=[cv2.resize(img, (128, 128),interpolation=cv2.INTER_NEAREST) for img in Xvc]

Yv=[cv2.resize(img, (128, 128),interpolation=cv2.INTER_NEAREST)[..., None] for img in
Yvc]
Yv=np.asarray(Yv)
ym=np.amax(Yv)
Yvf=(Yv/ym).astype(np.float32)
```

Train and test

Here we create the train and test variables. The example shown here it's just for variable one but similar would be for variable two. I will leave this here as a small exercise for the reader!

```
rng = np.random.RandomState(rand_sv)

#DRISHTI
X_train1t, X_test1t, Y_train1t, Y_test1t, F_train1t, F_test1t ,ind_train1, ind_test1=
train_test_split( X, Yb, FC, indDRI,test_size=0.25,random_state=rng)

#RIM ONE
X_train2, X_test2, Y_train2, Y_test2, F_train2, F_test2, ind_train2, ind_test2= train_
test_split( Xv, Yvf, FCv, indRIM, test_size=0.25,random_state=rng)

X_test_dri=np.copy(X_test1t)
Y_test_dri=np.copy(Y_test1t)
F_test_dri=np.copy(F_test1t)
```

Preprocessing for cup segmentation

Here for each image CLAHE is performed like this:

```
for i in range(16):
  X_traine[i*size:(i+1)*size]=X_train1
  Y_traine[i*size:(i+1)*size]=Y_train1

base= 16*size

for j in tqdm_notebook(range(X_train1.shape[0])):
  X_traine[base]=skimage.exposure.equalize_adapthist(X_train1[j], clip_limit=0.04)
  Y_traine[base]=Y_train1[j]
  base+=1
  X_traine[base]=skimage.exposure.equalize_adapthist(X_train1[j], clip_limit=0.02)
  Y_traine[base]=Y_train1[j]
  base+=1
  X_traine[base]=modify_brightness_p(X_train1[j],0.9)
  Y_traine[base]=Y_train1[j]
  base+=1
  X_traine[base]=modify_brightness_p(X_train1[j],1.1)
  Y_traine[base]=Y_train1[j]
  base+=1

X_testc=np.copy(X_test)
Y_testc=np.copy(Y_test)
```

Shuffling the dataset

Of course, we need to add a variable of randomness, so let's shuffle the dataset!

```
X_trains,Y_trains = unison_shuffled_copies(X_traine, Y_traine)
X_tests,Y_tests = unison_shuffled_copies(X_teste, Y_teste)

# Provide the same seed and keyword arguments to the fit and flow methods
seed = rand_sv
batch_s = 120

strategy = tf.distribute.OneDeviceStrategy(device="/gpu:0")

model = get_unet_light(img_rows=128, img_cols=128)
model.compile(
    optimizer=tf.keras.optimizers.Adam(learning_rate=.001, ),
    loss=log_dice_loss, metrics=[dice_coef]

# Understand the model
model.summary()
```

```
I won't include the output for brevity, but you'll see all the layers. Notice

X_tests.shape

(2380, 128, 128, 3)

# Model Training Step


if(train):
  history = model.fit(
    X_trains,Y_trains,
    epochs=Epochs,
    batch_size=128,
  )
```

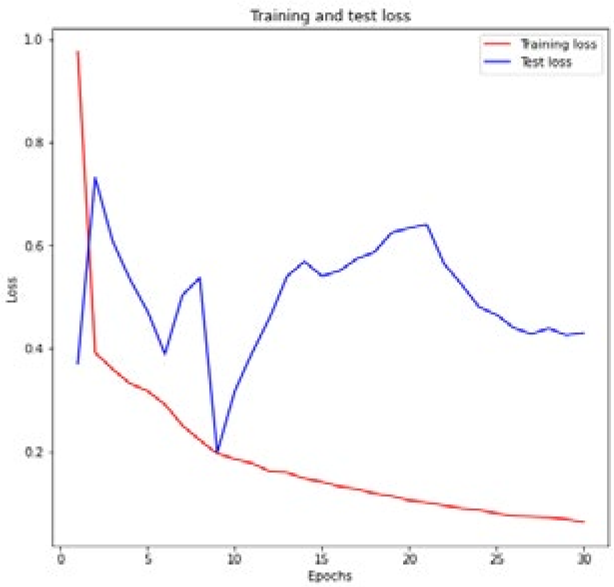
Saving the weights

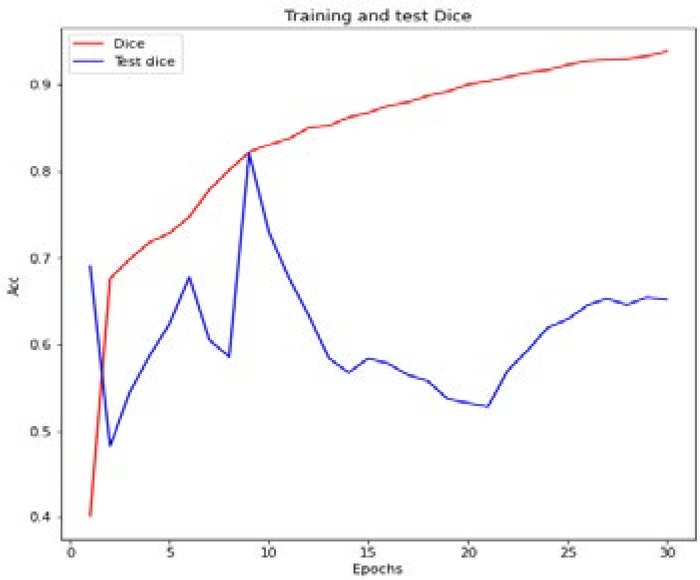
An important part which you should always remember is to save weights during training!

```
if(train):
  model.load_weights(os.path.join('/content/drive/My Drive/Medical Image/cnn_term_paper/
model2_cup_last_check_point.hdf5'))

# evaluate the model with the test data

result = model.evaluate(X_tests[0:700,:,:,:],Y_tests[0:700,:,:,:])
print("log dice loss for test set = ", result[0], ' ||| ',"dice coeffecient (accuracy)
for test set = ", result[1])
```

Loss visualisation

Here we visualize the change in accuracy and loss during training and plotting the log dice loss





Final stage!

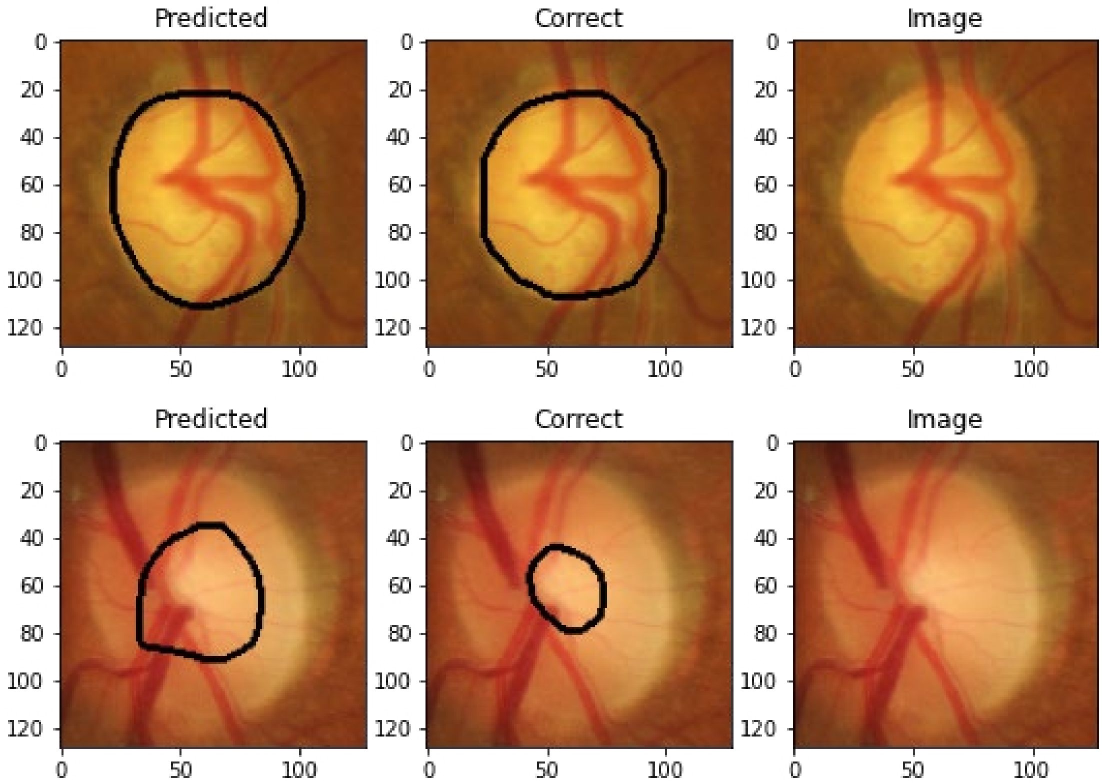
And now let's look at how to model and our segmentation is performing!

```
test_s=int(X_testc.shape[0])
test_s=test_s-test_s%8
test_s
```

```
112
```

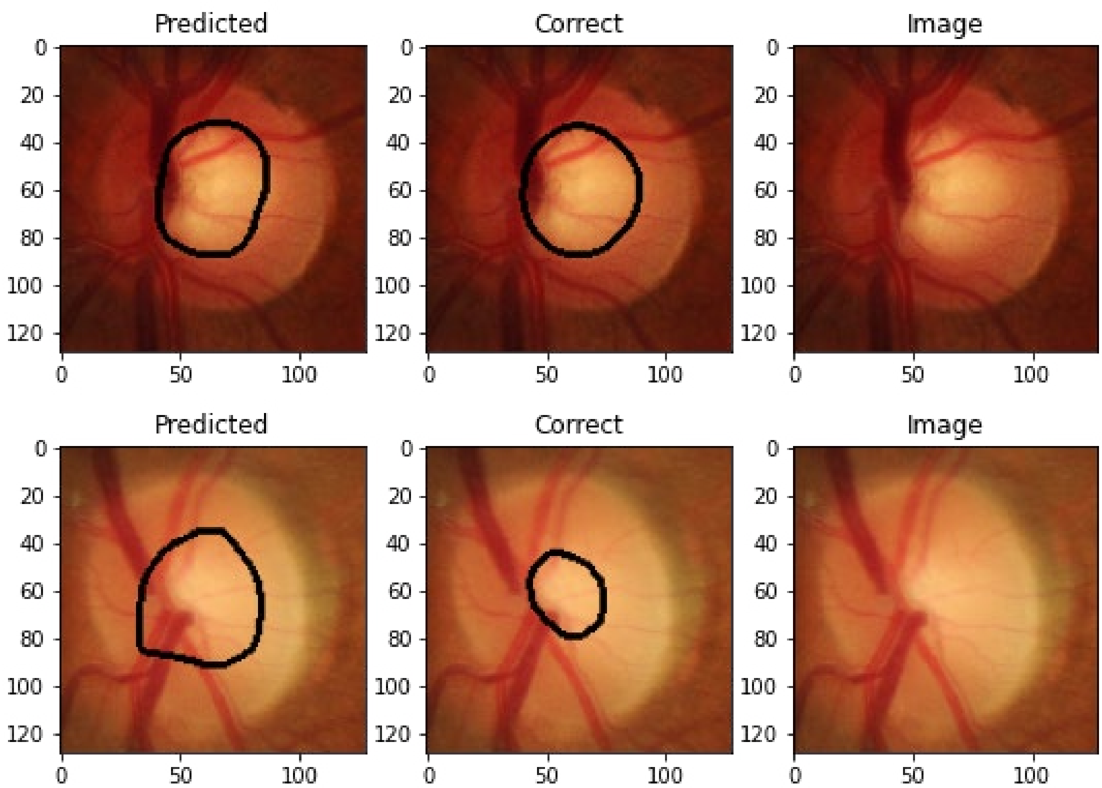
```
# Defining metrics again

smooth=1e-07
```

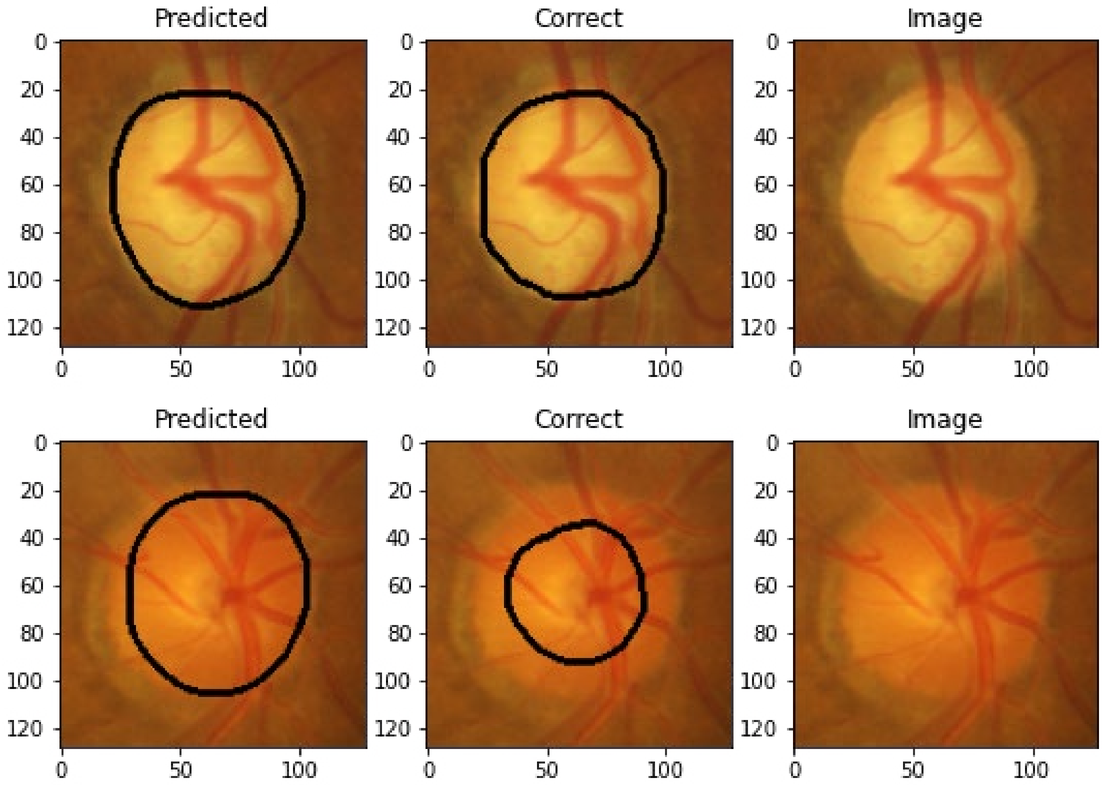
```
Dice mean= 0.83547586  Std= 0.103501074  best= 0.9616514  worst= 0.5241682
IOU mean= 0.73005486  Std= 0.14256199  best= 0.92613536  worst= 0.35516796
image #11
0.9616514
image #39
0.5241682
CPU times: user 4.77 s, sys: 25.1 ms, total: 4.8 s
Wall time: 5.63 s
```



And do the same for the RIM dataset:

```
Dice mean= 0.80241454  Std= 0.09940979  best= 0.9423816  worst= 0.5241682
IOU mean= 0.6809287  Std= 0.13192643  best= 0.8910413  worst= 0.35516796
image #26
0.9423816
image #0
0.5241682
```





And the DRISHTI

```
    edgesc = color.gray2rgb(1.0*edgesc)
    ax.imshow(img*edgesc)
    ax.set_title('Correct')
    ax = fig.add_subplot(1, 3, 3)
```

```
(13, 128, 128, 3)
```

{"model_id":"84b48c7de6f44158b8b43299c6ecf555","version_major":2,"version_minor":0}

```
Dice mean= 0.90240574  Std= 0.07061479  best= 0.9616514  worst= 0.67507553
IOU mean= 0.8287271  Std= 0.10222741  best= 0.92613536  worst= 0.50952005
image #11
0.9616514
image #7
0.67507553
```

Conclusion

This is an amazing result but of course, there's room for improvement. I intentionally change the last part of the network, to let you be able to have space to experiment and make the network better (both in UNet terms but also to achieve higher accuracy). What do you think? I would be very happy to see any images of yours using the same dataset.

We'll be back in two months for another coding article and a small challenge (it seems you do enjoy those). Until then, keep experimenting and be curious! ☺

BEST OF EC

MEDICAL COMPUTER VISION (MCV) WORKSHOP

by Tammy Riklin-Raviv and Moti Freiman

Moti Freiman, Ayelet Akselrod-Ballin, Tammy Riklin-Raviv, Tal Arbel

The ECCV Medical Computer Vision)MCV) workshop took place in Tel Aviv hosting 250 in-person attendees and additional 250 online participants. Being part of the long-standing series of workshops at CVPR the MCV workshop aims to bring closer the medical image computing, computer-assisted intervention and the computer vision communities by providing a dedicated forum to exchange ideas, encourage potential new collaborations and interdisciplinary efforts, and brainstorm new machine learning and computer vision applications in healthcare.

The full workshop day offered scientific presentations (oral and posters) of peer-reviewed workshop papers and a panel discussion as well as invited talks by distinguished speakers and a research perspective panel discussion hosting senior radiologists and researchers with diverse background in academia. The workshop organized by (add names and links) had been sponsored by the Technion Human Health Initiative, Ben Gurion University of the Negev, Nvidia, GE healthcare, and Teva.

The workshop started with a fascinating invited talk by **Yonina Eldar**, Professor at the Weizmann Institute. In her talk she demonstrated how current advances in deep-learning methods can benefit from combining model-based approaches leveraging classical approaches and prior physical models. Some examples she gave included the recovery of Ultrasound images from sub-nyquist sampled data, super-resolution ultrasound and tomographic imaging.

Alex Frangi, professor at the University of Leeds, gave an excellent invited talk about Precision Imaging for next Generation Regulation of Medical Products. He introduced the concept of **In-silico clinical trials (ISCT)** which are computer-based medical device trials performed on populations of virtual patients. ISCT use computer models/simulations to conceive, develop and assess devices with the intended clinical outcome explicitly optimised from the outset (a-priori) instead of tested on humans (a-posteriori). This includes testing for potential risks to patients (side effects) exhaustively exploring in-silico for medical device failure

Hybrid and in-person panel participants.

modes and operational uncertainties before being tested in human clinical trials. Alex demonstrated the ISCT concept by his work on the design of medical devices for vascular aneurysms.

The expert panel members: **Leo Joskowicz**, **Hayit Greenspan**, **Ilan Shelef**, **Ron Kikinis** and **William (Sandy) Wells** with **Tal Arbel** discussed the challenges they envisage for medical computer vision from both clinical and technical perspectives. Sharing their multiple-years perceptive in the field, they identified the need for more standard infra-structure for data analysis in hospitals, the need for improved confidence and explainability of deep-learning methods and the need to find large-scale solutions rather than focusing on specific applications. They also pointed out the need of a shared language between clinicians and technological people.

The peer-reviewed scientific program composed from 26 full papers, selected from 37 submissions by a rigorous peer-review process with at least three reviewers and an additional meta-reviewer for each paper. The geographic distribution of the submission well represented the international nature of the workshop with submissions from China, India, UK, USA, Israel, Germany, Belgium, Italy, Hong Kong, Korea, Abu Dhabi, among others.

Based on the reviews, the program committee awarded the best paper award to the paper: "**Anatomy-Aware Contrastive Representation Learning for Fetal Ultrasound**", authored by **Zeyu Fu** and colleagues from the University of Oxford, UK. The paper presents how a novel self-supervised contrastive representation learning approach combining anatomical information to design the latent space can be leveraged to improve downstream tasks such as cross-domain segmentation and classification of fetal ultrasound images.

Fetal imaging was also the topic of the runner-up paper by **Yael Zaffrani-Reznikov** from the Technion. Yael's presentation demonstrated how we can **non-invasively assess the functional maturation of fetal lungs by analyzing diffusion-weighted MRI data**. A key challenge in such analysis is the inevitable motion of the fetus. To address it, Yael presented **qDWI-Morph**, a physically-driven self-supervised deep-neural-network approach for simultnaous motion compensation and quantitative diffusion-weighted MRI analysis.

Another runner-up paper presented by **Alona Golts** of the Technion addressed the need for **cell-level automatic detection and classification required for personalized cancer treatment**. Alona presented a novel dataset of **Proteasomestained Multiple Myeloma (MM) bone marrow slides**, containing nine categories with unique morphological traits. With the relative difficulty of acquiring high-quality annotations in the medical-imaging domain, the dataset was intentionally annotated with only 5% of the cells in each tile. To tackle both cell detection and classification within a single network, she modeled these as a multi-class segmentation task, and trained the network with a combination of **partial cross-entropy and energy-driven losses**. It is encouraging to see that two out of the three first authors for the best papers are female.

Additional scientific oral presentations presented addressed the challenges of medical image reconstruction. On MRI reconstruction, the goal is the reduce overall acquisition time by undersampling the data frequencies. The novelty in the work presented by **Aniket Pramanik** from the university of Iowa resides in the combination of segmentation and reconstruction tasks to improve the model performance rather than focusing on the reconstruction itself and perform the segmentation as a subsequent task. Another work by **Bassel Hamoud** from the Technion focused on reconstruction of CT images from low-dose data. In this case the goal is to reduce the patient exposure to ionized radiation. While deep CNNs have demonstrated unprecedented performance in many image restoration tasks. However, their operation is inherently local, making them less suitable for handling CT reconstruction tasks, which suffer from global artifacts. While existing works circumvent this incompatibility using deeper networks with very large receptive fields, Bassel addressed this issue from the data perspective by proposing a new locality-preserving representation for the CNN's input data.

Other oral presentation topics included incremental 3D medical image segmentation using Multi-Scale Multi-Task distillation and predictive model to estimate withdrawal time in colonoscopies.

In addition to the talks there were 19 poster presentations on a variety of MCV topics which intrigued intense discussions.

Tal Arbel
McGill University

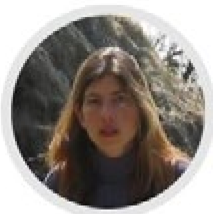
Ayelet Akselrod-Balin
Reichman University

Vasileios Belagiannis
Otto von Guericke University

Qi Dou
Chinese University of Hong Kong

Moti Freiman
Technion - Israel Institute of Technology

Nicolas Padoy
University of Strasbourg & IHU Strasbourg



Tammy Riklin-Raviv
Ben-Gurion University

Mathias Unberath
Johns Hopkins

Yuyin Zhou
University of California, Santa Cruz

BEST EC

AI-ENABLED MEDICAL IMAGE ANALYSIS WORKSHOP (AIMIA)

by Sara P. Oliveira

Sara Pires de Oliveira is finishing her PhD at the Faculty of Engineering of the University of Porto (FEUP) and a Research Assistant at INESC TEC. The AIMIA workshop was held virtually on October 24, 2022, on a full day of great talks and papers presentations, divided into the digital pathology and the radiology/COVID19 tracks. As part of the ECCV 2022, the workshop's goal was to provide a platform for scientific discussion on medical image analysis/processing, introducing the challenges of whole-slide images and CT/MRI/X-ray images to the computer vision and artificial intelligence community.

Besides the presentation of the submitted works, the workshop also hosted three invited keynote speakers, who shared their experiences on the development of AI models in digital pathology, with a journey from its initial steps, through what is being done now and also towards the future direction of the field.

At the beginning of the day, **Henning Müller** (Full Professor at the HES-SO Valais, Switzerland) started by presenting the ExaMode project, a practical approach for machine learning in digital pathology. Then, in the afternoon, **Inti Zlobec** (Professor at the University of Bern, Switzerland) led us on an odyssey through digital pathology with the question "**How will the digital revolution affect personalized medicine and what will the pathology of the future look like?**". Finally, Dimitris Metaxas (Distinguished Professor at Rutgers University, USA) concluded the workshop presentations with an overview of novel deep learning methods that provided new insights into the relationship between pathological image features, molecular outcomes and targeted therapies.

The digital pathology track included nine

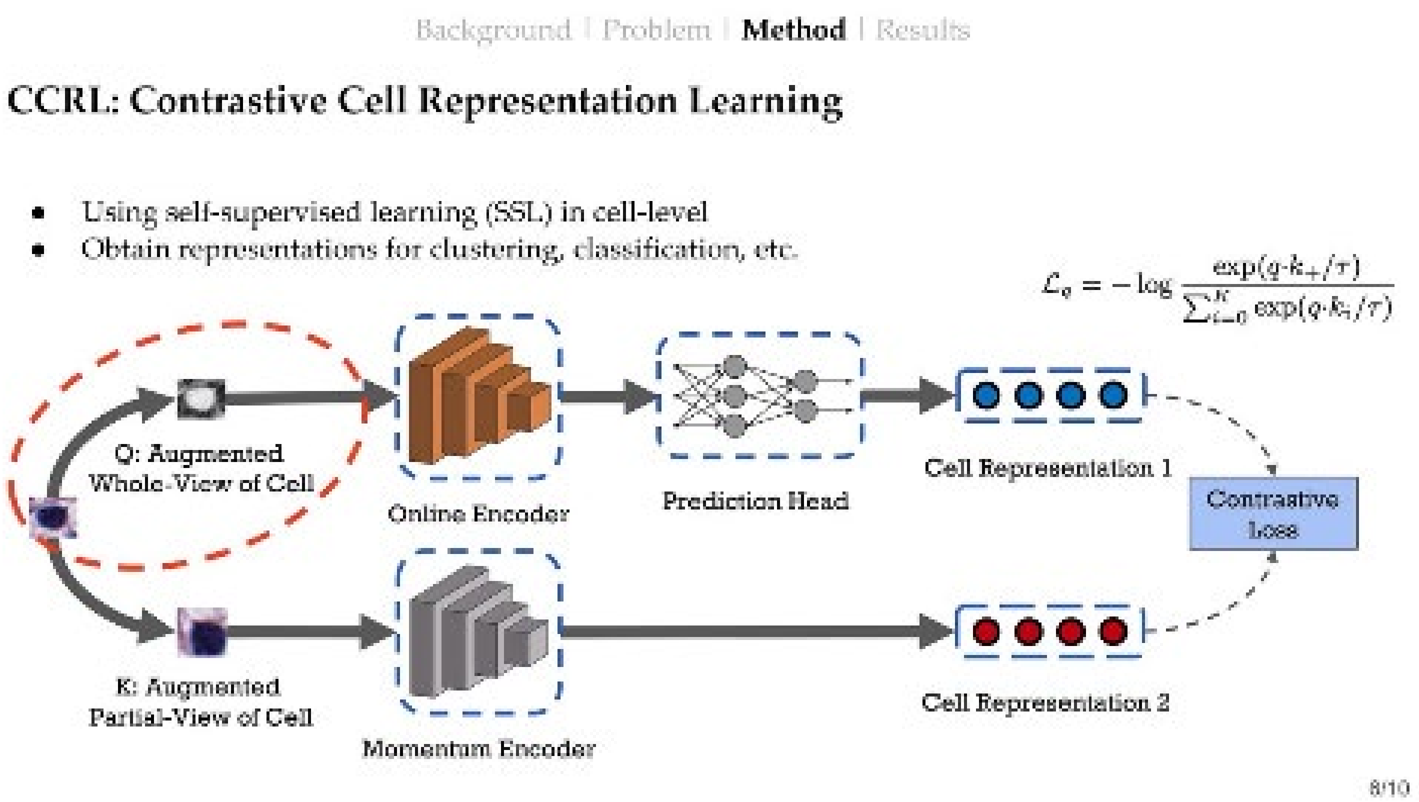
Prof. Henning Müller
HES-SO Vallais-Wallis

Prof. Inti Zlobec
University of Bern

Prof. Dimitris Metaxas
Rutgers University

presentations of cutting-edge research on both academia and industry. From attention-based models to contrastive learning and self-supervised learning, the authors presented several approaches to tackle tasks such as data annotation, cell identification or cancer grading.

Out of a set of five pre-selected papers, based on the scores given by the reviewers, the workshop participants voted for the **three best digital pathology works**. The best paper award went to Ramin Ebrahim Nakhli and colleagues, from the University of British Columbia, Canada. The paper proposes an approach to automatically identify cells on H&E-stained slides, without the time-consuming step of annotation. The authors propose a contrastive cell representation learning (CCRL) model for cell clustering that enables model training on larger datasets compared to previous methods. See video below.



The work presented by Dwarikanath Mahapatra, from the Inception Institute of Artificial Intelligence, UAE, came in second place, proposing an approach to improve the generalization capability of classification models for medical images. With graph neural networks, unsupervised learning, disentangled semantic and domain invariant structural features, the authors showed better performance across distribution shifts, in two types of data: chest X-ray images and histopathology images. The third prize was awarded to Romain Mormont *et al.*, from the University of Liège, France, for the work entitled "Relieving pixel-wise labelling effort for pathology image segmentation with self-training". In order to combine both exhaustively annotated and very sparsely-annotated images, the authors proposed a self-training based approach for image segmentation tasks,

Jaime Cardoso
INESCTEC, PT

Stefanos Kollias
NTUA, GR

Jeroen Van der Laak
Radboudumc, NL

Xujiong Ye
Uni. of Lincoln, UK

Luc Bidaut
Uni. of Lincoln, UK

Sara P. Oliveira
INESCTEC, PT

Mattias Rantalainen
Karolinska Institutet, SE

Cameron Chen
Google Health, USA

Dimitrios Kollias
QMUL, UK

Francesco Rundo
STMicroelectronics, IT

Giuseppe L. Banna
Portsmouth Uni., UK

Ana Monteiro
IMP Diagnostics, PT

which proved to be a good strategy to relieve labelling effort in the digital pathology domain. The authors were awarded with monetary prizes of €500, €300 and €150, respectively, kindly sponsored by Google.

The radiology/COVID19 track included seven presentations of works on explainable data models, medical image segmentation and classification, with a particular focus on attention-based and self-supervised models. Moreover, this track hosted six papers presenting the best-performing approaches developed for the 2nd COVID19 detection competition, four papers of the best-performing approaches developed for the COVID19 severity detection competition, as well as a paper describing the baseline models given in both competitions.

Isabel M. Pinto
IMP Diagnostics, PT

Diana Felizardo
IMP Diagnostics, PT

Pedro C. Neto
INESCTEC, PT

The first task of the competition was won by team ACVLAB, from the National Cheng Kung University, Taiwan, which proposed an effective spatial-slice feature learning (SSFL) for COVID-19 symptom classification on CT scans. The framework included a conventional CNN to first extract the feature embedding of each CT slice followed by a visual transformer-based sub-network to deal with feature learning between slices, leading to a joint feature representation. Then, the most relevant slices are automatically selected, which could effectively remove the uncertain slices as well as improve the performance of the SSFL model. This work was presented by Chih-Chung Hsu.

In this task, there was also another winning team, team FDVTS_COVID, from Fudan University and the Shanghai AI Laboratory, who came first in the second challenge as well. To detect COVID19 on CT scans, the authors have proposed an improved version of the strong 3D Contrastive Mixup Classification network, used as the baseline method, by adapting a pre-trained video transformer backbone to introduce natural video priors to COVID-19 diagnosis.

For the COVID-19 severity detection task, the authors adapted the same framework after a segmentation step to extract both the lung region and then the inner lesion region. The lesion segmentation mask serves as complementary information for the original CT slices. The team was represented by Jilan Xu.

Each winning group was awarded with a monetary prize of €200. The three prizes were sponsored by DERI, GRnet and ICCS, respectively.

The workshop was organised by an international committee that joined INESC TEC and IMP Diagnostics (Portugal), the National Technical University of Athens (Greece), Radboud University Medical Center (The Netherlands), Karolinska Institutet (Sweden), Google Health (USA) and the University of Lincoln and the Queen Mary University of London (UK).

Congrats, Doctor!!

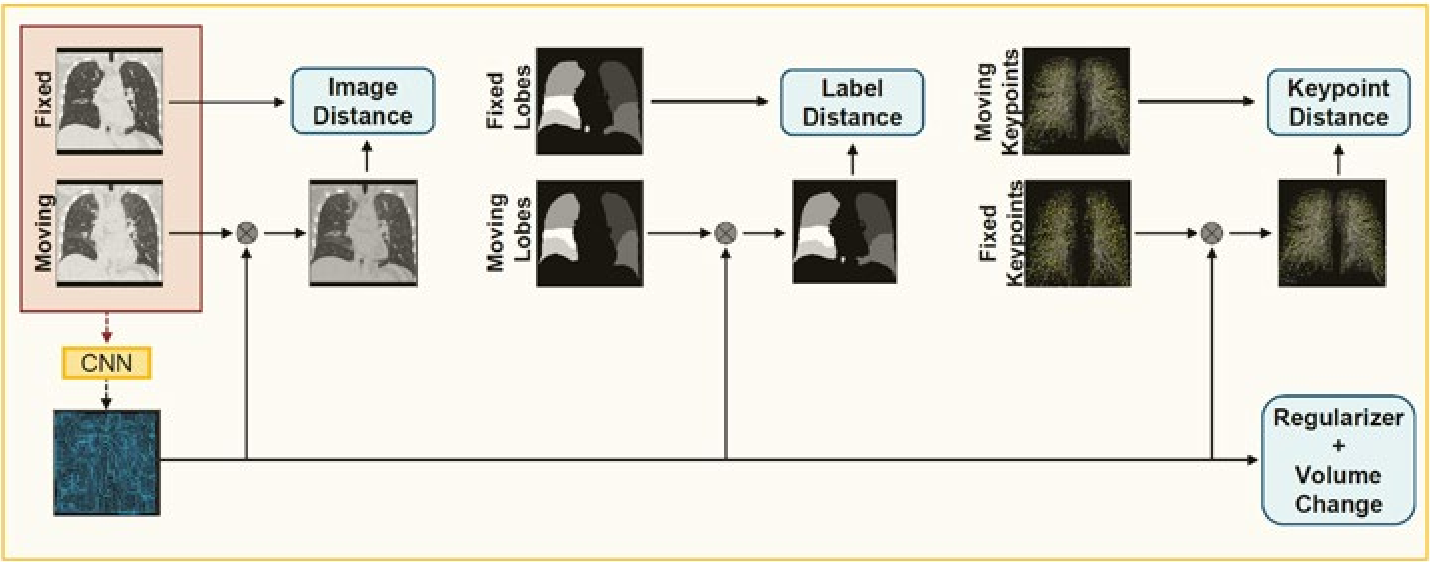
Alessa Hering has recently completed her PhD at RadboudUMC Nijmegen. Her research focused on the development of deep-learning-based image registration methods and tumor follow-up analysis (link). During her PhD, she mainly worked at Fraunhofer MEVIS. Currently, she pursues her career as a PostDoc at RadboudUMC but she still holds a position at Fraunhofer MEVIS. Congrats, Doctor Alessa!

Deep-Learning-Based Image Registration:

Medical image registration aims to align the anatomical structures of two images by establishing spatial correspondences. This is an important step for many tasks in medical image analysis as it links previously unrelated data and enables joint processing of these data. While deep learning has become a methodology of choice in many areas like segmentation and classification, image registration is often still based on conventional methods. This is because we cannot ask a medical expert to annotate a reference deformation field that establishes such correspondences for every voxel of the images.

In our research, we overcome this issue by using several loss functions that integrate prior anatomical knowledge as visualized in Figure 1. Furthermore, we developed a multi-level approach for deep-learning-based approaches to better handle deformations on different scales. The experimental results show very good performance on several publicly available datasets.

And as one opponent correctly pointed out during my defense: you cannot seem to finish your PhD in our group without organizing a challenge - so I co-organized the **Learn2Reg registration challenge** mainly together with Mattias Heinrich and Lasse Hansen to compare deep-learning-based to conventional registration methods on several registration tasks.



Efficient Tumor Follow-Up Assessment

Measurement of metastatic tumors on longitudinal CT scans is essential to evaluate the efficiency of cancer treatment. With yearly increases in the number of imaging studies conducted and the rising global cancer burden, the workload for radiologists keeps increasing. Therefore, there is an urgent need for systems that can speed up reporting and make structured reporting easier.

We developed a pipeline that automates the segmentation and measurement of lesions given a point annotation inside the lesion on the baseline scan. Our approach segments the baseline lesion and then uses an image registration method to find the corresponding location in the follow-up image. There, the lesion is also segmented using a CNN. Thereby, the follow-up measurements can be processed fully automatically. In the first step, we evaluated our algorithm with typical metrics like the Dice Score. Even though such metrics give us a rough indication of the performance, it does not tell us how our AI support can impact the follow-up assessment performed by radiologists. Therefore, we conducted a reader study with the goal of comparing a manual workflow with an AI-assisted workflow. In this study, we were able to verify the following three hypotheses:

- Assessment time is reduced using the AI-assisted workflow

- The inter-reader variability of the segmentation is reduced with AI assistance

- AI-assisted segmentation is as good as a manual segmentation

I find this research particularly exciting to work on because we work together with radiologists on a specific clinical application.

Roza is wrapping up her PhD studies this Fall. She studies Computer Science in the NEURDY Lab at Vanderbilt University. Her research focuses on developing computational models for advancing individual precision in brain mapping.

Her research has gained recognition through various awards. During her PhD, she also served on committees that advance women in STEM and Open Science.

She is now keen to find a job where she can continue 'making a difference' through her research in industry. And she's a great catch!

Developing Computational Models to Enable Individual Precision

Modern brain imaging techniques provide non-invasive windows into the anatomy and function of the brain. Since its inception, magnetic resonance imaging (MRI) has become a workhorse for investigating the human brain in health and disease. The primary focus of MRI has been to study similarities across subjects and to make inferences about groups of individuals. However, one movement in brain research is the study of individual precision. In my work, I leverage computational methods to model and interpret large-scale and high dimensional neuroimaging data at the level of individuals.

Mechanisms of Functional MRI Signals

Brain-body interactions underpin key functions including cognition and emotion, as well as the overall health of an organism. Natural fluctuations, such as breathing and heart rate, provide windows into these critical functions. Yet, many studies of the human brain using fMRI lack these physiological measurements. I developed and tested frameworks [1,2] to reconstruct respiratory variation (RV) and heart rate (HR) directly from whole-brain fMRI dynamics. To model our physiological signals of interest, I adapted an LSTM network to jointly learn these signals Fig.1 [3]. To achieve generalizability, we utilized large-scale datasets for training; and tested our models on external datasets. The models successfully infer both respiration and heart rate fluctuations; are transferable to scans of variable lengths and different experimental conditions.

During the process of investigating our physiological models, we also asked if individuals have unique physiological signal patterns? To assess this, we conducted a series of experiments to relate individual physiological maps to cognitive phenotypes and behavioral traits. We found that physiological variance in fMRI signals correspond to unique properties of individual subjects and are predictive of individual-specific traits.

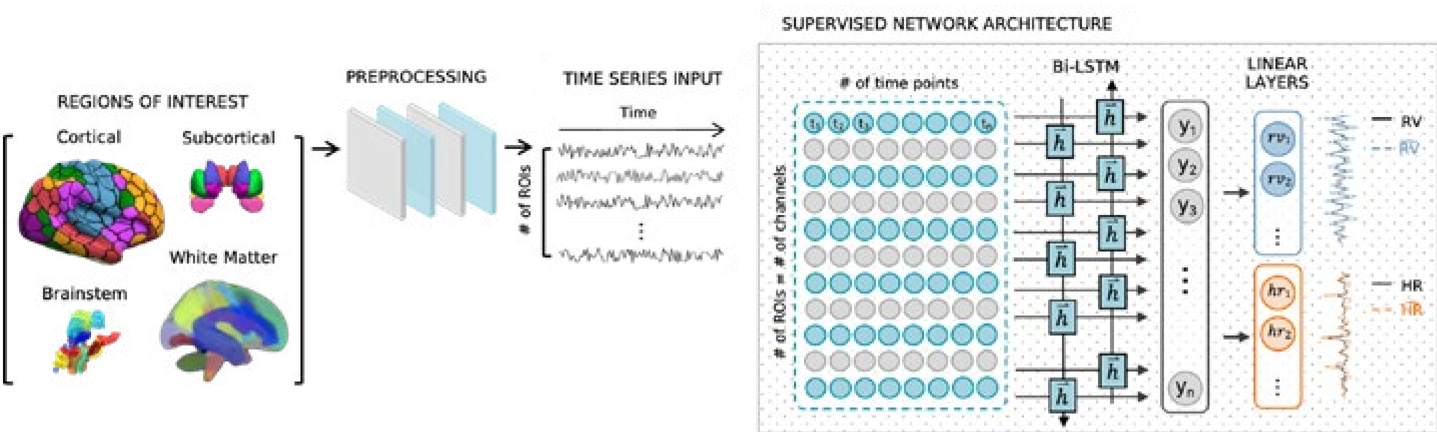


Figure 1: Framework for extracting respiration variation (RV) and heart rate (HR) signals directly from functional MRI data.

Functional Parcellations

My thesis also consists of a number of studies that investigate individualized brain mapping using different computational perspectives. In one of the projects, I developed an interactive visual analytics tool to explore and construct individual-specific, functionally-meaningful regions from group level atlases [4]. PRAGMA takes an existing reliable prior and allows experts to build functional parcellations that capture individual differences (Fig. 2).

It has been shown that the parcellations tailored to individual subjects provide useful information about brain function/dysfunction. Methods to derive these *functional* parcels commonly rely on fMRI scans that are not routine in clinical practice. In a recent work [5], I asked whether tissue properties (cortical thickness, curvature etc.) that are measurable from more routine clinical scans can tell us about functional regions in the brain. Specifically, I developed a framework to infer individualized functional parcellation from structural MRI measures. The proposed method achieved comparable parcellation accuracy against fMRI derived ground truth labels. In addition, our individual-level parcellations improved homogeneity (similarity of time-courses within parcel) over widely used group parcellations.

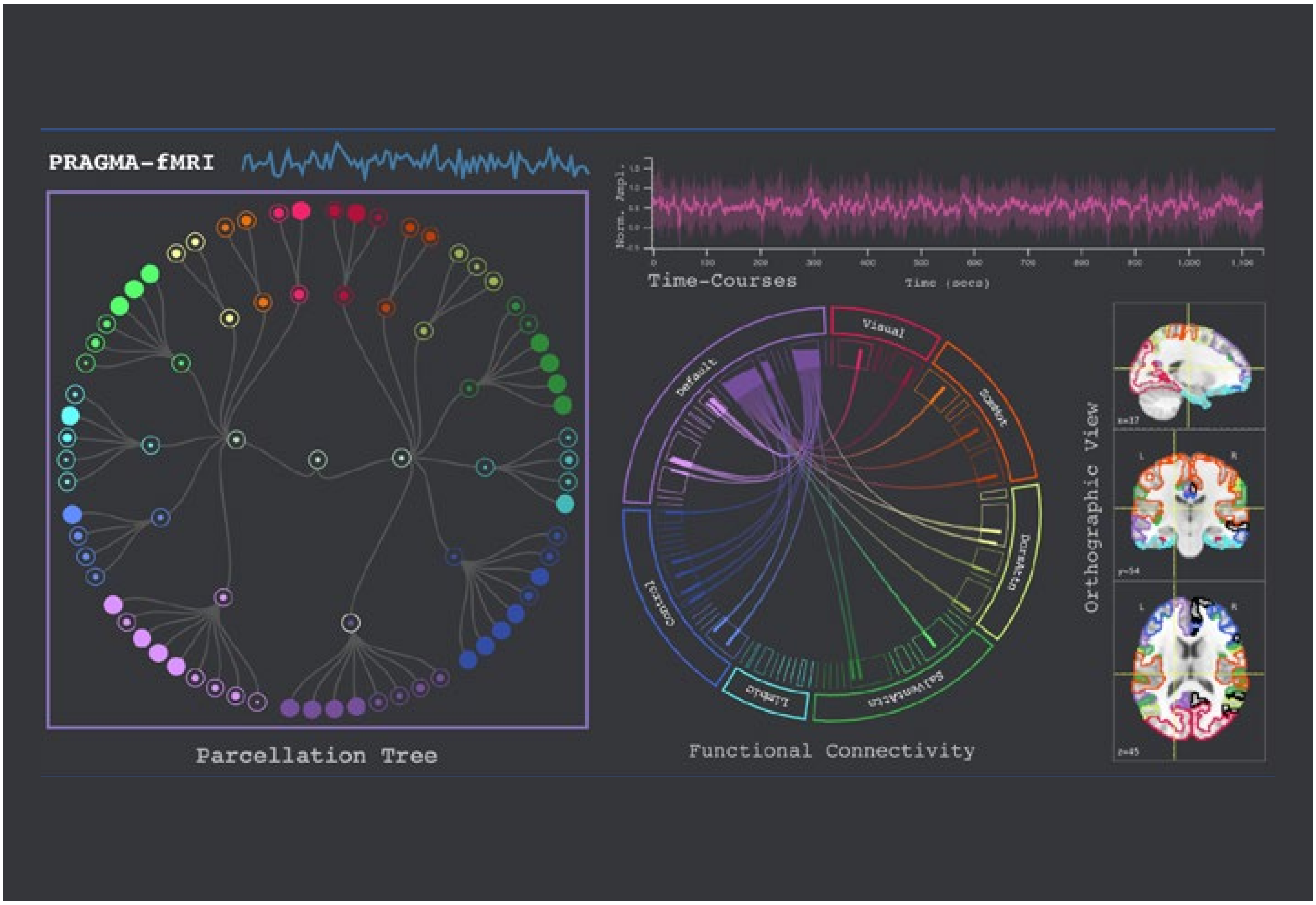


Figure 2: PRAGMA interface for interactively constructing functional brain parcellations.

GeoMedIA Workshop 2022

Geometric Deep Learning in Medical Image Analysis
18 November 2022, Amsterdam

First speaker will be awesome Emma Robinson from King's College London!

MICCAI Endorsed Event

ellis unit | AMSTERDAM

NWO

METRICS RELOADED

Paul Jäger

Minu Tizabi

Paul Jäger is the Principal Investigator of the Interactive Machine Learning Research Group at the German Cancer Research Center (DKFZ) in Heidelberg, where Minu Tizabi is a Scientist in Lena Maier-Hein's Intelligent Medical Systems Research Group. They speak to us about their follow-up work to Metrics Reloaded, a recommendation framework for biomedical image analysis validation.

Validation is critical to automated AI-based biomedical image analysis. However, while research focuses on developing new models and algorithms, less attention is paid to **whether these algorithms are being appropriately validated**. If an algorithm is not validated according to the relevant outcome measures, it is not possible to make a reliable statement about whether it has the potential to be used in a practical setting in the future or if, for all intents and purposes, it just looks good on paper.

"Our group found that validation is often performed in a way that's not entirely correct," Minu tells us. *"A big part of this is the inappropriate use of validation metrics, the measures against which algorithm performance is assessed. People wrongly apply or use metrics that aren't mathematically suited to the underlying problem. That was the starting point for us to investigate. How can we help people choose the right metrics?"*

The group found that rankings used in popular community challenges are sometimes computed in ways that are not ideal – for example, using two metrics with different names but the same mathematical formula while believing that they are two separate metrics with two separate statements.

They also suggest that the computer vision community has developed certain practices that may be disconnected from application over the years. There are many great algorithms and research successes, but not so many end up in a clinical setting. Part of the reason for this gap might be that algorithms are not directly evaluated for clinical translation.

"We don't have specific examples of where it's gone wrong in practice because it's still in the research bubble," Paul points out. *"We don't get to see many models being tried and failing, so we can't provide this empirical evidence that it failed because there was a wrong validation. More abstractly, there are **two big bottlenecks provoked by inappropriate metrics**. One is the general translation into clinics because the numbers just don't add up. Everyone does segmentation, but we talk to clinicians who don't want it. Also, research progress itself is impeded. If competitions use inappropriate metrics, then winning models are selected that aren't the best for the job. That could spark research in the wrong direction. If we can't select the best models, we can't use them again as baselines to develop new methods. The whole scientific progress is affected."*

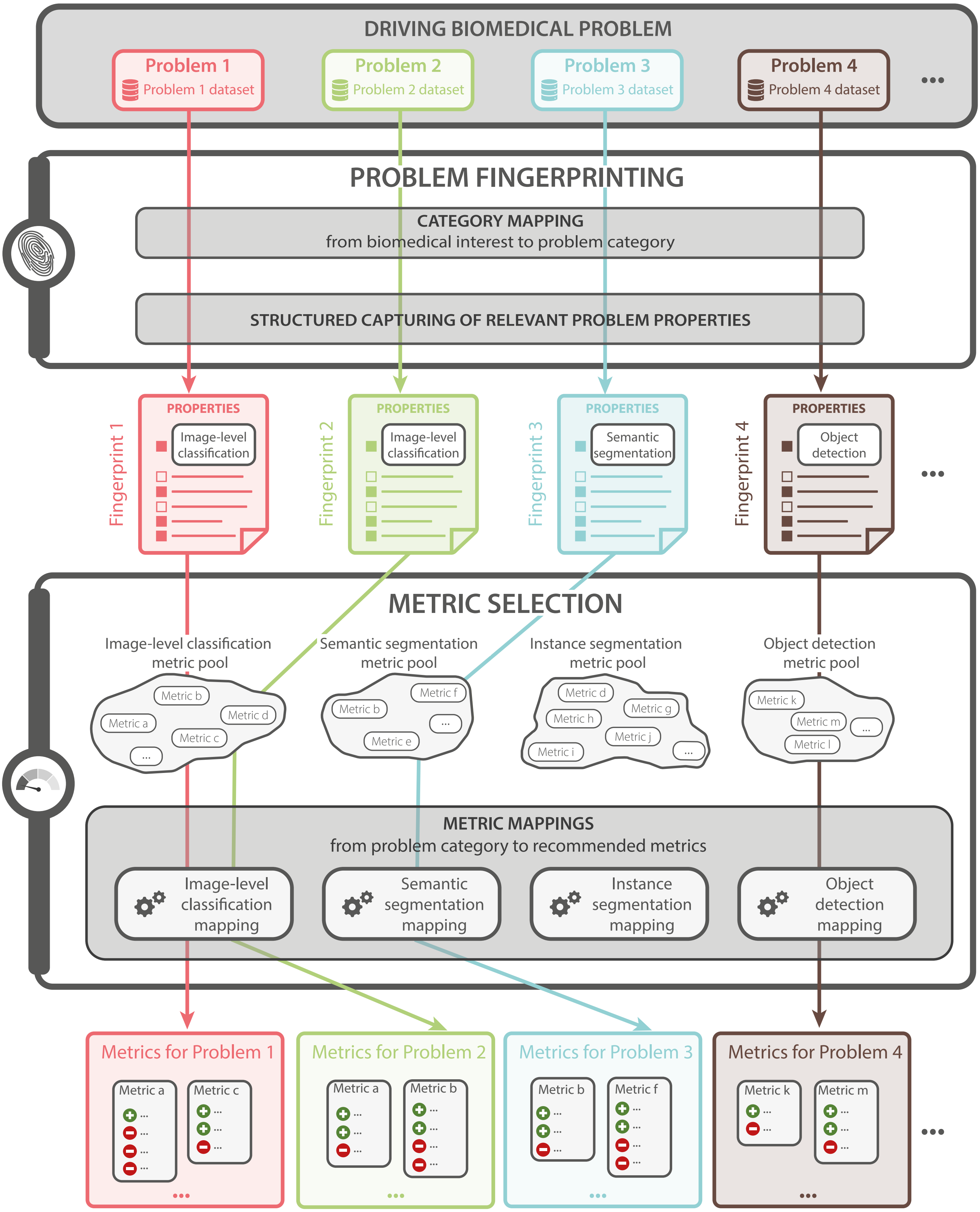
Lena and **Annika Reinke** did some analysis and found that 70% of competitions rely on semantic segmentation, where each pixel is classified individually.

"From a clinician's perspective, when they diagnose something in the clinic, they don't base their diagnosis on pixel values," Paul continues. *"They would not say to a patient – hey, you have a cancerous pixel here! At **MICCAI**, 80% of papers are new **U-Net variants**, and they all validate with **Dice score segmentation metrics**. But most often, the clinical need is not on the pixel scale, and rather decisions are made on objects or entire structures."*

Work has been ongoing in this direction for several years. Annika Reinke and a group of eminent co-signers impressed with their position paper at MICCAI 2018, showing how "**security holes**" related to challenge design and organization could be used to manipulate rankings. They proposed best practice recommendations to remove opportunities for cheating.

Once the problem was identified, a large international consortium of image analysis experts came together. Using a **multi-stage Delphi process**, with questionnaires and expert group meetings for consensus building, they began to figure out the best approach for fostering the convergence of validation methodology and changing common practice. Metrics Reloaded, a comprehensive framework guiding researchers toward choosing metrics in a problem-aware manner, grew out of all this work.

"We want to make sure that people who use our recommendation framework find out the best metrics for them," Minu states. *"They should be made aware of the pitfalls and educated transparently about what they're doing, rather than just putting something in and getting something out of a black box and not understanding why these metrics would be adequate. It takes time to change common practice and*



Choosing metrics for a biomedical problem is achieved in two main steps. First, the \textbf{problem fingerprint} is generated, which captures the characteristics of the driving biomedical problem relevant for metric selection. During this step, the category mapping assigns the given problem to the appropriate image processing category, namely image-level classification, semantic segmentation, object detection and instance segmentation. Finally, the \textbf{metric selection} step generates a set of suitable metrics from a category-specific pool of metrics in a domain interest-aware manner.

convince people that they may have been doing something wrong."

Ultimately, the team wants to be able to map every personal problem to a set of recommended metrics. Their rules need to catch every use case. They are looking to tackle all scales of classification at once, including **semantic segmentation, instance segmentation, object detection, and image-level classification**, and aim to be agnostic to application domains, incorporating **histopathologic images, radiological images, and microscopic images**. They see similarities and synergies between them all.

One can imagine that telling people where they have been going wrong is not without its own pitfalls. Did the team face any opposition?

"Yes, constantly!" Paul reveals. *"I'm currently part of parallel email discussions fighting about the best metric for certain use cases. It's not always clear. We could have made it easier for ourselves, but we defined a very broad scope and ambitious setting. It makes sense to address everything at once, but that also takes more time and effort."*

Although Metrics Reloaded occupies a large part of their time, Paul and Minu have other roles. Minu is a physician by background, having studied medicine, but is now working at the intersection of her department's technical and medical sides.

*"Metrics is my main occupation timewise, but in our department, we're trying to create **long-lasting structural efforts in surgical data science**, and I'm heavily involved in that too,"* she tells us.

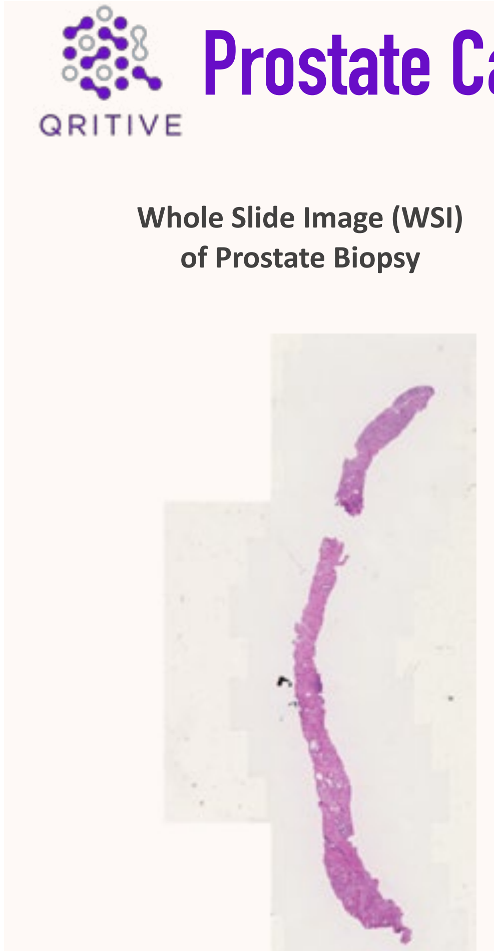
Paul adds: *"Now that I'm an independent research group leader, I supervise four PhD students and two master's students in my group. We're pursuing work on everything to do with human-machine interaction, including **active learning approaches, explainable AI, and detecting failures in systems**. When I'm not supervising my group, my hands-on work is mostly the Metrics Reloaded project."*

The team plans to submit two papers to **Nature Methods** shortly. One is about understanding and analyzing the pitfalls of metrics. The second is the recommendation framework, where users are guided through the process of selecting and applying appropriate validation metrics while being made aware of the potential pitfalls. It includes a user-friendly online tool, providing a point of access to explore the weaknesses and strengths of the most common metrics.

Project coordinator Lena Maier-Hein believes the work to be essential to understanding the limited translation of machine learning methods into clinical practice.

"This work is just the start," Paul teases. *"This online framework maps your problem to recommended metrics, but there's room for much more. It could be the central point for education and transparency about metrics and their limitations. If you're interested in a specific metric, we have a library where you can click on it and get all the information, see the connections, and generate tables with the correlations between metrics. I imagine **interactive educational tools where you can create simulations with data sets to shift around and see how the metrics behave**. We want to be the central hub for metric understanding in the community. That is the long-term vision."*

QRITIVE



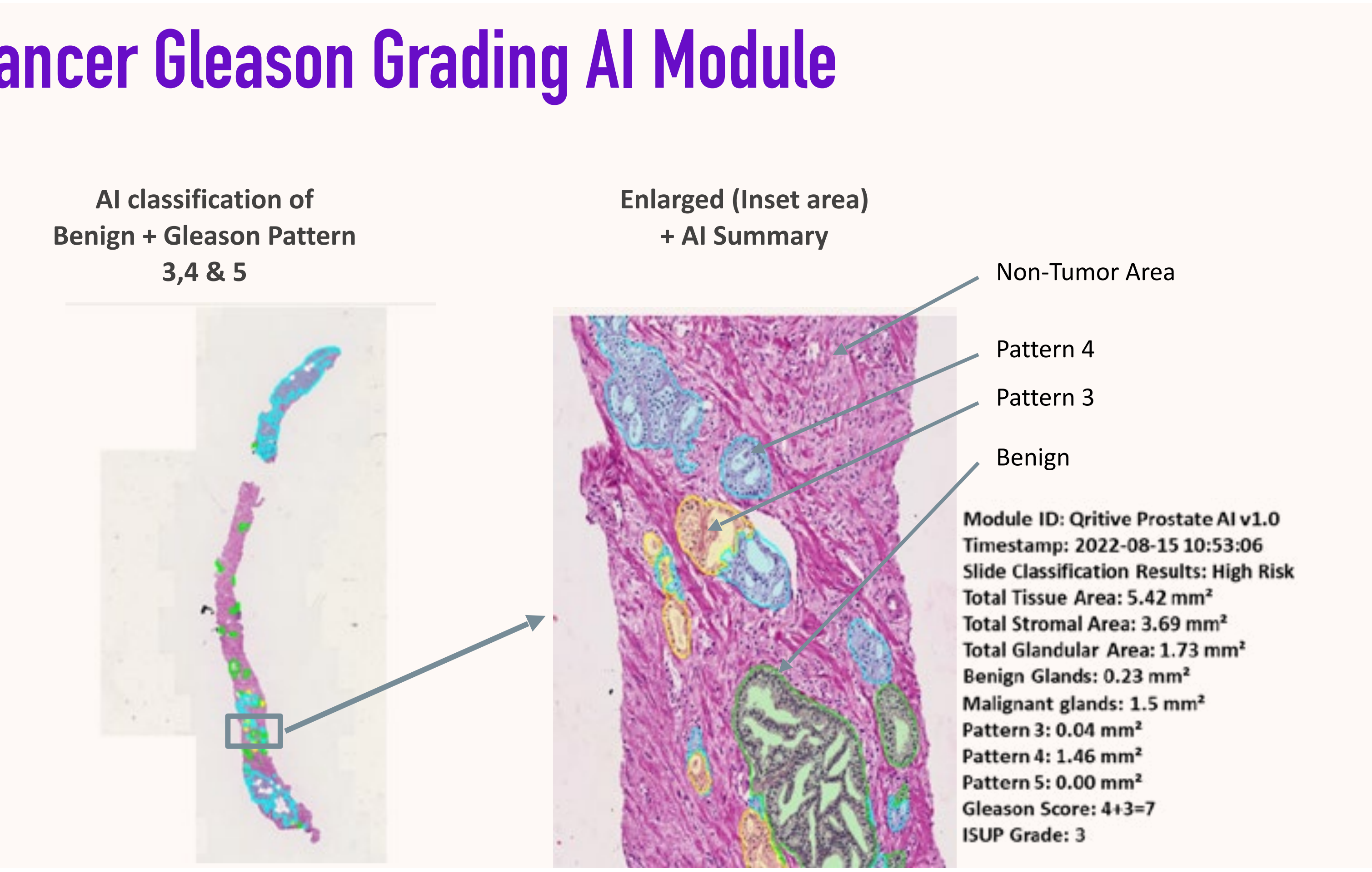
Bruno Occhipinti is CCO at Qritive, a Singapore-based health start-up harnessing the power of AI to improve cancer diagnosis. He speaks to us about Qritive's work and its plans for the future.

Qritive is on a mission to solve one of the most important problems in histopathology: inefficiencies and subjectivities in cancer diagnosis. It aims to make diagnosis fast, accurate, and affordable through the use of promising AI technology.

"We've built the infrastructure required for addressing the entire AI development and deployment lifecycle," Bruno tells us.

"Building AI modules might actually be relatively simple, but ***many challenges prevent AI modules from being deployed in a clinical setting****. From the moment we receive the data, we get expert knowledge into it, train our AI modules, and then deploy it in a way consistent with what's expected in a hospital environment."*

Qritive has a system called **AI Engine**, allowing its scientists to build their own AI modules, and run experiments



in a scalable Kubernetes cluster. They efficiently build these algorithms, test them, and take them through internal validation. The same system allows Qritive to **automatically deploy the modules for clinical validation to see if they work in a real-world environment**.

"When building algorithms, training AI models and running experiments, companies and academia face ***issues with scaling up***,*"* Bruno points out.

"With AI engine, we can run hundreds or even thousands of experiments with GPUs attached to the Kubernetes cluster and the whole system gets handled automatically. It even takes care of all the complex MLOps required to deploy this for clinical validation or clinical use."

Another issue is that **histopathology has many proprietary file formats, scanners, and image file types**, so it can be challenging for people to work with them all. However, AI Engine supports almost all the major slide scanner manufacturers.

Qritive has put a great deal of engineering into its platform over the past five years to build a solid AI infrastructure for histopathology.

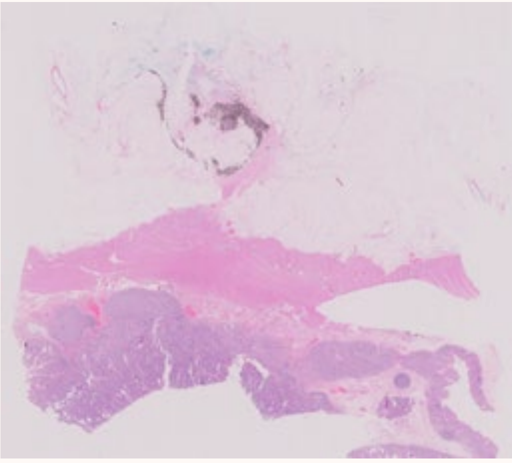
QRITIVE

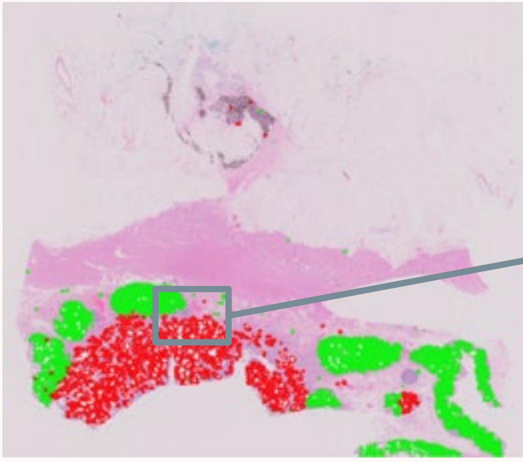
Colon Cancer Grading AI Module

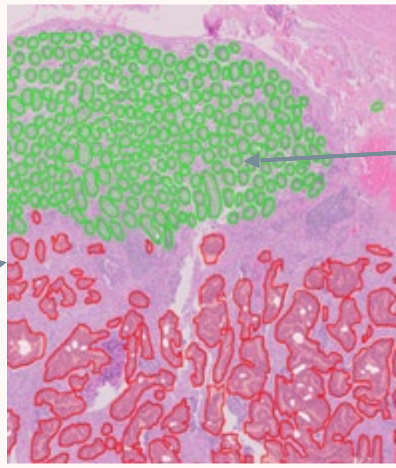
Whole Slide Image (WSI)
of Descending Colon Biopsy

AI classification of
Benign + Malignant areas

Enlarged (Inset are
AI summary







Module ID: Qritive Colon
Timestamp: 2020-10-15 0
Slide Classification Result

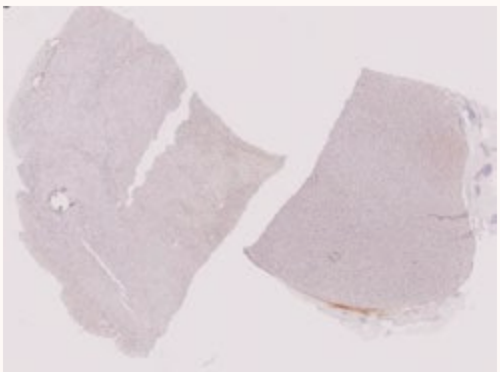
QRITIVE

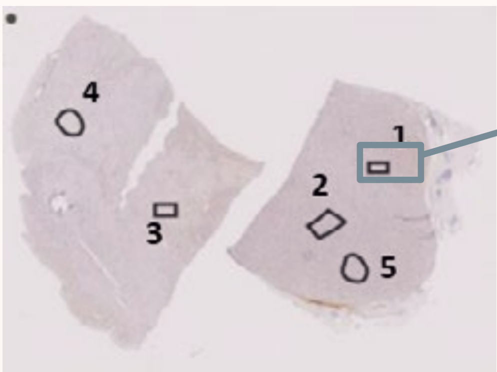
Sarcoma Cancer Ki-67 Classification and Co

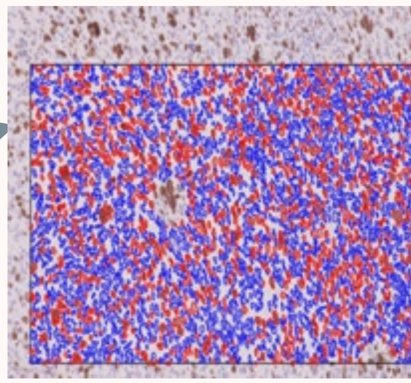
Whole Slide Image (WSI)
of right thigh mass resection
for Sarcoma (IHC)

Identification of ROI areas (1-5)
for AI classification for Ki-67
Positive and Negative counts

Enlarged (Inset area)
ROI 1 +AI summary

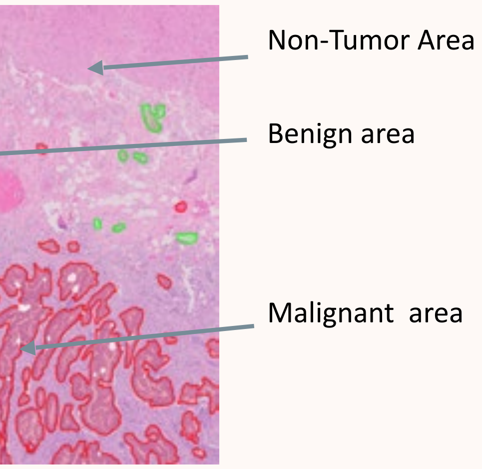






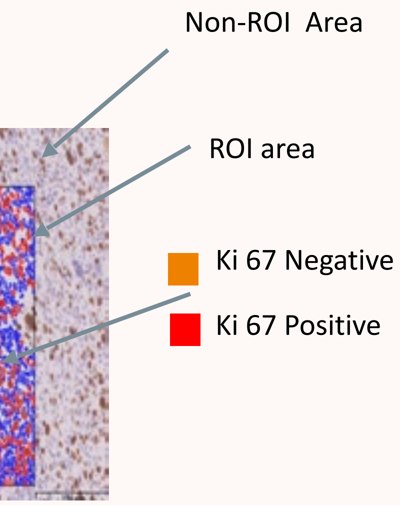
Module ID: Qritive IHC AI v1.0
Timestamp: 2022-07-29 08:27:20
Total Tissue area: 182.22 mm²
Total ROI area: 3.85 mm²
Slide Classification Results: Ki 67 positive

a) &



AI v1.0
9:07:44
: High Risk

unts AI Module



"Histopathological images tend to be huge, and to be able to process all this information with an AI module, specific engineering is needed," Bruno explains.

"If you open up one of these images, it looks like Google Maps. You can zoom in to the maximum resolution and look at the cell morphology and very subtle details that normal or tumor cells might show. If you zoom out, you'll see how huge the image is. To process this sort of image, you must address it at many different magnifications and levels. You start with very low magnification to get a broad sense of this, then go deeper and deeper, and aggregate all this information into insights the system generates."

The use of AI in healthcare is still in its early stages. Its adoption has been slow compared to other industries, despite increasing interest from stakeholders. Why does Bruno think this is?

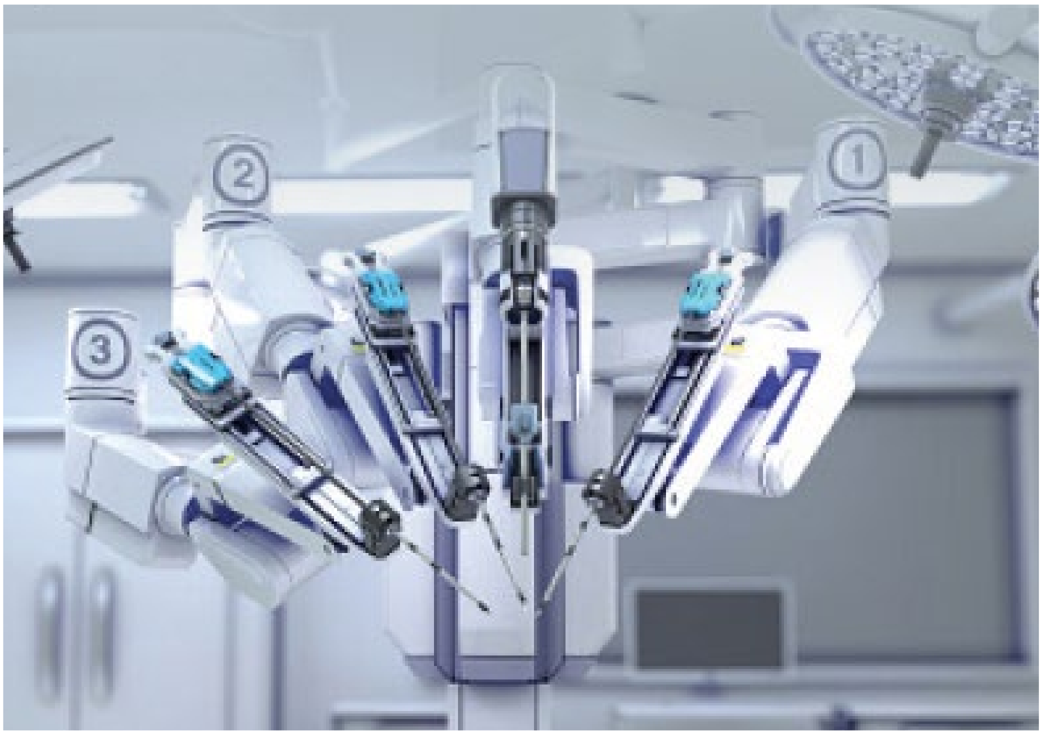
"AI in histopathology is a given. It's coming. There's no question about it," he responds.

*"One of the reasons it's been slow is because of a lack of infrastructure. At Qritive, **we're building all the important pieces of this infrastructure required to bring AI to histopathology**. We're pretty sure we'll be successful, but it's about the market, execution, and many other factors. We're doing our best to deliver it."*

Qritive has already built an extensive network of interest, but with global ambitions, it is always looking for diverse partners and collaborators worldwide.

"Other companies like us are mostly in the US or Europe, whereas Qritive is in Asia," Bruno points out.

"In Asia, access to a pathologist is a massive challenge. You don't have the skill sets. We want to use AI to unlock access to diagnostics. In a world where you see so many cases of cancer, the fact we can open access to diagnostics could impact healthcare dramatically, and we're on the right side of the world to do it."



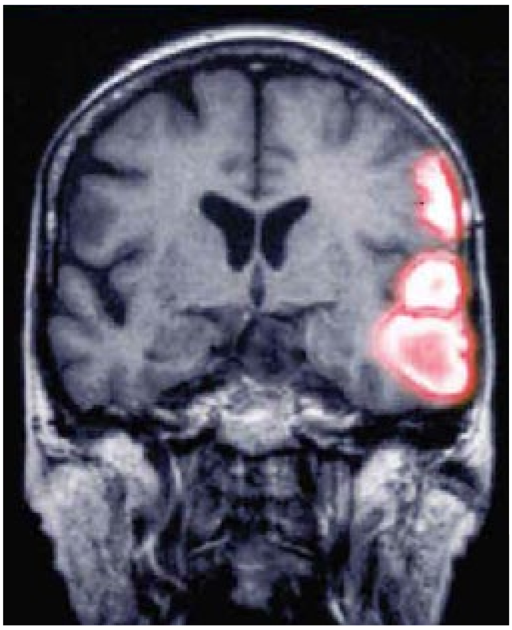
IMPROVE YOUR VISION WITH

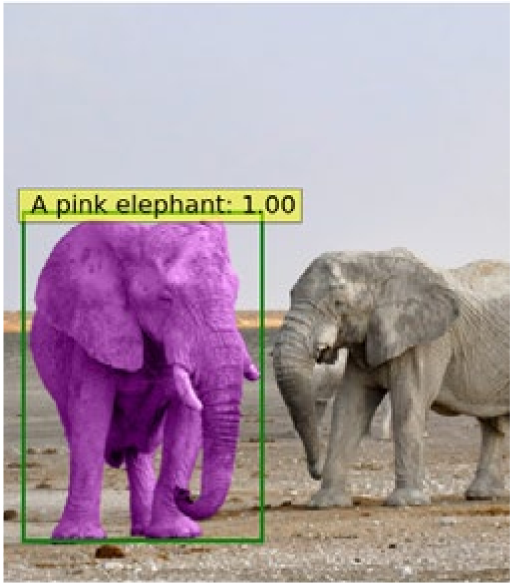
Computer Vision News

SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!





RSIP VISION