November 2023

Computer Vision News

& Medical Imaging News

The Magazine of the Algorithm Community

Yann LeCun
Exclusive Interview

A publication by
RSIP VISION

Yann LeCun was keynote speaker at MICCAI 2023. He was so kind as to give a second interview to Ralph, during his visit at ICCV 2023 in Paris.

Yann, thank you very much for being with us again. When we talked five years ago, you told me you had a clear plan for the next few years. Did you stick to it?

The plan hasn't changed very much – the details have changed, and we've made progress, but the original plan is still the same. The original plan was the limitation of current AI systems is that they're not capable of understanding the world. You need a system that can understand the world if you want it to be able to plan. You need to imagine in your head what the consequences of your actions might be, and for this, you need a world model. I've been advocating for this for a long time. This is not a new idea. The concept is very old, from optimal control, but using machine learning to learn the world models is the big problem.

Back when we talked, I can't remember if I'd made the transition between what I called latent variable generative models and what I'm advocating now, which I call JEPA, so joint embedding predictive architectures. I used to think that the proper way to do this would be to train a system on videos to predict what will happen in the video, perhaps as a consequence of some action being taken. If you have

a system that can predict what's going to happen in the video, then you can use that system for planning. I've been playing with this idea for almost 10 years. We started working on video prediction at FAIR in 2014/15. We had some papers on this. Then, we weren't moving very fast. We had Mikael Henaff and Alfredo Canziani working on a model of this type that could help plan a trajectory for self-driving cars, which was somewhat successful.

"This is, I think, the future of AI systems. Computer vision has a very important role to play there!"

But then, we made progress. We realized that predicting everything in a video was not just useless but probably impossible and even hurtful. I came up with this new idea derived from experimental results. The results are such that if you want to use self-supervised learning from images to train a system to run good representations of images, the generative methods don't work. The methods are based on essentially corrupting an image and then training a neural network to recover the original image. Large language models are trained this way. You take a text, corrupt it, and then train a system to reconstruct it. When you do this with images, it doesn't work very well. There are a number of techniques to do this, but they don't work very well. The most successful is probably MAE, which means masked autoencoder. Some of my colleagues at Meta did that.

What really works are those joint embedding architectures. You take an image and a corrupted version of the image, run them through encoders, and train the encoders to produce identical representations for those two images so that the representation produced from the corrupted image is identical to that from the uncorrupted image. In the case of a video, you take a segment of video and the following segment, you run them through encoders, and you want to predict the representation of the following segment from the representation of the previous segment. It's no longer a generative model because you're not predicting all the missing pixels; you're predicting a representation of them. The trick is, how do you train something like this while preventing it from collapsing? It's easy for this system to collapse, ignore the input, and always predict the same thing. That's the question.

So, we did not get to solve the exact problem we wanted?

It was the wrong problem to solve. The real problem is to learn how the world works from video. The original approach was a generative model that predicts the next video frames. We couldn't get this to work. Then, we discovered a bunch of methods that allow one of those joint embedding systems to learn when they're collapsing. There are a number

of those methods. There's one called BYOL from DeepMind – Bootstrap Your Own Latent. There are things like MoCo. There have been a number of contrastive methods to do this. I probably had the first paper on this in 1993, on a Siamese neural network. You train two identical neural nets to produce identical representations for things you know are semantically identical and then push away the outputs for dissimilar things. More recently, there's been some progress with the SimCLR paper from Google.

Then, I became somewhat negative about those contrastive methods because I don't think they scale very well. A number of non-contrastive methods appeared about four years ago. One of them is BYOL. Another one, which came from my group at FAIR, is called Barlow Twins, and there are a number of others. Then, we came up with two other ones called VICReg and I-JEPA, or Image JEPA. Another group at FAIR worked on something called DINOv2, which works amazingly well. Those are all different ways of training a joint embedding architecture with two parallel networks and predicting the representation of one from the representation of the other. DINOv2 is applied to images, VICReg is applied to images and short videos, I-JEPA to images, and now we're working on something called V-JEPA or Video JEPA, a version of this for video. We've made a lot of progress. I'm very optimistic about where we're going.

You have long been a partisan of the double affiliation model. Would you suggest young people today consider a career with hats in academia and industry, or would your advice for this generation be a little bit different?

I wouldn't advise young people at the beginning of their career to wear two hats of this type because you have to focus on one thing. In North America, if you go into academia, you have to focus on getting tenure. In Europe, it's different, but you have to focus on building your group, your publications, your students, your brand, your research project. You can't do this if you split your time.

Yann's interview with Ralph in 2018

Once you're more senior, then it's a different thing. Frankly, it's only in the last 10 years that I've been straddling the fence in a situation where I'm pretty senior and can choose what I want to work on.

At FAIR, we don't take part-time researchers who are also faculty if they're not tenured. Even the tenured, we tend only to take people who are quite senior, well established, and sometimes only for a short time, for a few years or something like that. It's not for everyone. It depends on which way you want to have an impact and whether you like working with students. In industry, you tend to be more hands-on, whereas in a university, you work through students generally. There are pluses and minuses.

You are one of the well-known scientists in our community who does not shy away from talking to younger and less experienced people on social media, in articles, and at venues like ICCV and MICCAI. Do you also learn from these exchanges?

The main reason for doing it is to inspire young people to work on interesting things. I've been here at ICCV for about an hour and a half, and about 100 people came to take selfies with me. I don't turn them down because they're so enthusiastic. I don't want to disappoint them. I think we should encourage enthusiasm for science and technology from young people. I find that adorable. I want to encourage it. I want to inspire people to work on technology that will improve the human condition and make progress in knowledge. That's my goal. It's very indirect. Sometimes, those people get inspired. Sometimes, that puts them on a good trajectory. That's why I don't shy away.

There are a lot of exchanges about the potential benefits and risks of AI, for example. The discussions I've had on social media about this have allowed me to think about things I didn't think of spontaneously and answer questions I didn't know people were asking themselves. It makes my argument better to have these discussions on social media and have them in public as well. I've held public debates about the risks of AI with various people, including Yoshua Bengio and people like that.

I think it's useful. Those are the discussions we need to have between well-meaning, serious people. The problem with social media is that there's a lot of noise and people who don't know anything. I don't think we should blame people for not knowing; I think we should blame people for being dishonest, not for not knowing things. I'm a professor. My job is to educate people. I'm not going to blame them for not knowing something!

You started in a place where you

knew every single scientist in your field. Now, you are meeting thousands and cannot learn all their names. What is your message to our growing community?

A number of different messages. The first one is there are a lot of applications of current technologies where you need to tweak an existing technique and apply it to an important problem. There's a lot of that. Many people who attend these conferences are looking for ideas for applications they're interested in medicine, environmental protection, manufacturing, transportation, etc. That's one category of people – essentially AI engineers. Then, some people are looking for new methods because we need to invent new methods to solve new problems.

Here's a long-term question. The success we've seen in natural language manipulation and large language models – not just generation but also understanding – is entirely due to progress in self-supervised learning. You train some giant transformer to fill in the blanks missing from a text. The special case is if the blank is just the last word. That's how you get autoregressive LLMs. Self-supervised learning has been a complete revolution in NLP. We've not seen this revolution in vision yet. A lot of people are using self-supervised learning. A lot of people are experimenting with it. A lot of people are applying it to problems where there's not that much data, so you need to pre-train on whatever data you have available or synthetic data and then fine-tune on whatever data you have.

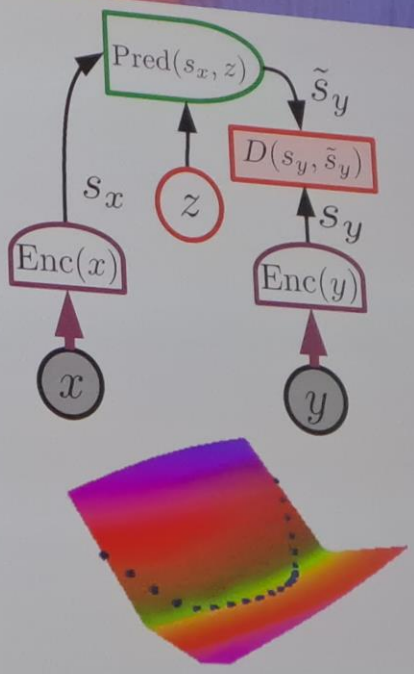
So, some progress in imaging. I'm really happy about this because I think that's a good thing, but the successful methods aren't generative. The kind of methods that work in these cases aren't the same kind of methods that work in NLP. In my opinion, the idea that you're going to tokenize your video or learn to predict the tokens is not going anywhere. We have to develop specific techniques for images because images and video are considerably more complicated than language. Language is discrete. It makes it simple, particularly when having to handle uncertainty. Vision is very challenging.

We've made progress. We have good techniques now that do self-supervised learning from images. The next step is video. Once we figure out a recipe to train a system to learn good representations of the world from video, we can also train it to learn predictive world models: Here's the state of the world at time T. Here's an action I'm taking. What's going to be the state of the world at time T+1? If we have that, we can have machines that can plan, which means they can reason and figure out a sequence of actions to arrive at a goal. I call this objective-driven AI. This is, I think, the future of AI systems. Computer vision has a very important role to play there. That's what I'm working on. My entire research is entirely focused on this!

Recommendations:

Y. LeCun

- **Abandon generative models**
  - in favor joint-embedding architectures
- **Abandon probabilistic model**
  - in favor of energy-based models
- **Abandon contrastive methods**
  - in favor of regularized methods
- **Abandon Reinforcement Learning**
  - In favor of model-predictive control
- Use RL only when planning doesn't yield the predicted outcome, to adjust the world model or the critic.



Problems to Solve

Y. LeCun

- **JEPA with regularized latent variables**
  - Learning and planning in non-deterministic environments
- **Planning algorithms in the presence of uncertainty**
  - Gradient-based methods and combinatorial search methods
- **Learning Cost Modules (Inverse RL)**
  - Energy-based approach: give low cost to observed trajectories
- **Planning with inaccurate world models**
  - Preventing bad plans in uncertain parts of the space
- **Exploration to adjust world models**
  - Intrinsic objectives for curiosity

Inspiring (and surprising) slides from Yann LeCun's keynote speech at MICCAI.

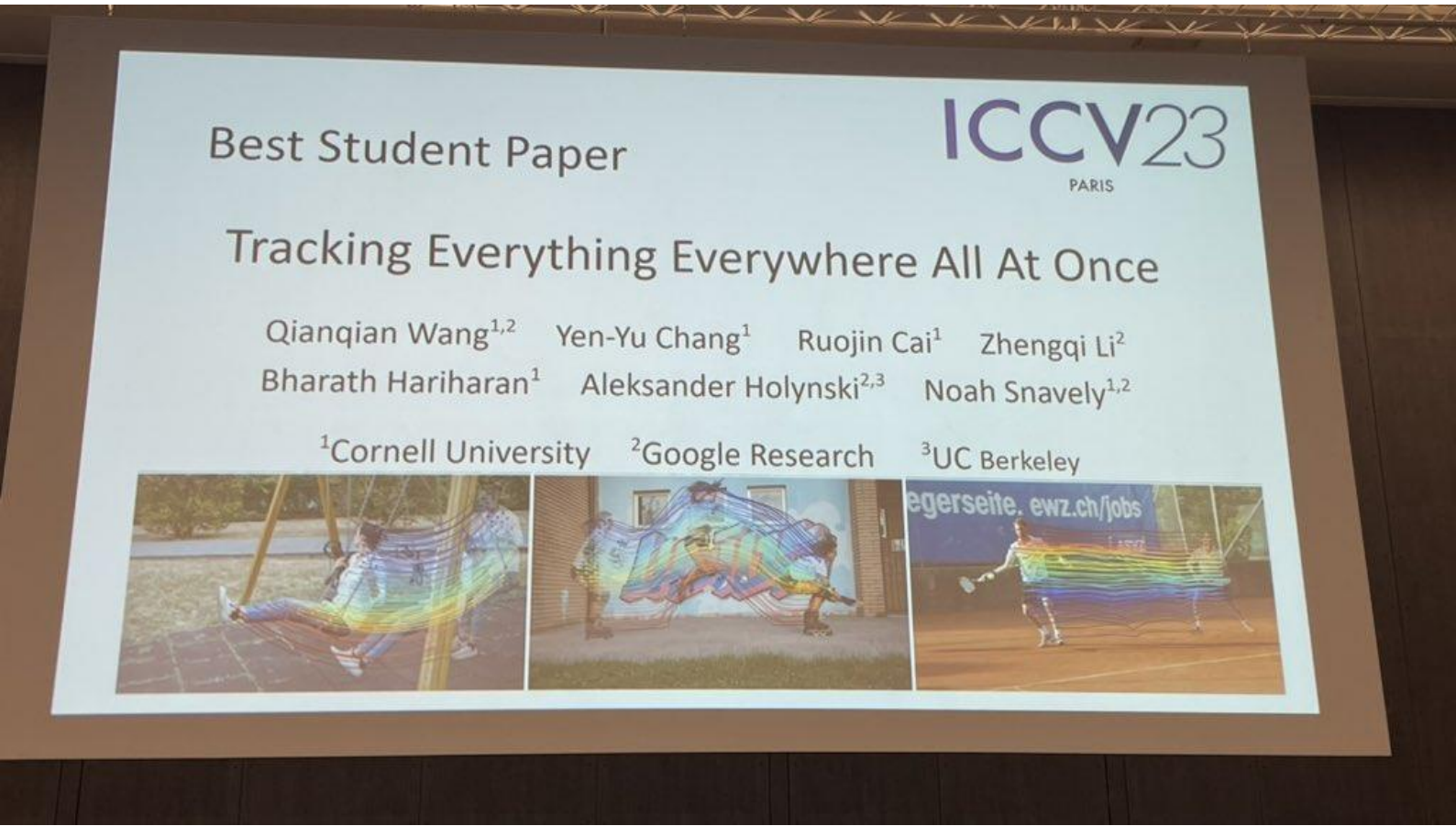
Tracking Everything Everywhere All At Once

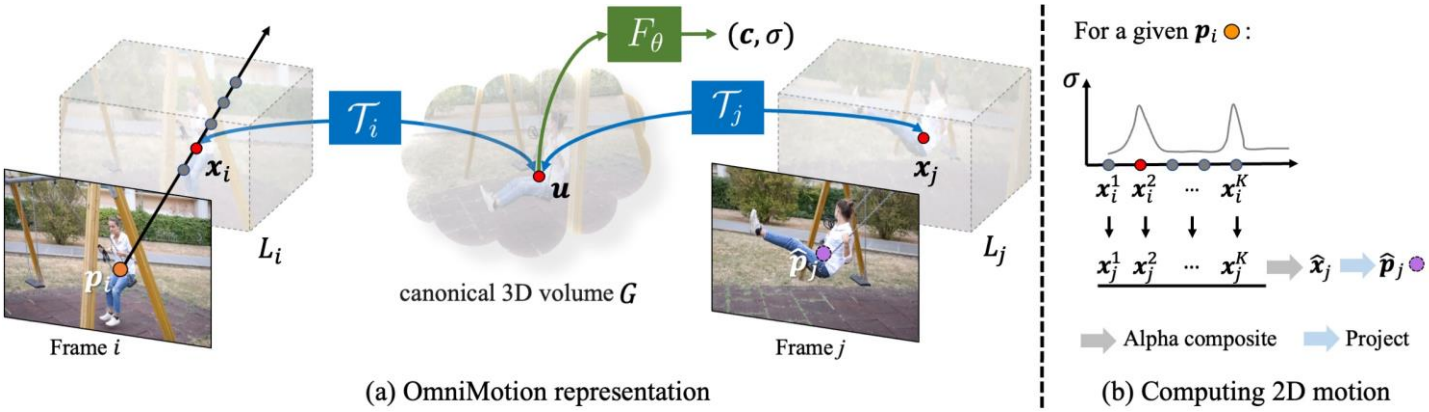
This exceptional work has just won the Best Student Paper Award at ICCV 2023. This interview was conducted before the announcement of the award. RSIP Vision continues a long tradition of selecting in advance the future award-winning papers for full feature! Congrats Qianqian!

Author Qianqian Wang is a postdoc at UC Berkeley. She recently completed her PhD in Computer Science at Cornell Tech.

She speaks to us about her work on estimating motion from video sequences ahead of her oral presentation and poster this afternoon. Read our full review of her winning work in the next pages!

In this paper, **Qianqian** proposes **a novel optimization method for estimating the complete motion of a video sequence**. It presents dense and





(a) OmniMotion representation (b) Computing 2D motion

and long-range motion representation that allows for **tracking through occlusions and modeling full-length trajectories**.

This method finds **correspondences between frames**, a fundamental problem in computer vision. These correspondences are the foundation for various applications, notably dynamic **3D scene reconstruction**, as understanding 2D correspondences between frames in a dynamic scene is essential for reconstructing its 3D geometry and 3D motion. The research also opens up exciting possibilities for video editing, allowing for seamless propagation of edits across multiple frames.

"I came up with this idea because, in my last project, I realized there was no such motion representation in the past," Qianqian tells us. *"It's not a new problem, but people don't have a good solution. The last paper I saw similar to our work was 10 years ago, but because it's too challenging, and people don't have new tools to work on it, progress has been suspended for a decade."*

Now, renewed interest in this problem has sparked concurrent research. While approaches may differ, the shared goal remains the same – **tracking points in a video**

over extended periods. However, the road to achieving that is not without its challenges.

"The first challenge was to formulate the problem because it's different from what most people did before," Qianqian explains. *"We have **sparse feature tracking**, which gives you long-range correspondences but they are sparse. On the other hand, we have **optical flow**, which gives you dense correspondences, but only for a very short period of time. What we want is **dense and long-range correspondences**. It took a little bit of time to figure that out."*

An important moment in the project was realizing the need for **invertible mapping**. Without it, the global consistency of estimated motion trajectories could not be guaranteed. It was then a challenge to determine how to represent the geometry. Parameterizing the quasi-3D space was far from straightforward,

which led to the team exploring the concept of **neural radiance field**, a dense representation offering the flexibility needed to optimize scene structure and the mapping between each local and canonical frame.

The work opens up opportunities for future extensions, including using similar principles for reconstructing dynamic scenes and enhancing video editing techniques with speed and efficiency improvements.

"*Compared to other correspondence work, our approach guarantees **cycle consistency**,*" Qianqian points out. "*We're mapping it to 3D space, which allows it to handle occlusion. That's a nice property because most works on motion estimation remain in 2D. They don't build a consistent 3D representation of the scene to track.*"

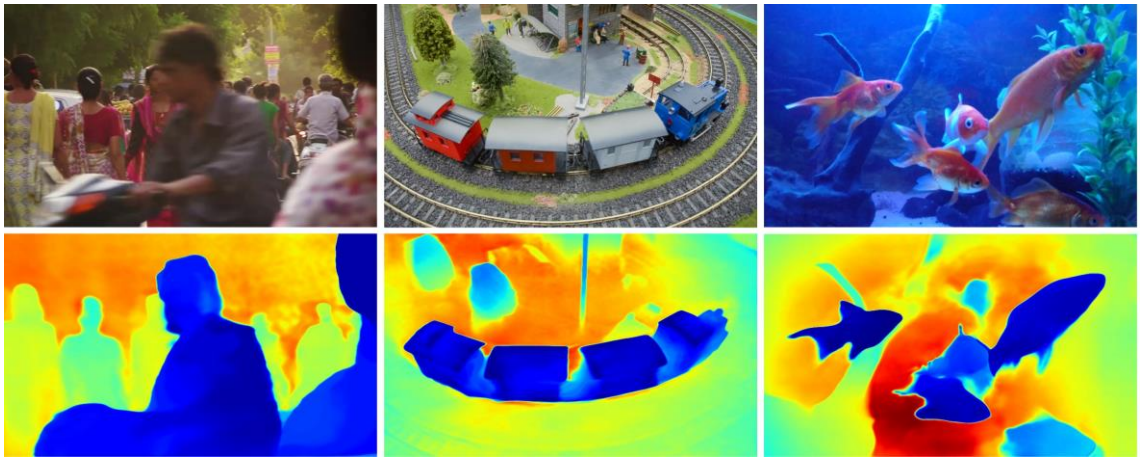
Qianqian is originally from China but has been in the US since starting her PhD in 2018 and says it is a "*very welcoming and inclusive*" environment. Her advisors on this project at **Cornell Tech** were **Noah Snavely** and **Bharath Hariharan**.

"*Noah is the most wonderful advisor in the world,*" she smiles. "*He's super caring. He has very creative ideas and guided me through the whole process. We discussed along the way and then figured out the right formulation for the problem. He pointed me to important papers that inspired me to work on this. He's super helpful, and I appreciate his guidance!*"

In her new position at **UC Berkeley**, Qianqian works with two exceptional professors, who are also great friends of our magazine: **Angjoo Kanazawa** and **Alyosha Efros**. She is in a transition stage but plans to continue working on motion estimation, 3D reconstruction, and video understanding, particularly **fine-grained and deep video understanding**. She adds that if we better understand motion in a video, we'll better understand higher-level information, like semantic information.

Where does she see herself in 10 years?

"*That's a very hard question to answer,*" she ponders. "*I still want to do research and contribute to the field of computer vision. I hope to find a faculty position in a university and stay in academia, but if that doesn't work out, I'm also fine to find a research position in industry. I'll keep doing research. That's something I know!*"



The Inline Microscopic 3D Shape Reconstruction is the winner of the ICCV 2023 best demo award for its contribution to computer vision and industrial inspection.

The demo was presented by scientists from the Austrian Institute of Technology (AIT) including Christian Kapeller, Lukas Traxler and Doris Antensteiner (in the photo with Ralph, from left to right). The demo showcased a novel optical microscopic inline 3D imaging system, which can be utilized by future industries for micro-scale object analysis.

by Doris Antensteiner

Our innovation resulted from a Computer Vision research project aimed at **retrieving accurate 3D shape at micro-scale for industrial inspection applications**. Our system fuses **light-field imaging with photometric stereo** to simultaneously capture detailed 3D shape information, object texture, and photometric stereo characteristics. The integration of light-field imaging and photometric stereo offers a holistic approach to 3D shape reconstruction.

Light-field imaging captures the angular information of the incoming light, allowing for a multi-view perspective of the object. Photometric stereo complements this by analyzing the way light interacts with the object's surface, providing crucial information about surface normals and reflectance properties.

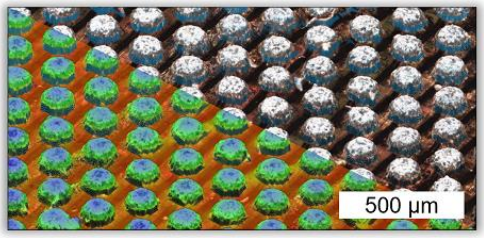
A notable feature of our system is its ability to perform **3D reconstructions without the need for relying on traditional scanning or stacking processes**, setting it apart from technologies like confocal scanning or focus stacking.

This functionality allows for **data acquisition during continuous object motion**, making it versatile for applications like inspecting moving parts on a production line or in roll-to-roll manufacturing processes. Traditional methods, such as confocal scanning or focus stacking, involve capturing multiple images from varying focal depths or perspectives and subsequently combining them to generate a 3D model. These techniques can be time-consuming and less suitable for dynamic or moving objects. In contrast, the **Inline Microscopic 3D Shape Reconstruction system excels in capturing 3D data while the object is in continuous motion**, eliminating the need for costly setup changes or interruptions.

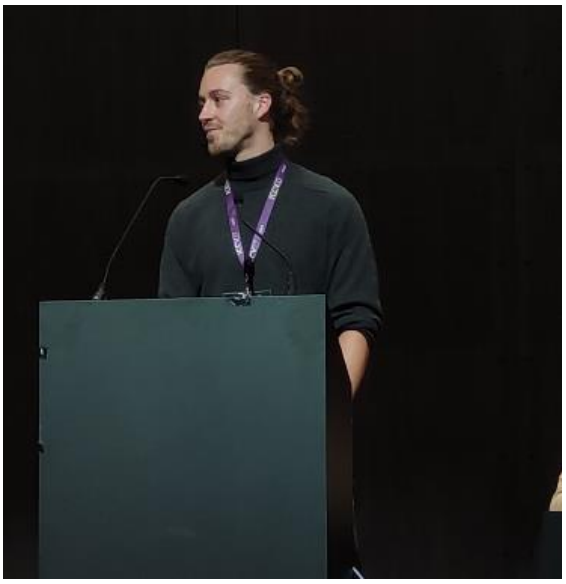
In our demo, various samples were demonstrated to showcase **a high variety of scanning scenarios**. The samples which were chosen are typically considered challenging for 3D inspection (metal surfaces, ball-grid-arrays form integrated circuits, security prints, etc.). These samples were placed on a translation stage to simulate object motion during inspection. The system demonstrated its capabilities by capturing objects with a point-to-point distance of 700nm per pixel and an acquisition speed of up to 12mm per second, equivalent to 39 million 3D points per second.

Our Inline Microscopic 3D Shape Reconstruction system has the potential to show great impact in the field of microscopic inspections in various industries. It reaches a high level of precision and efficiency in inspection processes and has a wide range of practical applications. **Our innovation has the potential to enhance micro-scale object analysis and industrial inspection in real-world scenarios.**

ICI Microscopy setup, computing 3D depth and capturing RGB data with a lateral point-to point distance of 700nm.



FunnyBirds: A Synthetic Vision Dataset for a Part-Based Analysis of Explainable AI Methods



Robin Hesse is a third-year PhD student at the Technical University of Darmstadt under the supervision of Simone Schaub-Meyer and Stefan Roth.

In his research, he works on explainable artificial intelligence with a particular interest in intrinsically more explainable models and how to evaluate explanation methods.

by Robin Hesse

Today's strong performance of deep neural networks coincides with the development of **increasingly complex model**s. As a result, humans cannot easily understand how these models are working, and therefore, only have **limited trust** in them. This renders their application in safety-critical domains such as autonomous driving or medical imaging difficult. To counteract this issue, the field of **explainable artificial intelligence (XAI)** has emerged which aims to shed light on how deep models are working. While numerous fascinating methods have been proposed to improve the explainability of vision systems, the evaluation of these methods has often been limited by the **absence of ground-truth explanations**. This issue naturally leads to the lingering question: "*How to decide which explanation method is most suited for my specific application?*", which is the motivating question for our work.

To answer this question, so far, the XAI community resorted to **proxy tasks to approximate ground-truth explanations**. One popular instance are **feature deletion protocols** where pixels or patches are incrementally removed to measure their impact on the model output and approximate their importance. However, these protocols come with several limitations, such that they introduce out-of-domain issues that could interfere with the metric, they only consider a single dimension of XAI quality, and they work on a semantically less meaningful pixel or patch level. The last point is especially important considering that **explanations aim to support humans, and humans perceive images in semantically meaningful concepts rather than pixels.**

Motivated by these limitations, our paper proposes a synthetic classification dataset that is specifically designed for the part-based evaluation of XAI methods. It consists of renderings of funny-looking birds of various 'species' on which 'semantically meaningful' image-space interventions can be performed to approximate ground-truth explanations.

Following a similar idea as the above feature deletion protocols, the dataset allows to delete individual parts of the birds, e.g., their beak or feet, to measure how the output of the model drops. If a deleted part causes a large drop in the output confidence, one can assume that the part is more important than one that only causes a minor output drop **(Fig. 1)**. This allows to move from the above pixel level to a semantically more meaningful part level and, as the training set now includes images with deleted parts, all interventions can be considered in domain.

To thoroughly analyze various aspects of an explanation, the FunnyBirds framework considers three dimensions of XAI quality and two dimensions of model quality **(Fig. 2)**. Various interesting findings were made, using the proposed FunnyBirds framework. First, architectures that were designed to be more interpretable, such as **BagNet**, often achieve higher metrics than the corresponding standard backbone networks. Second, the **VGG16 backbone** appears to be more explainable than the **ResNet-50** backbone,

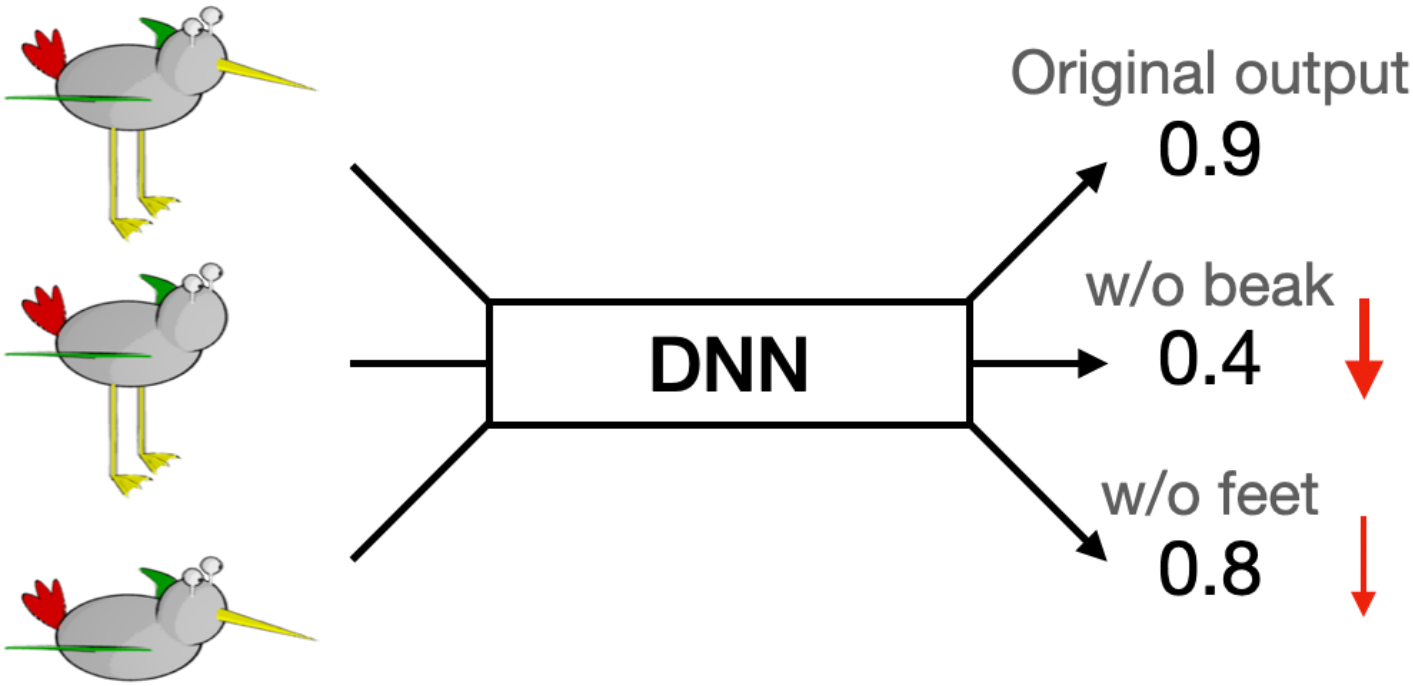


Fig 1. Removing individual bird parts and measuring the output change allows to approximate ground-truth importances for each part. In this example, the beak is more important than the feet.

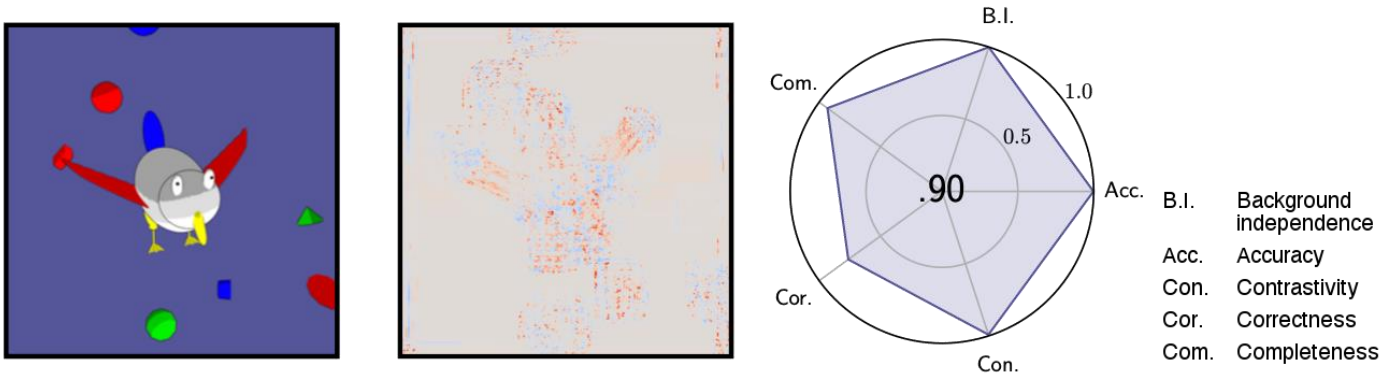


Fig 2. (left) Example image of the FunnyBirds dataset. (center) BagNet explanation for the left image. (right) Quantitative results for the examined BagNet model in the FunnyBirds framework.

indicating that different architectures or model components are more or less explainable. Third, the ranking of XAI methods may change across different backbones, so it is crucial to consider multiple backbones for the future evaluation of XAI methods.

Left: supervisor Simone Schaub-Meyer

Dear Computer Vision community,

It was **CVPR 2019** when **Vitto**, Andrew, and I sat down at a coffee shop and started brainstorming on crazy places where we could organize **ECCV in 2024**. A year later we presented the bid, and three years later here we are, full steam ahead to what will hopefully be a fun, exciting, and engaging ECCV.

We will meet in less than a year in **Milano, Italy**, capital of fashion and spritz, and now also computer vision. The Program chairs, **Olga**, **Elisa**, Gül, **Torsten**, **Ales**, and **Stefan** have been working non-stop to ensure that our scientific program will be innovative and extremely interesting.

We are looking forward to the event and we hope that you will enjoy every minute of what we have prepared for you! **Be ready to be surprised!**

Ci vediamo a Milano!

Laura

Let's all have a spritz with Laura in Milano!

NB: awesome Laura Leal-Taixé is a co-General Chair of ECCV 2024

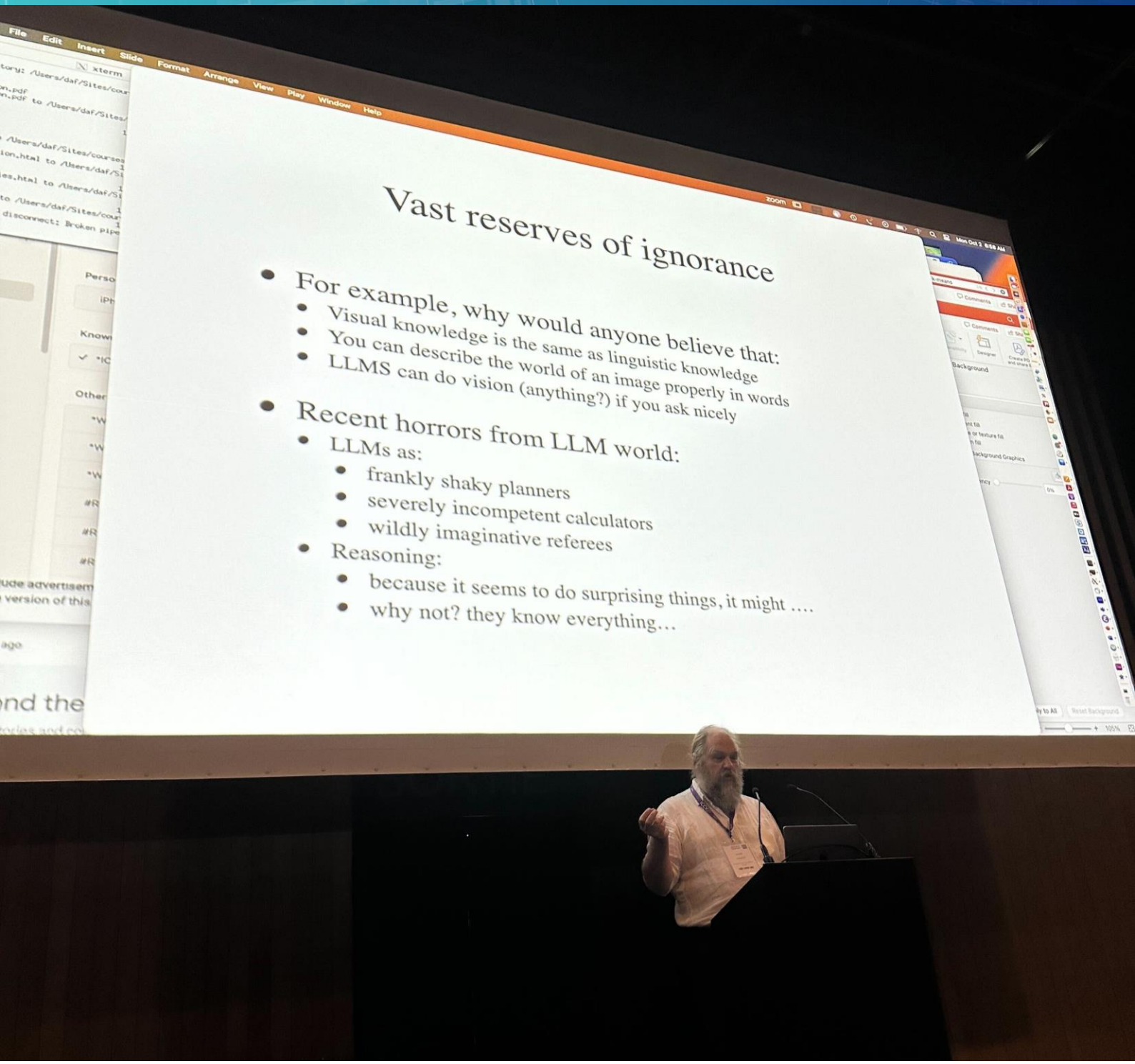
Quo Vadis, Computer Vision?

"Quo Vadis, Computer Vision?" means "Where are you headed, Computer Vision?". That's the name of a very successful workshop at ICCV 2023, showcasing a fantastic line-up of speakers. Did you miss it? Awesome Georgia got you covered!

by Georgia Gkioxari

We stand at a pivotal juncture. The past two years have been an exhilarating ride, brimming with innovation and creativity. The dawn of Generative AI (the author of this piece loves some drama!) has ushered in an epoch few could have foreseen just three years prior. Anyone claiming the contrary is with high certainty lying!

Amidst the exhilaration, there's discernible concern regarding the direction of Computer Vision research. The industry's aggressive investments, both in talent and computing power, signal a rush to capitalize on the latest technological advances. This surge is welcome; it offers more opportunities for our community members. This is nothing but a sign of a healthy field. But simultaneously, it instills a sense of uncertainty in many about their next steps.



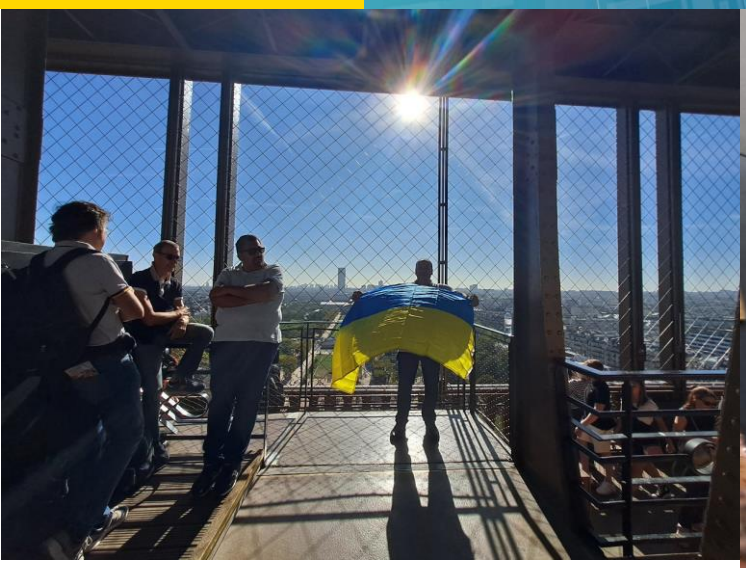
These concerns came under the spotlight and were extensively discussed at the **“Big Scholars”** workshop during CVPR, sparking debates about the trajectory of academic versus industrial research and its implications for **the future of Computer Vision.**

Arguably, our field’s fast pace is distilling our budding talents with a sense of agony around how they could make their own significant mark in this new landscape of research. This is where our **"Quo Vadis, Computer Vision?"** workshop enters the scene aspiring to guide

and galvanize young researchers in navigating this transformed research milieu. We've asked experts from diverse backgrounds and research focus to share their insights. We've posed to them an important question: "In today's landscape, what would you, as a grad student, focus on?". Many of us, including the organizers of this workshop, are staunch in our belief that countless challenges in CV await solutions. To put it another way, the most crucial problems remain unconquered. But we are concerned that this sentiment isn't universally shared by our emerging scholars. We are optimistic that our seminar will inspire them to think, delve deeper, and find their place in this ever-evolving landscape of Computer Vision.

“ … countless challenges in CV await solutions. To put it another way, the most crucial problems remain unconquered…”

Paris supports Ukraine



ICCV's sister conference **CVPR** adopted a motion with a very large majority, condemning in the strongest possible terms the actions of the Russian Federation government in invading the sovereign state of **Ukraine**. One day before ICCV in Paris, we decided to involve the **Eiffel Tower** and the **Mona Lisa**. Photos credit to awesome **Doris Antensteiner**!

Computer Vision News

Editor:
Ralph Anzarouth

Ralph's photo on the right was taken in lovely, peaceful and brave Odessa, Ukraine.

Publisher:
RSIP Vision

Contact us

Give us feedback

Free subscription

Did you subscribe to Computer Vision News? It's free, click here!

Read previous magazines

Copyright: **RSIP Vision**
All rights reserved
Unauthorized reproduction is strictly forbidden.

RSIP VISION Visual Intelligence for MedTech

Nadiya Shvai is currently a Senior Data Scientist responsible for research at Cyclope.AI.

Where shall we start? From Nadiya or from Cyclope.AI?

[*laughs*] Let's start with Cyclope.AI because I think we'll have plenty to talk about.

Perfect!

Cyclope.AI is a relatively small company that works on artificial intelligence-based solutions for smart road infrastructure and safety. For example, among the products that we do is the security system for the tunnels. You've probably heard about the accident in the Mont Blanc Tunnel that happened some years ago. After this, the regulations

Read 100 FASCINATING interviews with Women in Computer Vision

for the safety of tunnels have been reinforced a lot in France. We are working on the automation of the system to make sure that they are as fault-proof as possible. At the same time, they do not generate a lot of false alarms because a system that generates a lot of false alarms finally becomes not useful at all.

What do you do there as the data scientist?

My work really considers almost all the aspects of deep learning product development. Starting from the data collection to data selection, to supervising the data labeling, to model training and testing. Then, we put the deep learning models into the pipeline and finally prepare this pipeline for deployment. This is additional to the other research activities that we do.

Is this what you wanted to do when you studied? Or was it an opportunity that came to you, and you took it?

[*Thinks a little moment*] It's an opportunity that came to me, and I took it. This has more or less been happening throughout my professional path. I think it's normal for opportunities to come our way, and it's important to recognize them and grab them. Recognize those that speak to you, that are close to your spirit.

During your studies, what did you think you would be doing when you grew up?

Ahh, it's a very good question!

Thank you. I didn't come for nothing. [*both laugh*]

Well, deep learning as a mainstream activity is relatively new. It comes from signal processing, but this was not my specialization when I was studying. At the core, I'm a mathematician. You can think about this as being relatively far from what I do because I was doing linear algebra, and my PhD is also on linear algebra. But then, slowly, I drifted towards more applied skills, which is how I came to where I am today.

So it's not true that women are weaker in mathematics, or are you

special?

[*laughs*] No, I really don't think that I'm special. I honestly don't think that women are weaker in mathematics. However, I think we have to talk about the point that we are coming from. We're coming from the point that there is enough of the existing bias of what women should be occupied with and the lack of the examples of the women researchers. That's why the interviews that you do are so important. They provide examples to other women and young girls to broaden their spectrum of possibilities and realize, yes, I can do this. This is possible for me!

You've told us something about the present and something about the past. Let's speak about the future. Where are you planning to go?

Currently, I'm increasing the amount of research activities in my day-to-day work. This is my current vector of development. But where it will bring me, I don't know for now. I do know that this is what I am enjoying doing, and this is important for me.

Can you be a researcher all your life?

[*hesitates a moment*] Hopefully. If we're talking from the mathematician's point of view, there is this preconception that mathematicians usually are most fruitful in their 20s, maybe 30s. Then, after this, there is some sort of decline in activity.

I never heard that. That would be terrible if it were true.

[*laughs*] This is a conception that I have heard, and I'm not sure if there are actually some sort of statistics regarding this. But in one form or another, I would like to continue doing research as much as possible. Because for me, one of my main drives is curiosity. That's what makes research appealing to me. I don't think this curiosity is going to go away with time.

Are you curious about learning new things to progress yourself or to make progress in science? What is your drive?

I'm not that ambitious to think that I'm going to push science forward. For me, it's to discover things for myself or for the team, even if it's a small thing. I also enjoy seeing the applied results of the research that we do, because I believe that deep learning is the latest wave of automation and industrialization. The final goal is to give all the repetitive tasks to the machine, so we as humans can enjoy more creative tasks or just leisure time.

You just brought us to my next question!

[*laughs*] Please go ahead.

What has been your greatest success so far that you are most proud of?

If we're talking about automation, I was the person responsible for training and testing the model that right now does the vehicle classification according to required payment at the tolls all over France. It means that every day, literally hundreds of thousands of vehicles are being classified using the computer vision models that I have trained. So, I'm at least partial to the final product, and it means less of a repetitive job for the operators. Before, there was a need for the operator because physical sensors were not able to capture the differences between some classes, so humans had to check this. This is a job very similar to labeling. If you ever did the labeling of images and videos, you know how tough it actually is. You have to do it hour after hour after hour; it's quite tough. So right now, I'm super happy that a machine can do it instead of a human.

What will humans do instead?

Something else. [*laughs*] Hopefully, something more pleasant or maybe more useful.

That means you are not in the group of those who are scared of artificial intelligence taking too much space in our lives.

In our workload? No, I don't think so. First of all, as a person working with AI every day, I think I understand pretty well the limitations that it has. Definitely, it

cannot replace humans, but it's just a tool. It's a powerful tool that enables you to do things faster, to do things better. And am I worried about AI in regard to life? Maybe to some extent. Sometimes, when I see some things, I think, do I want this for myself or for my family? The answer is no. But again, it's rather a personal choice. For example, I think a couple of years ago, I saw this prototype of an AI-powered toy for really young kids who can communicate with the kid, etc. And honestly, I am not sure that this is something that I would like for my kids. I don't think that we are at the point that it's A, really safe, and B I think it might be a little bit early for the child to present this to them. It might create some sort of confusion between live beings and AI toys. But again, this is just my personal opinion, and here everyone chooses for themselves.

Nadiya, up until now, our chat has been fascinating. My next topic may be more painful. You are Ukrainian, and you do not live in Ukraine. How much do you miss Ukraine, and what can you tell us about how you have been living the past 18 months?

[*hesitates for a moment*] You feel inside as if you are split in two. Because for me, I live in France, and I have to continue functioning normally every day. I go to work, I spend time with my family, I smile at people, etc. Then there's a second half that reads news or gets messages from friends and family that are facing the horror and the tragedy and the pain of war. Of course, it cannot be even closely compared to people's experience who are in Ukraine right now. But I believe there is no Ukrainian in the world that is not affected by the war.

How can life go on when somebody is burning down your house?

I honestly don't know. But it has to, as you cannot just stop doing whatever you are doing and say I'm going to wait until the war is over.

Can you really say, okay, business as usual? Sometimes, don't you feel the whole world should stop and say, hey, come on, this can’t go on?

[*hesitates for a moment*] I wish it could be like this, but it's not like this. We have to do our best in this situation that we are in.

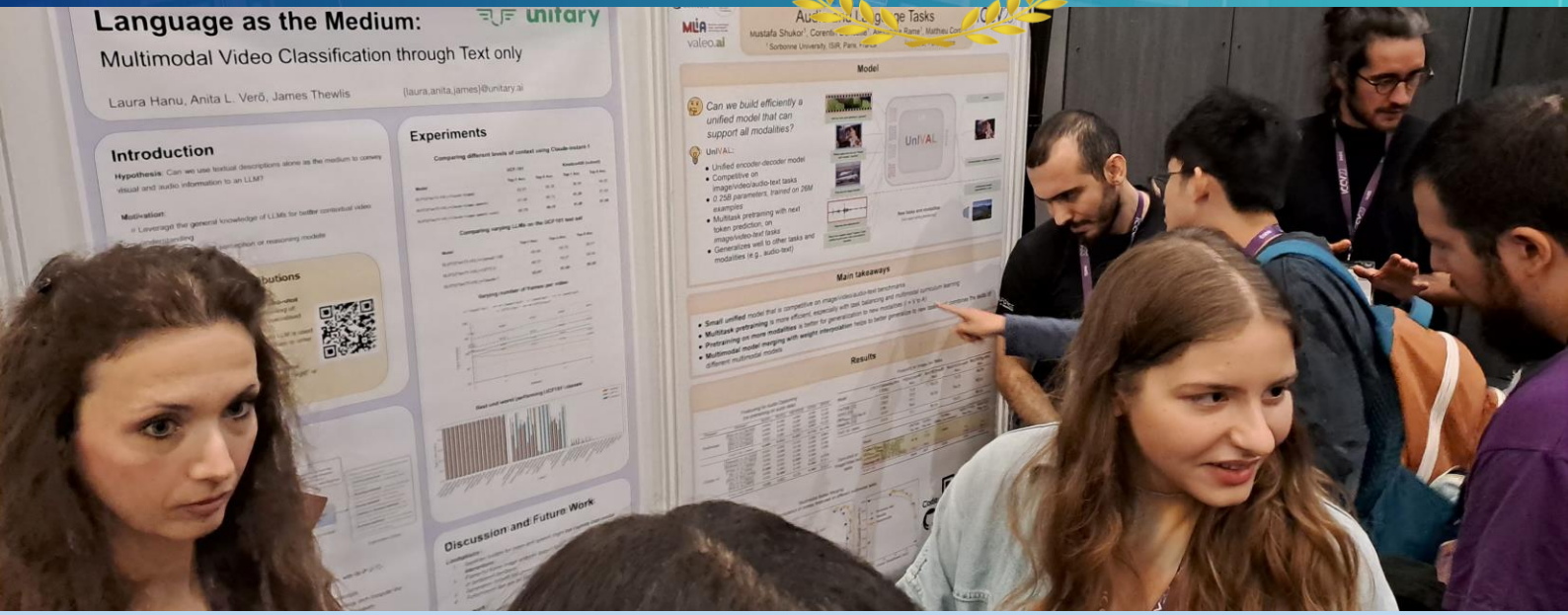
"I do feel the support of the research community, and I appreciate a lot the work that they are doing. It means a lot to me personally!"

Do you know, Nadiya, that one and a half years ago CVPR passed a resolution condemning the invasion of Ukraine and offering solidarity and support people of Ukraine? You enjoy a lot of sympathy in this community. Can you tell all of us what you expect from us to make things easier for you?

I do feel the support of the research community, and I appreciate a lot the work that they are doing. It means a lot to me personally, and I'm sure that it means a lot also for other Ukrainians. Being seen and heard is one of the basic human needs, particularly in the worst situation that we are in right now. To use our conversation as a stage, I think that the best the community can do is to provide support to Ukrainian researchers, particularly for those who are right now staying in Ukraine. For collaborations and projects, this is probably the best approach.

Do you have anything else to tell the community?

[*hesitates for a moment*] Sure, it's not work-related, but what I want to say is that a couple of days ago, I was reading a book. There was a quote in the book that I liked a lot that I'd like to share: "*There are no big breakthroughs. Only a series of small breakthroughs.*" And I'm saying this to support young researchers, particularly young girls. Just continue there, and you're going to achieve. Right? This is also my word of support to all Ukrainians who are going to read this. Every day brings us closer to victory.



Laura Hanu (right) and Anita L Verő are both Machine Learning Research Engineers at Unitary, a startup building multimodal contextual AI for content moderation. Laura told us that in this work, they demonstrate for the first time that LLMs like GPT3.5/Claude/Llama2 can be used to directly classify multimodal content like videos in-context with no training required.

"*To do this,*" she added, "*we propose a new model-agnostic approach for generating detailed textual descriptions that capture multimodal video information, which are then fed to the LLM along with the labels to classify. To prove the efficacy of this method, we evaluate our method on action recognition benchmarks like UCF-101 and Kinetics400.*"

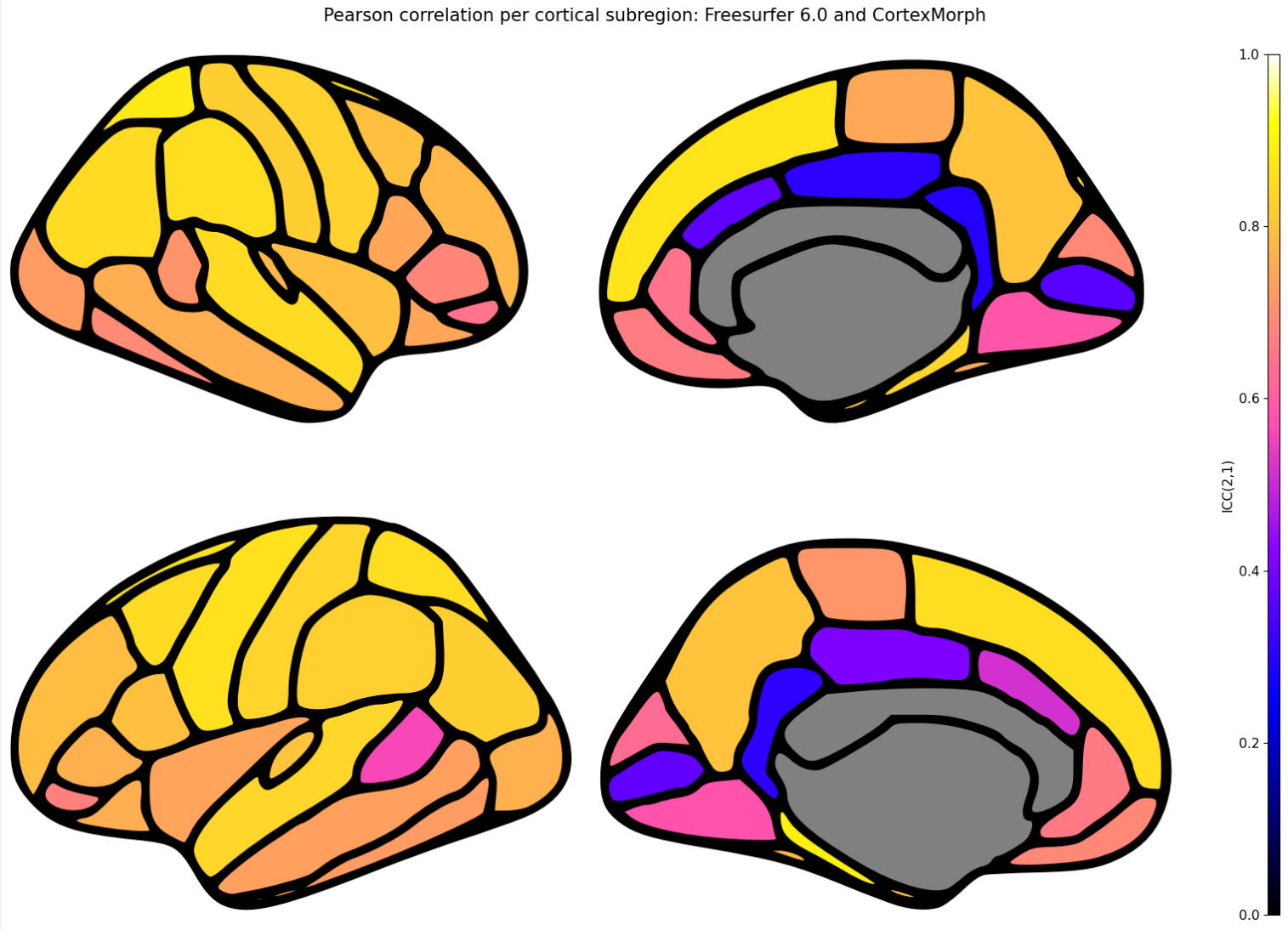
Ivan Reyes-Amezcua is a PhD student in Computer Science at CINVESTAV, Mexico. He is researching adversarial robustness in deep learning systems and developing defense mechanisms to enhance the reliability of models.

He presented his poster at the LatinX workshop, demonstrating how subtle changes to images can fool a model: shifting its confidence from identifying an image as a pig to confidently labeling it as an airliner.

BEST OF MICCAI 2023

MEDICAL IMAGING NEWS

NOVEMBER 2023



What's this?
Find out on page 50!

Automated AMD biomarker discovery

by Christina Bornberg
@datascEYEnce

It is time for another **deep learning in ophthalmology** interview as part of the datascEYEnce column here in the **Computer Vision News** magazine!

I am Christina and through my work with retinal images, I come across a lot of amazing research that I want to share with you! This time, I interviewed **Robbie Holland** from Imperial College London on his work on age-related macular degeneration (AMD)!

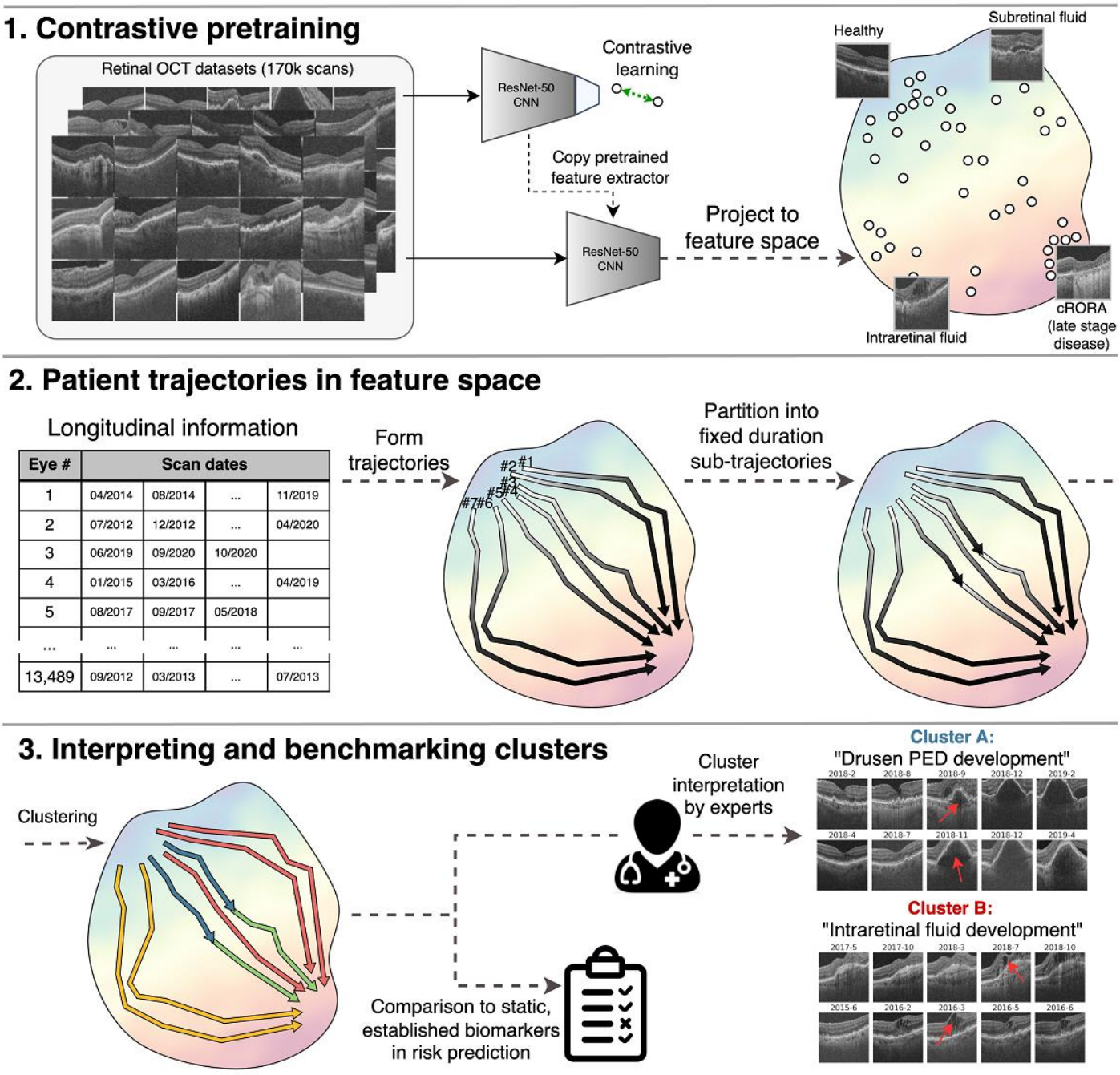
featuring Robbie Holland

Robbie's decision to get involved with deep learning for ophthalmology was influenced by both, his interest in modelling complex systems as well as reading a publication by DeepMind in 2018: "**Clinically applicable deep learning for diagnosis and referral in retinal disease**". Following his undergrad in Mathematics and Computer Science as well as a project on abdominal MRI analysis, he started his PhD with a focus on the early detection of AMD under the supervision of Daniel Rueckert and Martin Menten in the BioMedIA lab, Imperial College London.

Leaving behind a world of completing Jira tickets as a software engineer led him to work on finding disease trajectories for AMD, which he was able to present at last month's **MICCAI in Vancouver**!

In case you missed out on it, I am covering the key points of "Clustering Disease Trajectories in Contrastive Feature Space for Biomarker Proposal in Age-Related Macular Degeneration" here!

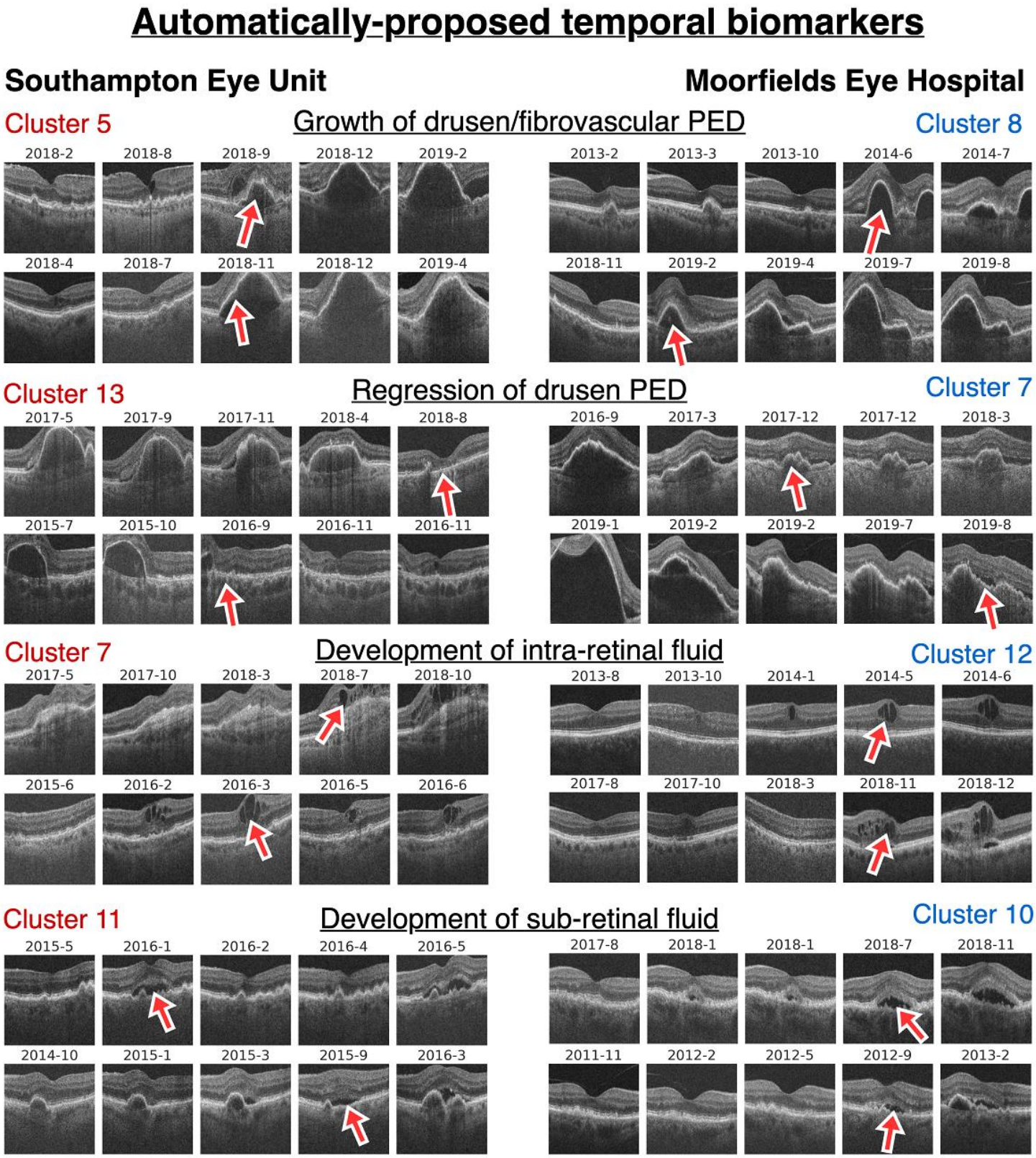
Let's start with the application. Robbie explained the limitations of common grading systems for AMD to me - they lack the ability of prognostics. Simply said, it is unclear how long it will take until a patient transitions from early-stage to a late stage of AMD. Some patients progress quicker than others.



To account for this, his research focuses on using deep learning for automatic temporal biomarker discovery. In other words, **clustering disease progression trajectories in a pre-trained feature space**.

The choice of a self-supervised approach, specifically **contrastive learning**, was made in order to identify trajectories, or more precisely sub-trajectories. Contrastive learning methods have shown their capability to autonomously learn disease-related features (including previously unknown biomarkers) without the need for clinical guidance.

For the setup, a ResNet-50 backbone was trained on cost-effective yet informative OCT scans. The self-supervised loss function **BYOL contrastive loss** makes it possible to train only on positive samples. The decision to use this



specific loss function for a self-supervised learning task is based on his previous work (and by chance got emphasized by Yann LeCun!!).

Now the ResNet-50 backbone can be used to extract features which are subsequently projected in a feature space. The next step is clustering.

Clustering sub-trajectories allows to find common paths for disease progression among the patients. In this work, spectral clustering was applied. Why not k-means? Because the distance function is not Euclidean.

The trajectories have an agnostic number of points and therefore it is difficult to find an average value for representation.

So, what is the final outcome? Robbie and his collaborators were surprised at how well contrastive learning can model human concepts. With the help of their deep learning pipeline, they were able to isolate patterns of disease progression that are suspected by clinicians to be related to AMD. In other words, the clinicians were able to relate the automatically generated clusters to known and additionally yet unknown temporal biomarkers.

As a last question, I asked Robbie about advice for prospective PhD students. For everyone who also wants to pursue research in deep learning for ophthalmology, Robbie emphasized research in self-supervised learning applied to retinal images. It is an exciting field, where you can try new big ideas - another very good example is RetFound which was recently published by researchers from UCL/Moorfields. In general, there is a high demand for understanding eye-related diseases and a lot of problems that remain to be solved!!

More about AI for Ophthalmology

Do you enjoy this November issue of Computer Vision News?

We are glad that you do!

We have one more important community message to tell you. It's an advert for something free ☺

Just go to page 64,
and you'll know.
Keep in touch!

BEST OF MICCAI 2023

An Explainable Geometric-Weighted Graph Attention Network for Identifying Functional Networks Associated with Gait Impairment

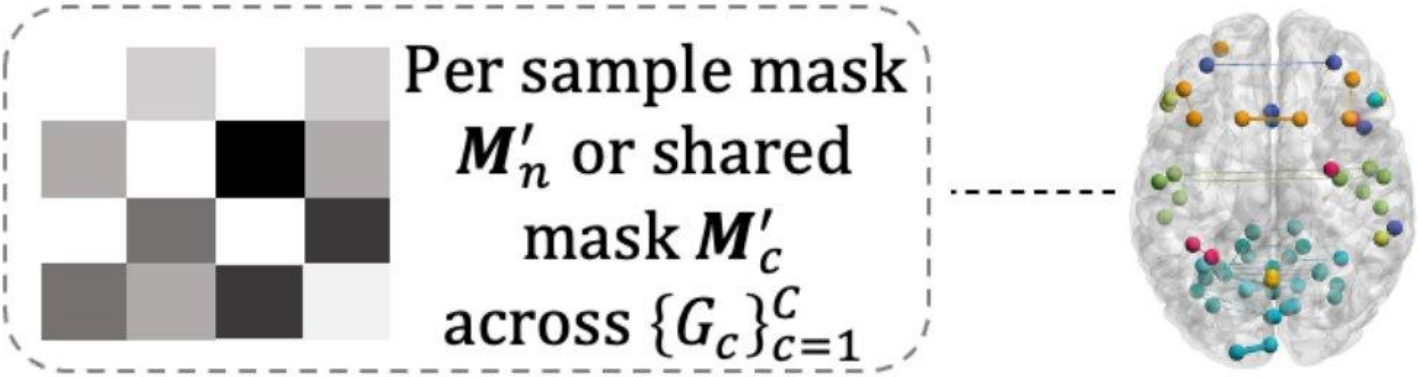
Favour Nerrise is a PhD candidate in the Department of Electrical Engineering at Stanford University under the supervision of Ehsan Adeli and Kilian Pohl.

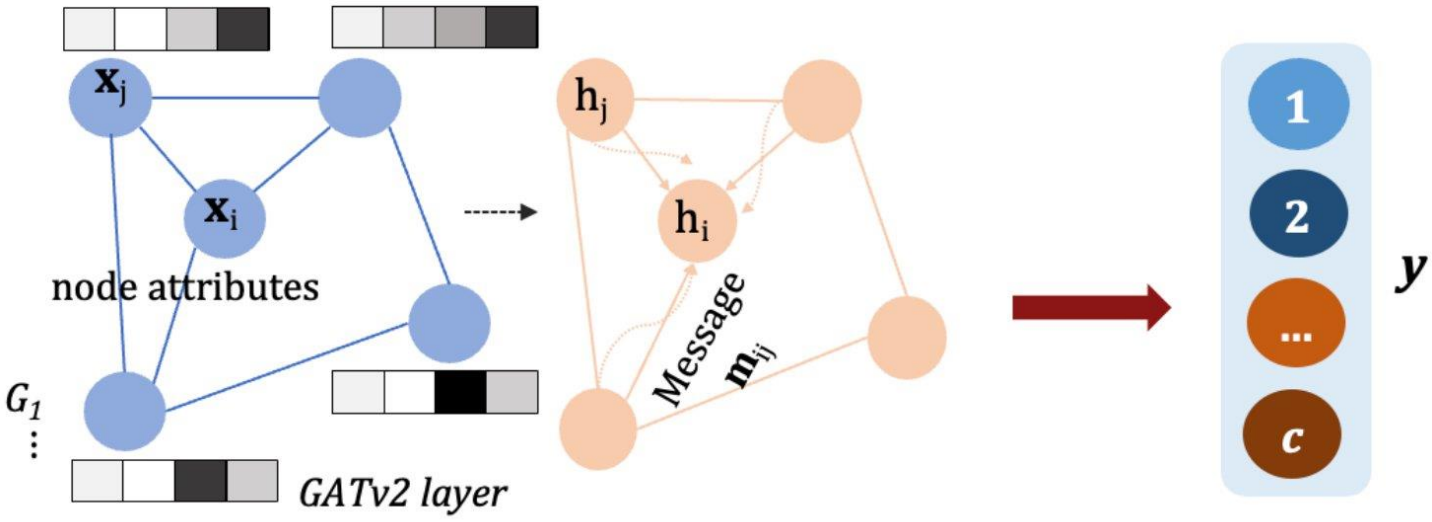
Her work proposes a new method to derive neuroimaging biomarkers associated with disturbances in gait for people with Parkinson's disease and has been accepted for an oral and poster presentations.

She spoke to us ahead of her presentations at MICCAI 2023.

Parkinson's disease, a neurodegenerative disorder affecting millions worldwide, has long been a focus of research to improve diagnostics and treatment. In this paper, Favour presents a method capable of **deriving neuroimaging biomarkers associated with disturbances in gait** – a common symptom in individuals with Parkinson's disease. However, the significance of this work extends beyond the laboratory. **Favour is determined to make a tangible impact on clinical practice**.

"A big piece of the work is an explainability framework, meaning that it's not only computational for how other medical physicists can





use the work, but more importantly, we created visualizations that you can turn on and off through our pipeline that allow people with relevant clinical or neurological knowledge to interpret what our computational model is doing on the back end," she explains. *"We hope that helps computational scientists, neuroscientists, or clinicians who want to adopt and expand this work translate it into clinical understanding."*

One of the most significant challenges was the scarcity of clinical data, a common issue in research involving rare diseases like Parkinson's. Clinical datasets are typically small and imbalanced, with more healthy cases than diseased ones. **Favour developed a novel statistical method for learning-based sample selection** to address this. This method identifies **the most valuable samples in any given class for training, oversampling them to achieve a balanced representation across all classes**.

"There were some existing methods that did sample selection either randomly or through synthetic oversampling, but we thought it'd be better to address this directly in a stratified way," she tells us. *"Making sure there's* ***equal representation of that strong sample bias across every class*** *before applying an external method like random oversampling."*

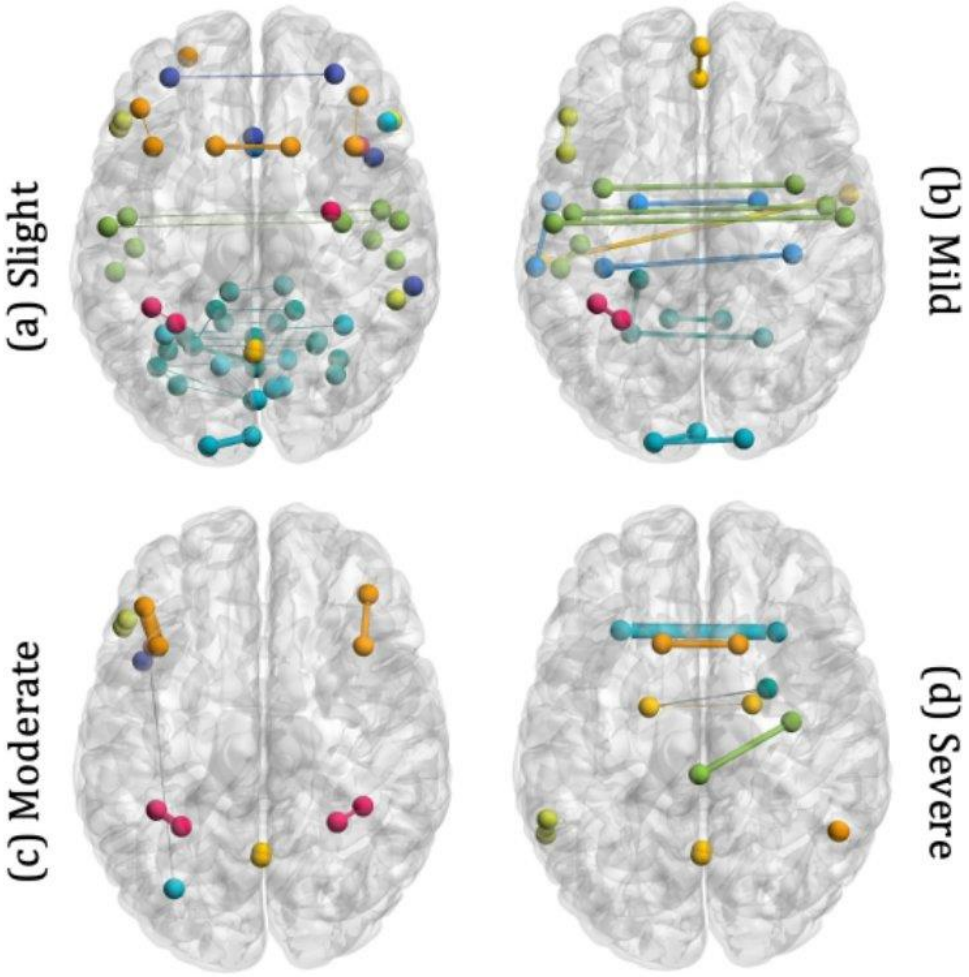
The method performed very well, surpassing existing approaches, including those that inspired Favour to take up the work in the first place, with up to **29% improvement for area under the curve (AUC)**. The success can be attributed to **a solution comprising various techniques** rather than a linear approach of only visualization explainability or sample selection.

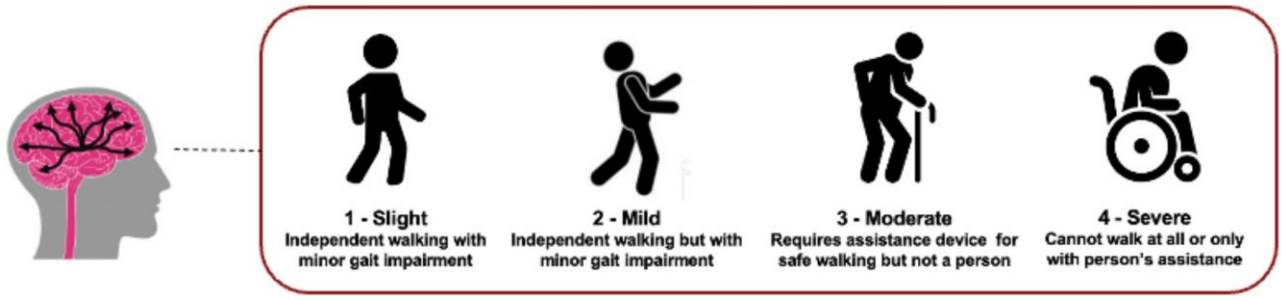
Favour plans to expand the research by evaluating it on a larger dataset, **the Parkinson's Progression Markers Initiative (PPMI) database**, and exploring new methods involving **self-attention and semi-supervised techniques**.

BEST OF MICCAI 2023

"*We also want to do something that I'm really passionate about, which is trying to identify* ***cross-modal relationships between our various features,***" she reveals. "*In this work, we focus on the neuroimaging side, taking the rs-fMRI connectivity matrices and then optimizing that using Riemannian geometry and leveraging some features. Now, I'm interested in combining some of the patient attributes and seeing how that could better inform linkages that could be learned during training amongst other types of techniques we'll try.*"

Favour has several follow-up projects in the pipeline that promise to push the boundaries further, leveraging attention-based methods and geometric deep learning techniques. Beyond neuroimaging, she aims to incorporate video data from the same patients. This multimodal approach is a significant next step. She intends to derive specific motion biomarkers that can be associated with the existing features. This expansion aims to optimize the learning process and further enhance the understanding of those linkages. The ultimate goal is to combine all these modalities into a comprehensive framework that can be generalized to a broader population. She envisions creating **a foundation model** that can serve as a valuable resource for researchers and clinicians in various downstream tasks.





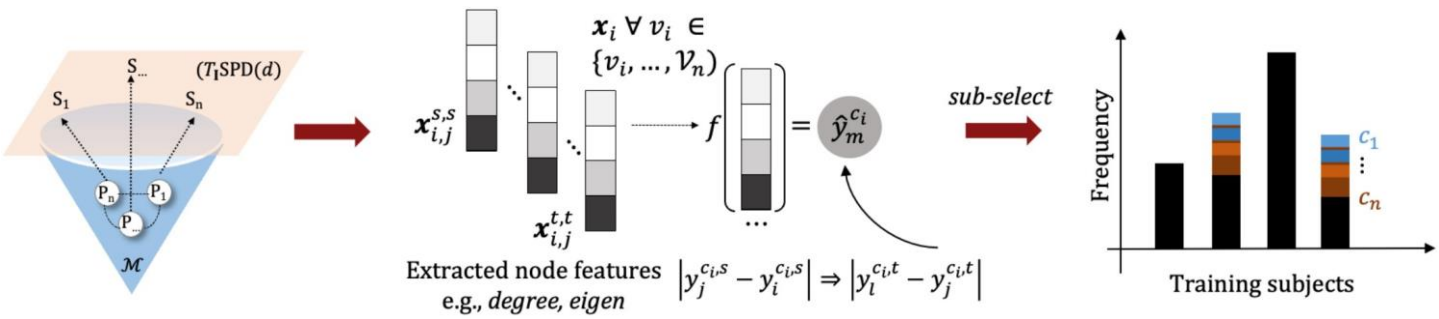
Research journeys are rarely smooth sailing, and Favour tells us a remarkable aspect of this one was the need for constant course correction in her coding efforts. As the deadline for MICCAI submission approached, the results were not exactly where she needed them to be.

"It was weighing on my heart so heavily," she admits. *"At first, we were even doing an approach of* ***comparative binary classification and multi-class classification****, and things just weren't making sense. Then, I just focused on the multi-class classification. Once I did that and started to look into how I could directly optimize my metrics, ensuring everything was weighted in my loss functions, sampling techniques, and all those things, we started to see consistent results that could be repeated over trials. I was so concerned about that because I'd get good results here and there, but I couldn't repeat them. I was so happy once it got* ***stable enough to have reproducible results****. That literally happened a few weeks before we were supposed to submit! I kept updating my paper every day until the night of submission."*

Favour remains dedicated to her academic journey. With a couple of years to go until she completes her PhD, she is committed to ongoing research in the field and leadership roles both on and off campus.

As we wrap up our time together, she acknowledges the importance of the support she has received along the way. This work has been a significant milestone as her first self-owned project and paper to be released during her graduate school career. It now has the honor of being accepted as an oral at a prestigious conference like MICCAI.

"I literally can't believe it!" she smiles. *"To any other PhD students thinking, I don't know what I'm doing, does this even make sense? I don't understand what I'm writing. Just have faith in your work because when I read my paper, I'm like, did I write this?! Just have faith in the work you're doing, and somebody will love it, and you're absolutely brilliant, and it's going to be worth it!"*

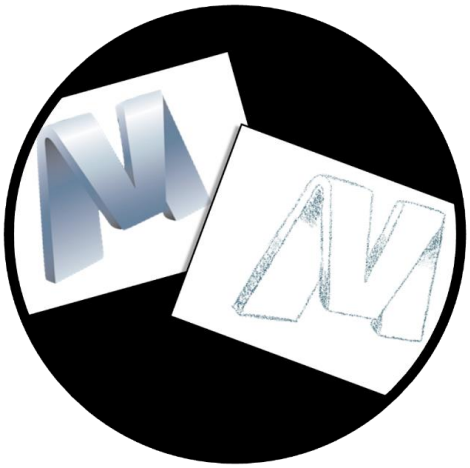




by [Ninon Burgos](), Camille Brianceau and Elina Thibeau-Sutre

Reproducibility is a key component of science, as the replication of findings is the process by which they become knowledge. It is widely considered that many fields of science, including medical imaging, are undergoing a reproducibility crisis. This has led to the publications of various guidelines to improve research reproducibility, such as the checklist proposed by MICCAI.

During the tutorial, the participants were asked to comment on the reproducibility of various fake MICCAI-like papers that we invented. They were guided by the MICCAI checklist, which covers several elements: Models and Algorithms, Datasets, Code and Experimental results. As there is no absolute answer when analysing reproducibility, this led to rich discussions and a few conclusions.



During the tutorial, the participants were asked to comment on the reproducibility of various fake MICCAI-like papers that we invented. They were guided by the **MICCAI** checklist, which covers several elements: **Models and Algorithms, Datasets, Code and Experimental results**. As there is no absolute answer when analyzing reproducibility, this led to rich discussions and a few conclusions.

First and foremost, it is not easy to understand what is expected in terms of reproducibility. Should the authors aim for exact reproducibility or rather conceptual reproducibility? Should the items of the reproducibility checklist to focus on be adapted to the type of paper? For example, a paper presenting a new method may need to particularly focus on the "Experimental results" section to adequately demonstrate the improvements reached compared with the state-of-the-art. The second main point of discussion was **the role of the reviewers**. Should they comment on the reproducibility of the paper in general, verify the consistency between the checklist and the paper, or both? Finally, the use of the checklist in the decision-making process was unclear to the participants, who were both authors and/or reviewers, which may impede the growth of reproducible research within the MICCAI community.

We hope that the participants of the tutorial now have a better understanding of the concept of reproducibility and how to implement it in practice. We are also calling on the community to help clarify the points raised during our discussions for the future MICCAI conferences!

- Data set ✗
- Code ✓
- Computational setup ✓
- Training ✓
- Evaluation ✗

BEST OF MICCAI 2023

Detecting the Sensing Area of a Laparoscopic Probe in Minimally Invasive Cancer Surgery

Baoru Huang is a PhD candidate at the Hamlyn Centre, Imperial College London, supervised by Daniel Elson and Stamatia (Matina) Giannarou.

Her work explores an innovative and groundbreaking visualization technique that holds significant promise for advancing cancer surgery.

She spoke to us ahead of her oral and poster presentations at MICCAI 2023.

Cancer remains a significant global challenge, with one diagnosis every two minutes in the UK alone. Due to a lack of reliable intraoperative visualization tools, **surgeons often rely on a sense of touch or the naked eye** to distinguish between cancerous and healthy tissue. Despite advances in preoperative imaging methods such as PET, CT, and MRI, pinpointing the precise location of cancerous tissue during surgery remains a formidable task.

Recently, minimally invasive surgery has garnered increasing attention for its potential to minimize blood loss and shorten recovery times. However, this approach presents another unique challenge for surgeons as they lose tactile feedback, making it even more difficult to locate cancerous tissue accurately.

Lightpoint Medical Ltd. has introduced a miniaturized cancer detection probe named **SENSEI**. This advanced tool, the first of its kind, leverages the cancer-targeting capabilities of nuclear agents typically used in nuclear imaging. By detecting the emitted gamma signal from a radiotracer that accumulates in cancerous tissue, **surgeons gain real-time insights into the location of cancer during surgery**.

"This probe can be inserted into the human abdomen and then grasped by a surgical tool," Baoru tells us. *"However, using this probe presents a visualization challenge because it's non-imaging and is air-gapped from the tissue, so it's challenging for the surgeon to locate the probe-sensing area on the tissue surface. Determining the sensing area is crucial because we can have some signal potentially indicating the cancerous tissue and the affected lymph nodes."*

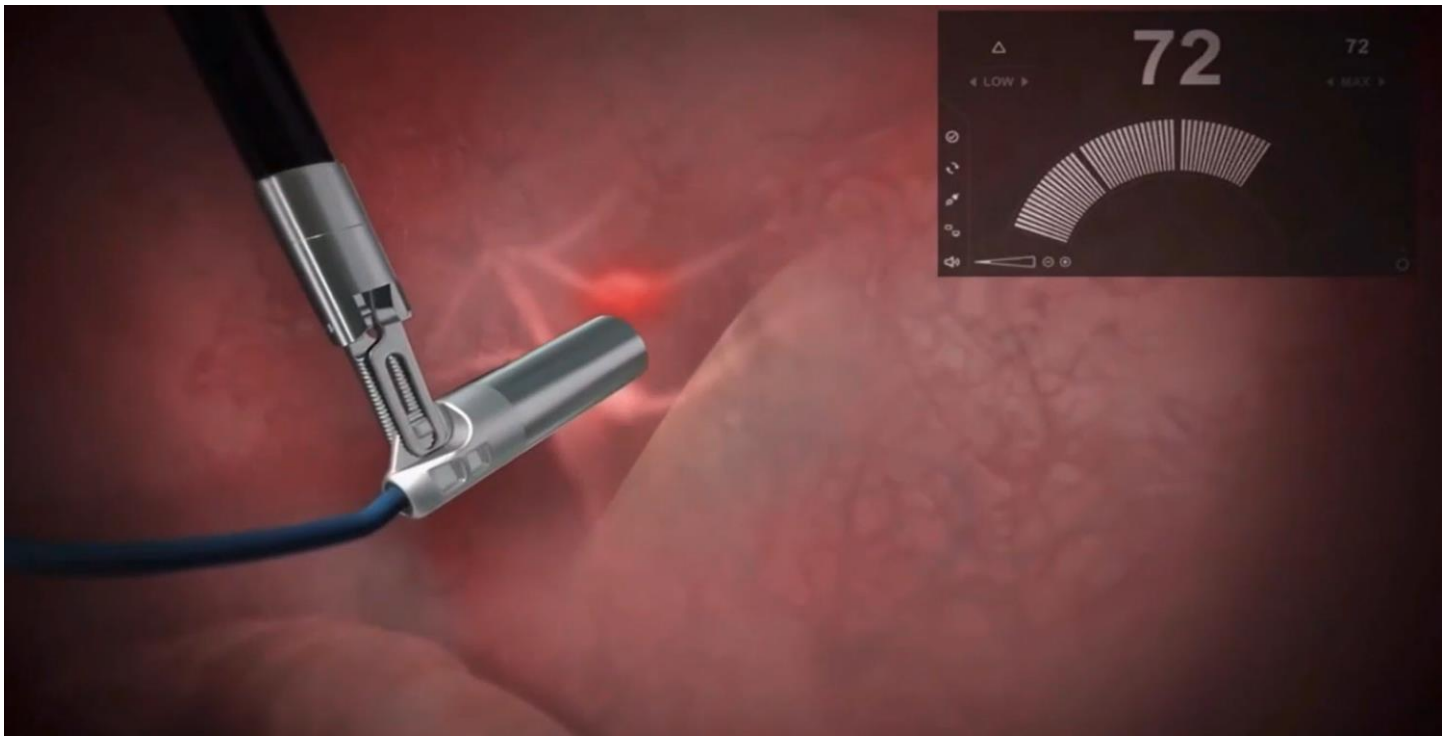
Geometrically, the sensing area is

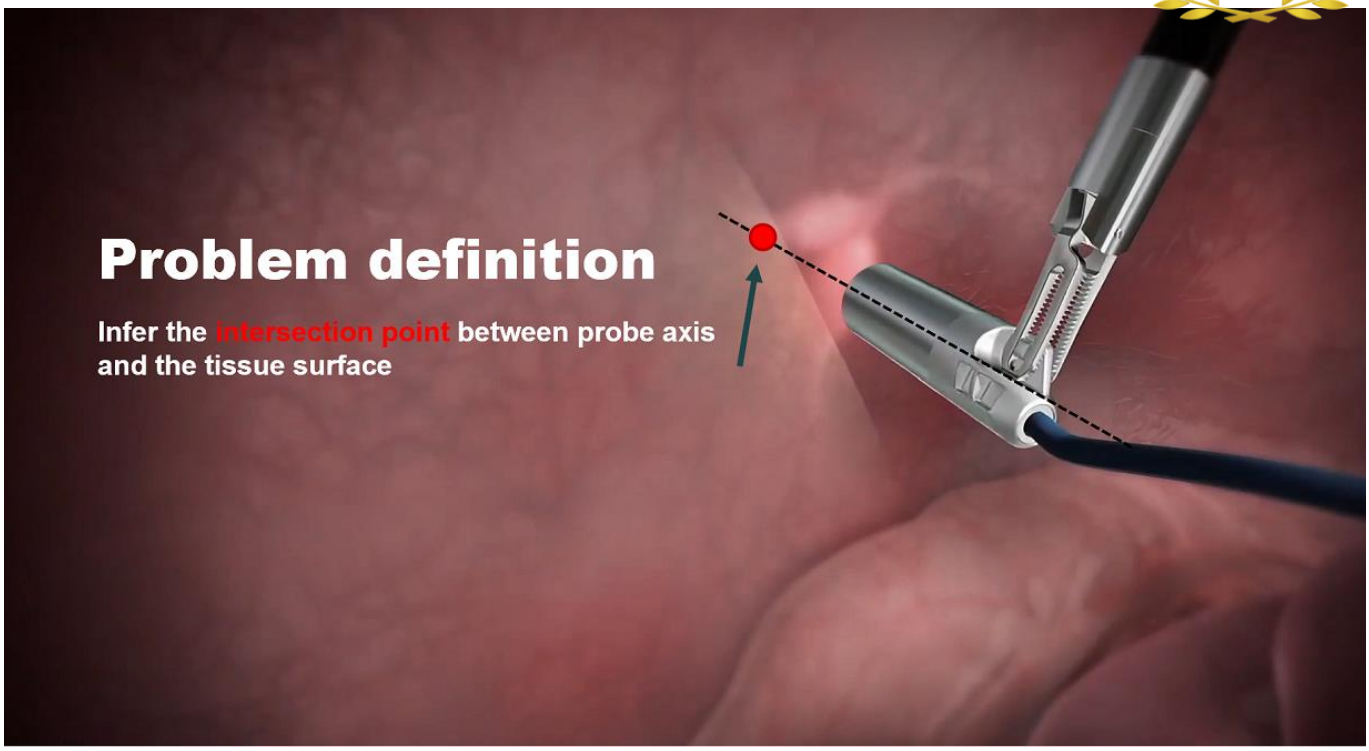
defined as the intersection point between the gamma probe axis and the tissue surface in 3D space but then projected onto the 2D laparoscopic image. It's not trivial to determine this using traditional methods due to the lack of textural definition of tissues and per-pixel ground truth depth data. Also, it's challenging to acquire the probe pose during the surgery. To address this challenge, Baoru redefined the problem from locating the intersection point in 3D space to finding it in 2D.

"*The problem is to infer the intersection point between probe access and the tissue surface,*" she continues. "*To provide the sensing area visualization ground truth, **we modified a non-functional SENSEI probe by adding a DAQ-controlled cylindrical miniaturized laser module**. This laser module emitted a red beam visible as red dots on the tissue surface to optically show the sensing area on the laparoscopic images, which is also the probe axis and the tissue surface intersection point. This way, we can keep the adapted tool visually identical to the real probe by inserting a laser module inside. We did no modification to the probe shell itself.*"

Baoru's solution involves a multi-faceted approach. Firstly, she modified the probe. Then, she built a hardware platform for data collection and a software platform for the learning algorithm to facilitate the final sensing area detection results.

With this setup, it is possible to find the laser module on the tissue surface, but the red dot is too weak compared with the laparoscope light. To solve this, she used a shutter system to control the laparoscope's illumination, closing it when the laser is turned on and opening it when it is turned off.





This process ensures the laser point is visible on the tissue surface despite the ambient lighting conditions.

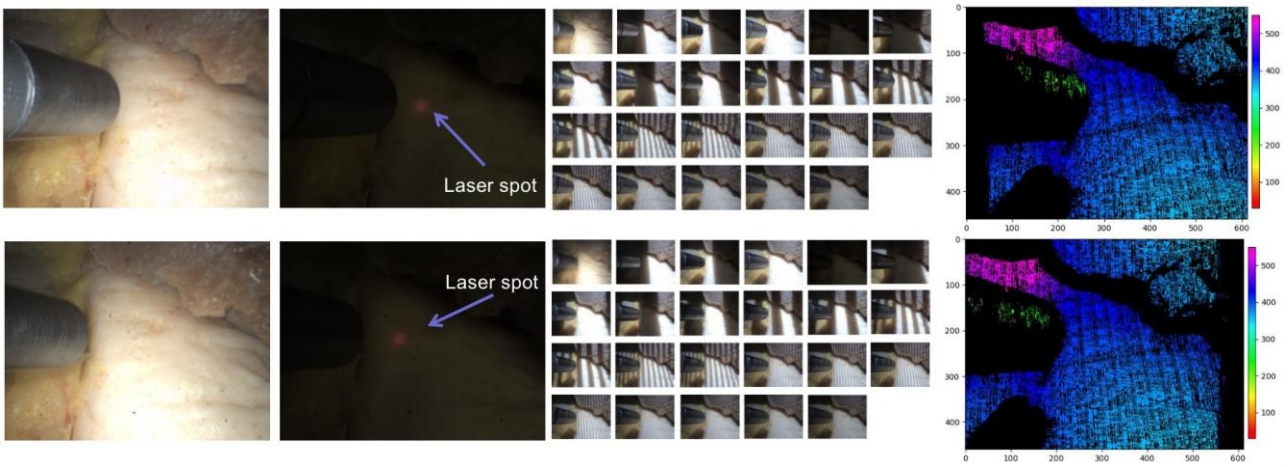
"*Our network includes two branches,*" she explains. "*For the first branch, the images fed to the network were the 'laser off' stereo RGB images, but crucially, the intersection points for these images were known a priori from the paired 'laser on' images. Then, we use the PCA, the Principal Component Analysis, to extract the central axis of the probe on the 2D. Then, we want to feed this information to the second branch. We sampled 50 points along this axis as an extra input dimension.*"

The network employed **ResNet** and **Vision Transformer** as backbones, and the principal points were learned through either a **multi-layer perceptron (MLP)** or a **long short-term memory (LSTM) network**. These features from both branches were then concatenated for regressing the intersection point, with the network being trained end-to-end using the mean square error loss.

"*Since it's important to report the errors in 3D and millimeters, we also recorded the ground truth depth data, just for evaluation, for all frames,*" Baoru adds. "*We used a custom-developed structured lighting system and the corresponding algorithms developed by us. With 23 different patterns projected onto one frame, we can get the depth map for this frame. We've released this dataset to the community.*"

Overall, what makes this work truly special is its **innovative use of the gamma probe to detect gamma signals and locate cancerous tissue, enhancing the accuracy of resection and diagnosis**. Moreover, its ability to

transform a 3D problem into a 2D one without acquiring highly accurate ground truth depth data or precise probe pose estimation sets a new benchmark in the field. The simplified network design allows for real-time application during minimally invasive surgery, achieving an impressive inference time of **50 frames per second**.

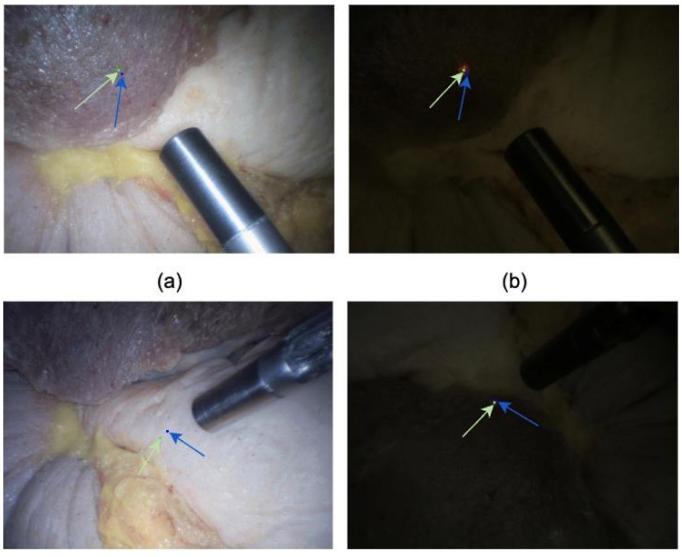


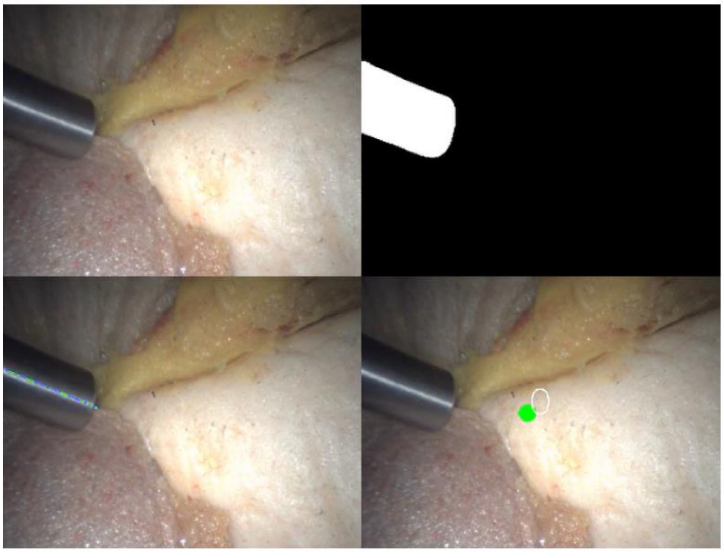
Originally from China, Baoru has been in the UK for eight years and completed her bachelor's degree, master's degree, and PhD there.

"*I really enjoy it – to be honest, I like the weather!*" she laughs.

Finally, we cannot let Baoru go without asking her about her experiences working with Matina Giannarou, her mentor and now second supervisor at the Hamlyn Centre, who is also a good friend of this magazine.

"*Matina has lots of enthusiasm for research,*" she reveals. "*As a female researcher, she gave me tips on balancing research and life and grabbing every chance you can. She's been the Winter School chair for many years, and then in 2020, she asked me to be the mentor, and since 2021, she's asked me to be co-chair of the Hamlyn Winter School. She's really encouraged me.*"





BEST OF MICCAI 2023

M&M: Tackling False Positives in Mammography with a Multi-view and Multi-instance Learning Sparse Detector

Yen Nhi Truong Vu (left) and Dan Guo (right) are Research Scientists at Whiterabbit.ai working under the Director of Research, Thomas Paul Matthews.

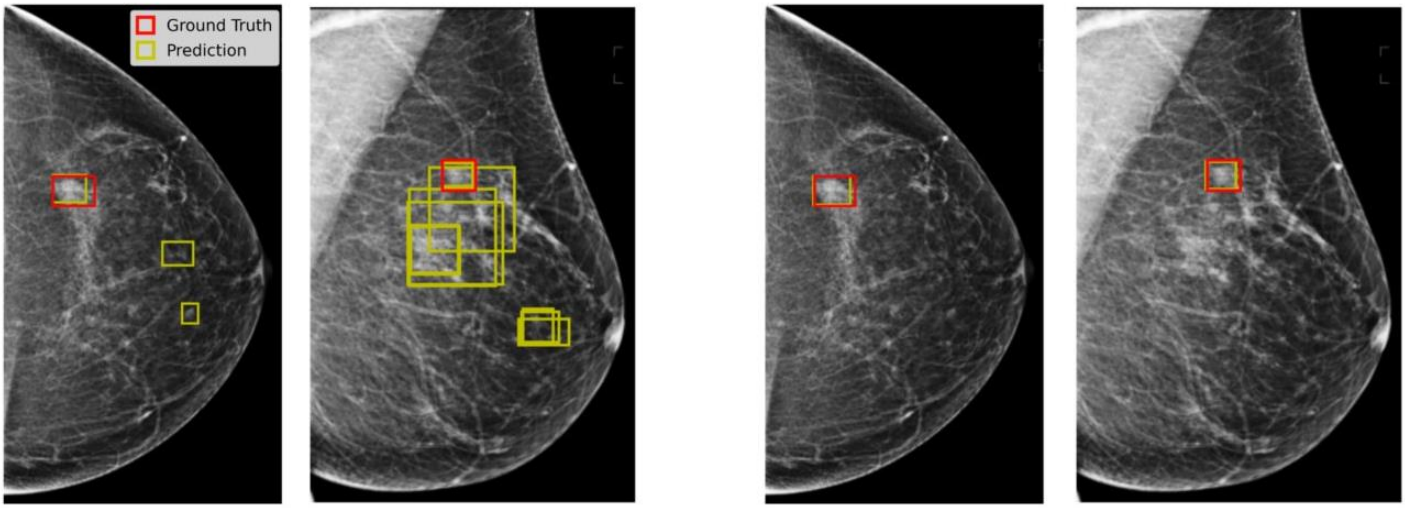
Their paper proposes a system to reduce the number of false positives for breast cancer screening.

They speak to us ahead of their poster presentation at MICCAI 2023.

Breast cancer screening is an important aspect of healthcare for women aged 40-75. While it is an invaluable tool, the prevalence of cancer in this age group is relatively low, with only about five out of every 1,000 women being diagnosed. However, one of the challenges of breast cancer screening lies in the occurrence of **false positives**, which can lead to anxiety and unnecessary medical procedures for patients. The key to improving the screening process is addressing these false positives, and researchers Nhi and Dan are turning to **computer-aided detection (CAD) software** to assist radiologists in this mission.

"*When a woman goes into screening, they will take four images, two images of each breast, and usually only one breast will be cancerous,*" Nhi explains. "*If CAD software produced one false positive per image, radiologists must* ***dismiss 400 false positive marks for every meaningful cancer detected****. That's a huge workload. It's a lot of resources wasted. Women have to go through a lot of extra procedures, anxiety, and in some cases, even biopsy, which is a very serious procedure.*"

The primary objective is to develop **high-sensitivity CAD software** while maintaining a low false positive rate. Achieving this goal is far from straightforward, as medical imaging



Qualitative Evaluation. **Left:** A dense detector (Cascade R-CNN) produces many false positives at 90% recall. **Right: At the same recall, M&M produces significantly less false positives.**

presents unique challenges compared to traditional natural image processing. Unlike natural images, where multiple objects are present, mammograms often contain **only a single focal point**, even in cancerous cases. Consequently, applying models designed for natural images to medical imaging is not advisable.

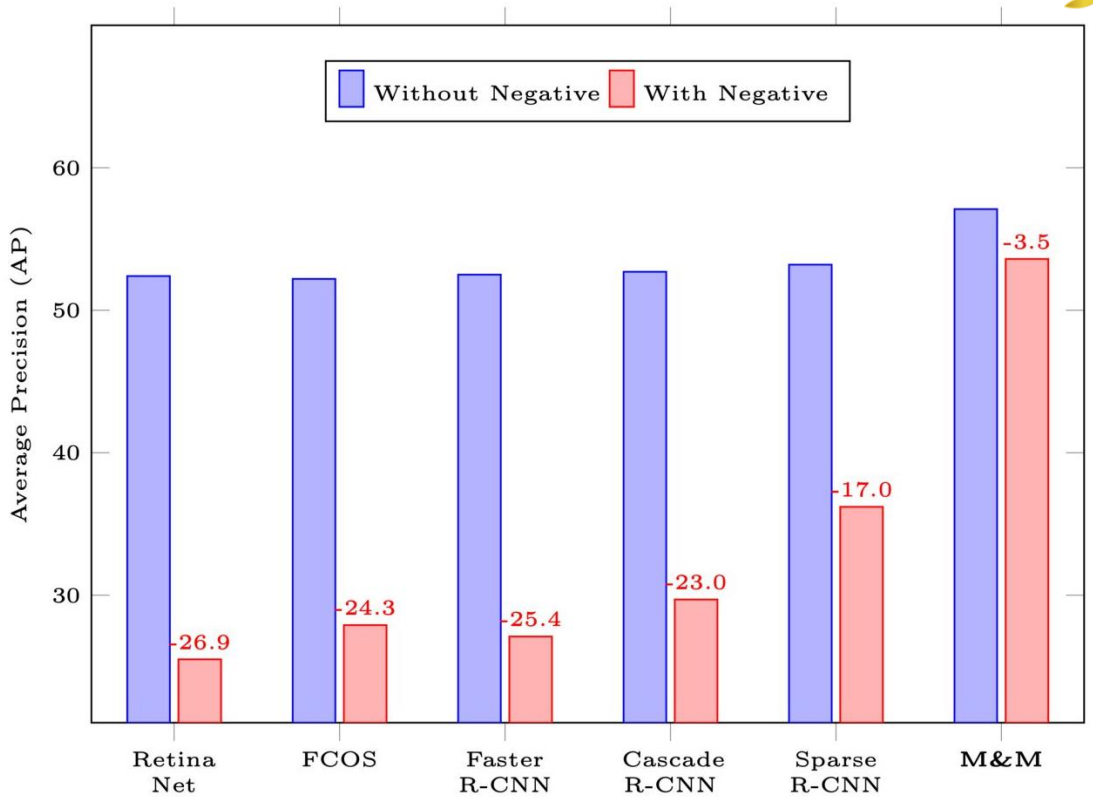
"*If you take a model designed for natural images, every image has an object,*" Nhi continues. "*Every image can be used to train the model. In medical imaging, many images are negative, so there's no object at all. How can you use these images to train the model?*"

Also, breast cancer screening typically involves capturing two images of each breast, providing complementary views of potential findings. A finding can look suspicious on one view but normal on the other. These views must be considered together to make accurate decisions.

"*Our director, the last author, is very interested in multi-view reasoning, so he came up with this idea of* ***working with two different views****,*" Nhi recalls. "*To make multi-view reasoning work, we must limit how many boxes or proposals we have on each view. That's when I started to look into the sparsity. At this point, Dan joined and did a lot of work on* ***multi-instance learning****.*"

As each malignant image typically contains only one finding, M&M uses **a sparse detector** with a set of sparse proposals, limiting the number of false positives. Dense detectors, where you have many dense proposals and anchors, tend to generate many false positives. Furthermore, introducing a multi-view attention module facilitates reasoning between the two views, enhancing the accuracy of detection, and multi-instance learning allows the team to harness **seven times**

BEST OF MICCAI 2023



antitative evaluation on OPTIMAM. Dense models produce too many false positives) on negative images: they suffer large average precision (AP) gap when evaluated h versus without negative images. Sparsity of proposals (Sparse R-CNN) reduces FP, s narrows the AP gap. By adding Multi-view reasoning and Multi-instance Learning, M closes this gap, i.e. **M&M produces much less FP on negative images.**

more negative images for training, a substantial improvement in the model's robustness.

While the team was optimistic about finding solutions to these challenges, it was a journey that required time and dedication. Dan, who is from China, initially joined the project as an intern. She returned the following year, and together with Nhi, who hails from Ho Chi Minh City in Vietnam, they refined their methods.

"*I think for all of these solutions, intuitively, we expected those methods to work, but how well it works was a somewhat unexpected component,*" Dan reveals. "*Overall, the model's performance was surprisingly good at the end.*"

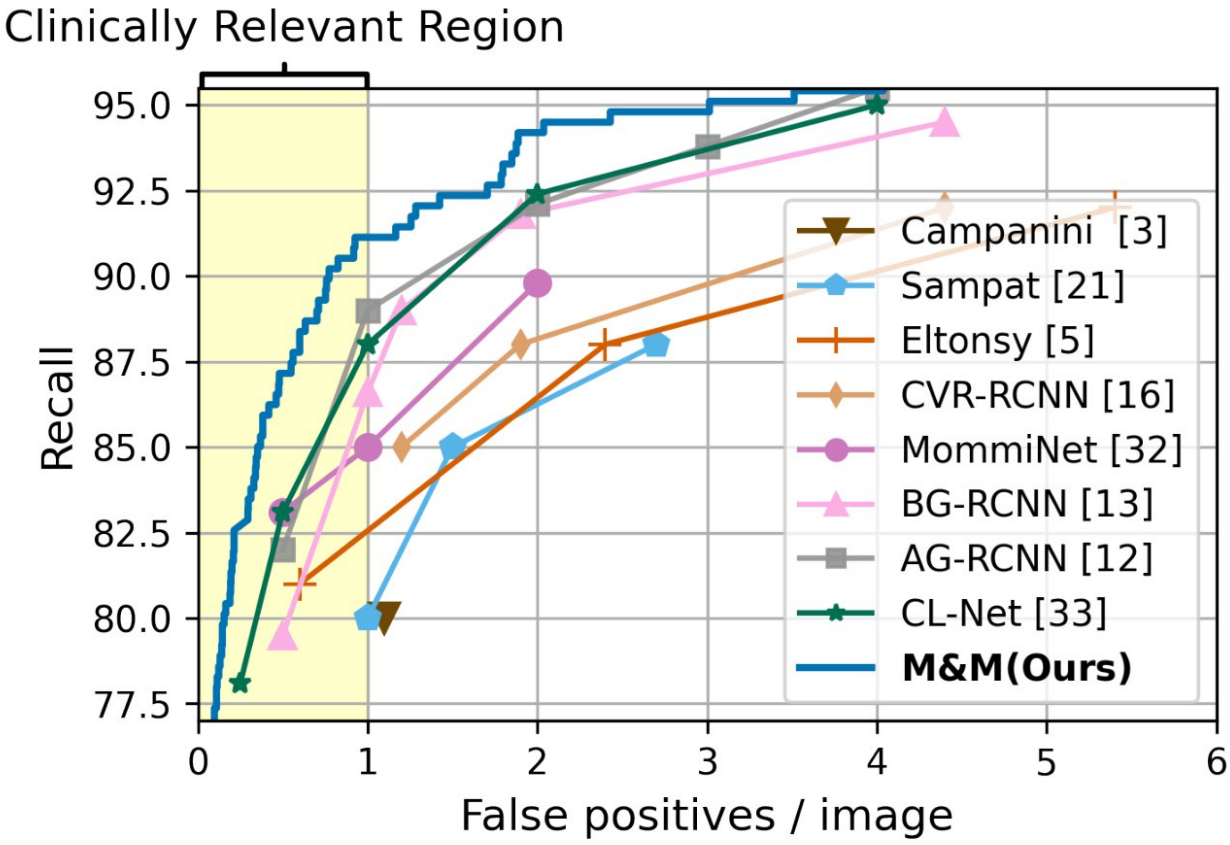
Nhi and Dan are working to bring their innovative solution into the real world. They have submitted their product to the Food and Drug Administration (FDA) for approval, a crucial step in ensuring its safety and effectiveness. Their optimism stems from pilot studies and positive feedback from radiologists, who have long grappled with the issue of false positives from CAD systems, which slow down the screening process and introduce unnecessary complexity. The team's model holds the potential to alleviate this burden significantly.

"We want to achieve something that, even at a very low false positive, like 0.1 or 0.2 false positives per image, gets a high recall," Nhi tells us. *"We've made progress in this part. If you look at the figure below,* ***our blue curve is quite a bit better than the others****. All the curves get close to each other as you go to three or four false positives per image, but that's not a point you want your software to operate at. We care about the part between zero and one, highlighted in yellow, and you can see that our curve is better than our competitors. It must be around 5% higher in terms of recall at 0.5 or 0.2 false positives."*

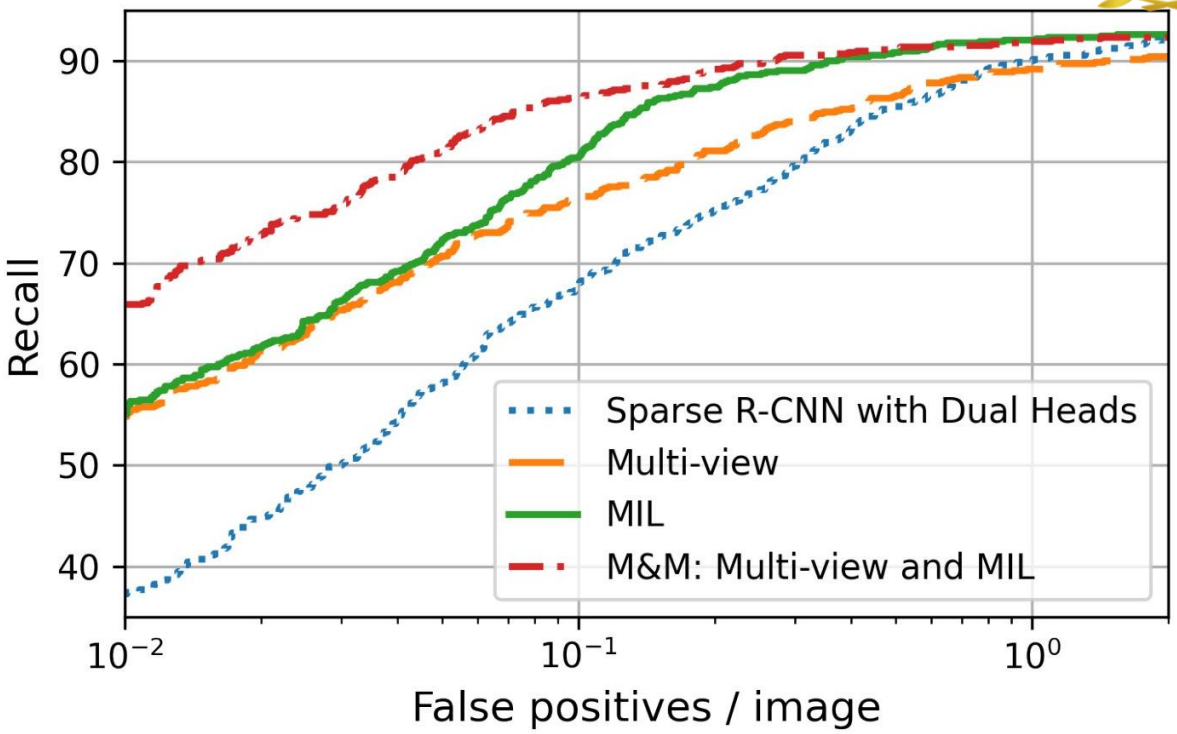
We have to ask about the origin of the model's title: **M&M**. Where could the inspiration for that possibly have come from, we wonder?

"Nhi came up with the title M&M, which makes me think of M&Ms, the chocolates!" Dan laughs.

Whiterabbit was founded in 2017 with a unique approach to early-stage breast cancer detection through mammography image data. Unlike traditional AI companies solely focused on algorithms and data, Whiterabbit initially owned and operated nine radiology clinics. This direct engagement with the clinical aspect of breast cancer



Quantitative evaluation of M&M on DDSM (4056 breasts). We compare M&M with recent literature on mammography object detection. **M&M consistently outperforms previous works, especially at low FP/image.**



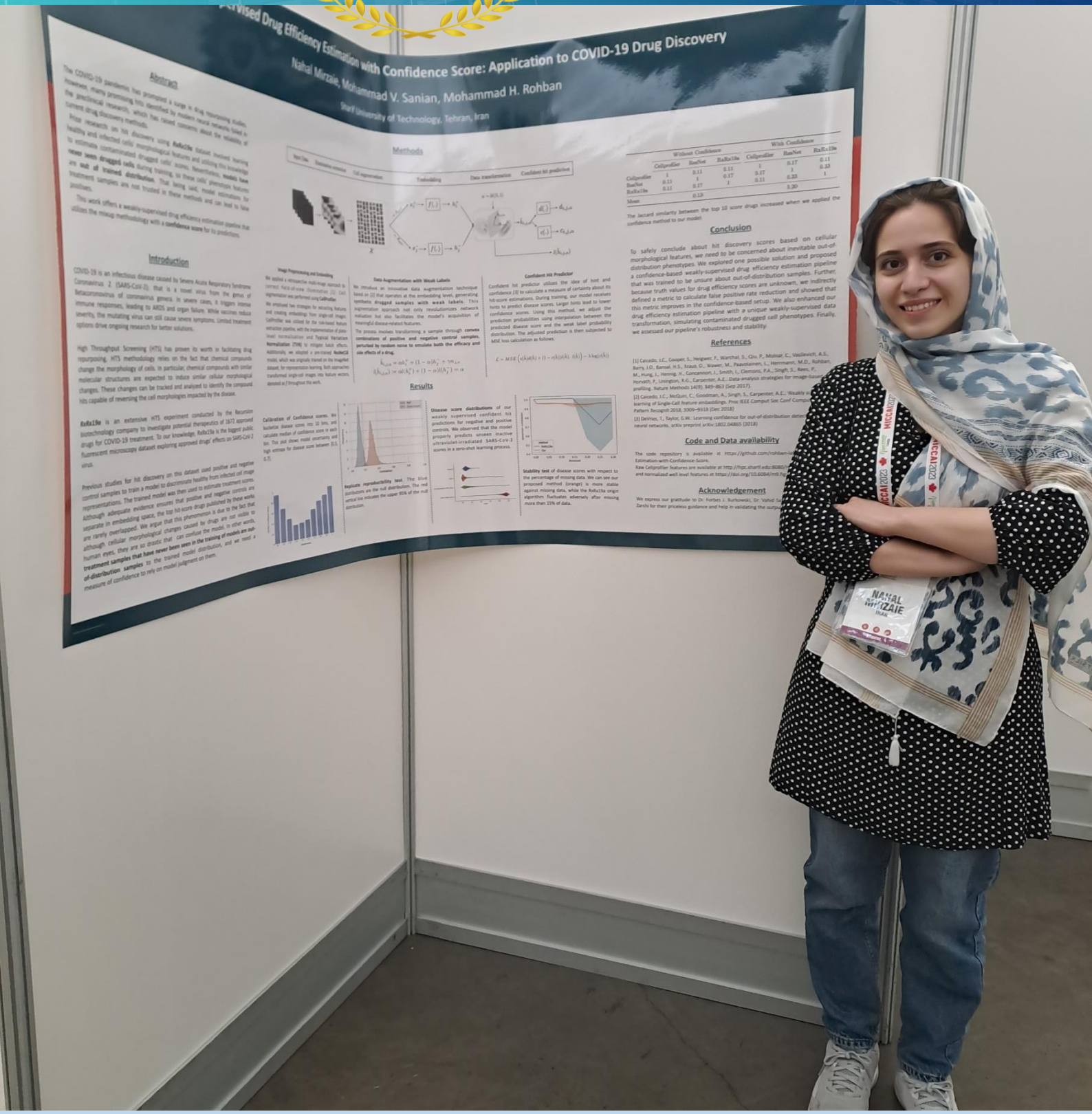
Quantitative evaluation of M&M's components. Without using any extra training samples, **Multi-view reasoning** improves recall at 0.1 FP/image (R@0.1) by 8.6%. By leveraging **Multi-instance learning,** M&M can train with images without bounding boxes. Thus, M&M sees 7x more images, improving R@0.1 by 9.9%. **Overall, M&M improves R@0.1 by 21.2% over vanilla Sparse RCNN.**

detection allowed the founders to forge close relationships with radiologists, staff, and patients. It cemented their desire to make products that have a meaningful impact on clinical outcomes.

"*We want to come up with products that help detect breast cancer earlier,*" Nhi says. "*This doesn't just come from something like our paper, which looks at the mammogram and detects the cancer, but we also try to improve the experience. It's not just about accuracy. You can say your model has very high sensitivity, but what if it creates a lot of false positives? As a company, we're very careful about the different aspects and consequences related to mammography. I feel lucky to be surrounded by people like Thomas and our CTO, Jason Su. They care about the problem and the clinical outcome as opposed to just making the biggest, most data-efficient model.*"

In closing, the message from Nhi and Dan's research is clear. When addressing a meaningful problem like reducing false positives in breast cancer screening, it's essential to identify and tackle the underlying challenges.

"*Essentially, the key idea is to find aspects of medical imaging that may not exist in the usual natural image computer vision sphere and try to use them to tackle these very special aspects of mammography,*" Nhi adds. "*It's not like we can just bring something from detecting dogs and cats over and apply it to patients!*"



Nahal Mirzaie is a PhD student at Sharif University of Technology with major in AI.

"Our paper Weakly-supervised Drug Efficiency Estimation with Confidence Score Application to COVID-19 Drug Discovery has been accepted at MICCAI 2023 and marks the culmination of my work during my M.Sc. journey. Our research addresses a crucial issue related to the reliability of existing drug discovery based on cellular morphological features methods when there are out-of-distribution phenotypes. I extend my heartfelt gratitude to the MICCAI RISE committee, whose generous support made my in-person participation possible."

Many thanks to awesome Esther Puyol for the intro

BEST OF MICCAI 2023

CortexMorph: fast cortical thickness estimation via diffeomorphic registration using VoxelMorph

Richard McKinley is a Senior Researcher at Inselspital, the University Hospital of Bern.

His paper on cortical thickness has been accepted as a poster, and he spoke to us ahead of his presentation at MICCAI 2023.

Cortical thickness, the thickness or depth of a thin ribbon of gray matter surrounding the white matter in the cerebrum of the brain, has emerged as **a potential biomarker for a range of neurodegenerative diseases and psychiatric conditions**. One notable example is multiple sclerosis, where the rate at which the cortex thins provides essential information on whether a patient's disease is being well controlled.

There are several open-source tools available for **quantifying cortical thickness**. Although AI-based tools have started making their way into clinical settings for evaluating neurodegenerative diseases, they primarily focus on cortical volume rather than thickness. It's a subtle yet significant distinction, as cortical volume comprises two key components: surface area and thickness.

"These measures are typically used in large cohort studies," Richard says. *"For example, you can take a large cohort of epilepsy patients and identify that relative to matched healthy members of the population, particular regions of their brains tend to show a reduction in cortical thickness. These are often regions functionally linked to each other. We want to leverage these changes in thickness to determine at an earlier stage whether a patient will likely have one of these diseases."*

While these tools are primarily used for research purposes today, it is hoped that, in the future, they could be applied more reliably to individual patients to aid in early diagnosis and personalized treatment.

"These tools are not yet sufficiently accurate on an individual patient

level," he confirms. "*We can use them as a research tool to see, in general, epilepsy patients will have this particular pattern of atrophy or one pattern of atrophy in frontotemporal dementia and another in Alzheimer's. It allows us to study the different mechanisms by which these diseases occur.*"

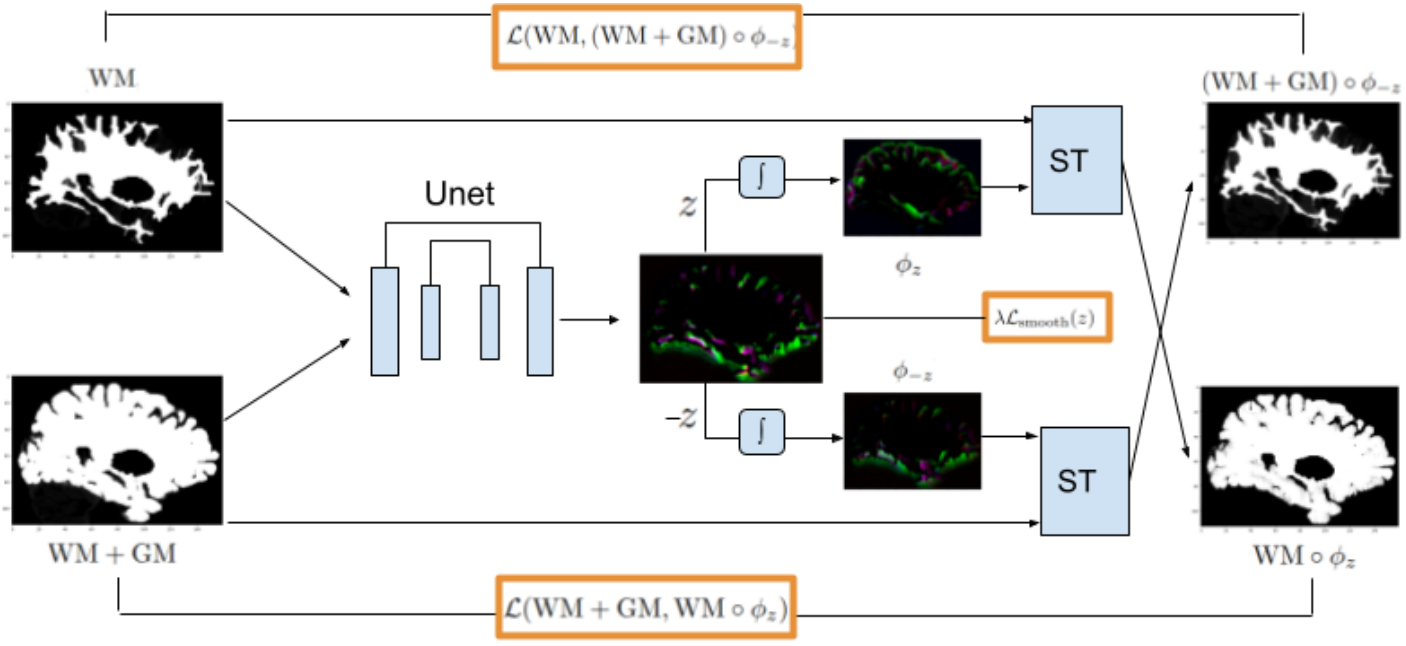
Richard adds that if cortical thickness analysis were applied to patients, it would be one element of a comprehensive diagnostic process considering various factors, including neurological tests and clinical assessments.

"*These methods only give us one piece of the puzzle,*" he asserts. "*If that piece doesn't fit with everything else, then you don't diagnose a patient with a particular disease.*"

Existing commercial tools, like brain imaging software **FreeSurfer**, provide volumes of the different lobes of the brain as biomarkers to radiologists. Though valuable, these markers are less sensitive to diseases than cortical thickness. Also, they take a significant amount of time – upwards of 8-10 hours – to process data for a single patient, presenting a clear challenge as clinical workflows demand quicker and more precise solutions.

"*Our goal initially was to produce something able to estimate cortical thickness, which was as reliable as the existing tools,*" Richard recalls. "*Now, we have evidence to suggest we have something more sensitive and reproducible while also* ***running in a matter of seconds!***"

He adopted an approach that is becoming increasingly important in the medical imaging field by reframing a problem previously solved using lengthy iterative algorithms. In doing this, he drew



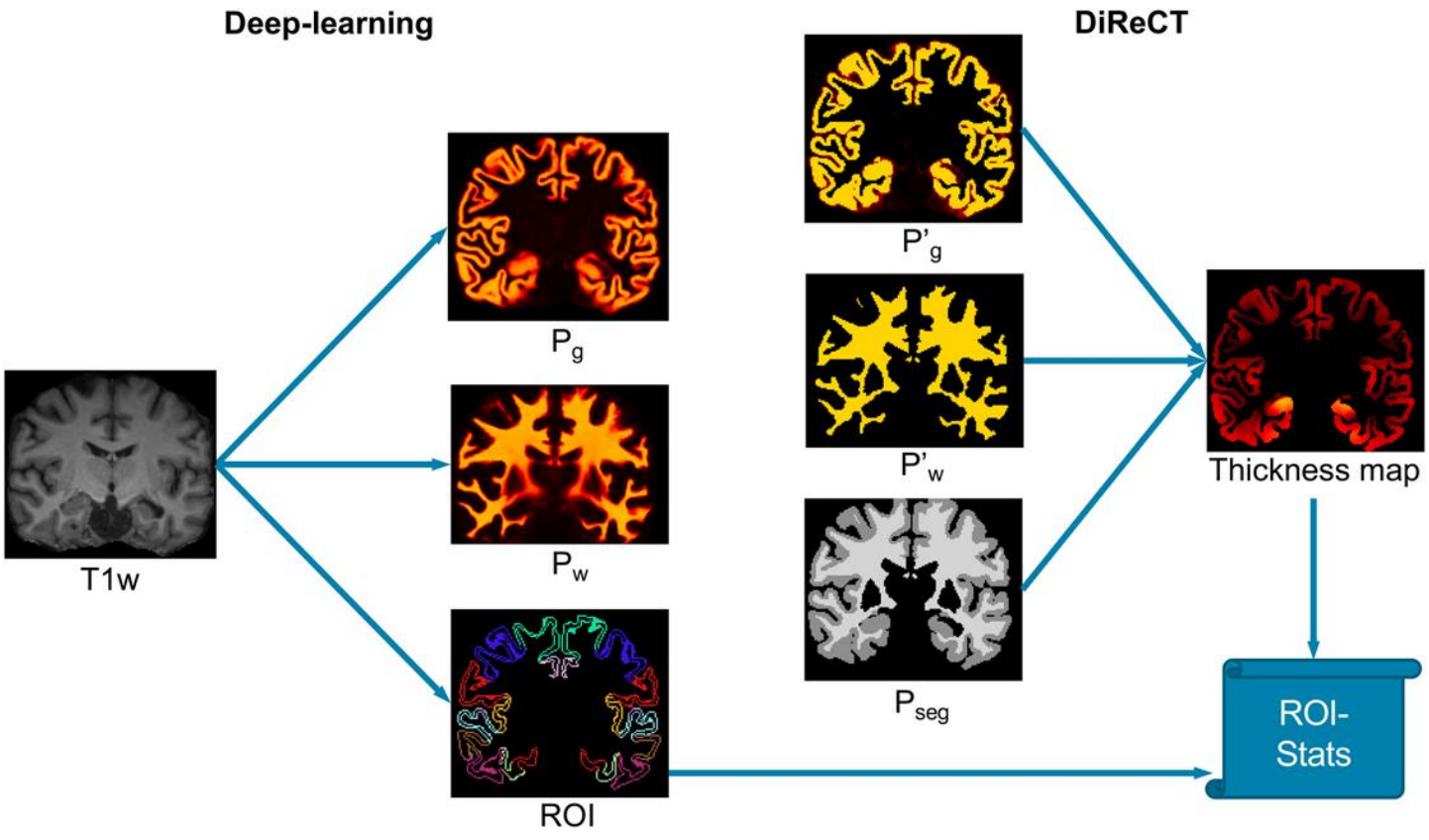
BEST OF MICCAI 2023

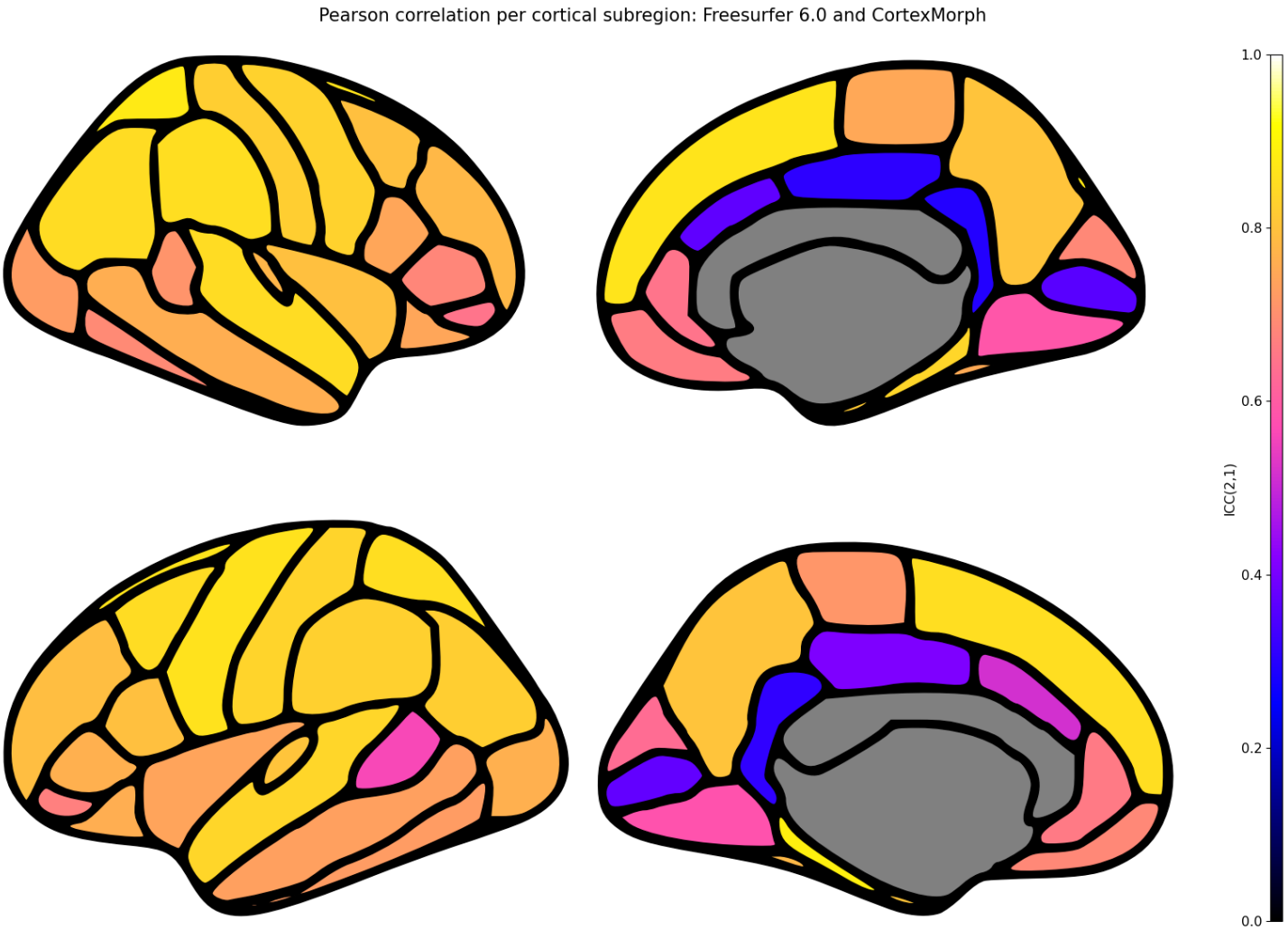
inspiration from **VoxelMorph** and its successor works, which took the problem of deformable image registration, **formulated it as the loss function of a deep neural network, and then trained the neural network to solve it**.

"*This, for me, is an excellent empirical but also elegant way to derive a solution to a problem,*" he says. "*You can have 20 pages of excellent theory explaining why your iterative algorithm works and converges well, but in the end, **the proof of the pudding is in the validation**. You don't start from the principles of trying to solve your problem; you define what the parameters are for your problem to have been solved, and then, using **gradient descent**, you search for a solution.*"

Interestingly, Richard built the prototype for this method several years ago, but when you build a system for calculating cortical thickness, how do you validate it if there is no ground truth? Luckily, a paper by **Rusak et al.** last year had the answer: **a synthetic phantom built using a GAN**.

"*They generated 20 subjects with different levels of cortical thickness reduction and showed that a predecessor method to ours was very sensitive to these reductions,*" he tells us. "*It's more sensitive than FreeSurfer. This work gave me the final piece of the puzzle. Now, I have a dataset coming from an independent group. It's one thing to say my deep learning approach is close to the existing approach, but can it do the same job of resolving*





these cortical thickness differences? Indeed, it does. In fact, it is even slightly more sensitive."

Did he get the chance to talk to Filip Rusak about his work?

"*Yes, a little bit,*" he reveals. "*After this paper was published, I was invited to be the examiner of his PhD thesis, which was very nice.*"

In terms of the next steps for this work, Richard says there are three key directions he hopes to pursue. First and foremost is **validation** and determining whether this method offers a similar or even superior ability to distinguish between different diseases compared to existing methods. Christian Rummell, the other author on this paper, has a currently funded project with the **Swiss National Science Foundation** to build **an open-source suite of neuroimaging tools**, allowing people to apply methods like **CortexMorph** in their own studies in their own hospitals.

"*We want to integrate these tools into this platform so people can use this method for free and, most importantly, easily,*" he adds. "*Before we do that, we need to validate*

BEST OF MICCAI 2023

whether it really works – not just on healthy controls, but also on the ability to distinguish between diseases."

The folding of the human brain poses another significant challenge to surmount in accurately measuring cortical thickness. **The brain sort of folds in on itself, and from the perspective of MRI, its surfaces virtually touch**. From the segmentation, you effectively do not see a difference between the two banks of the sulcus, which present as a thick mass of gray matter.

"We believe, but have to validate, that our method gives a better resolution

"*FreeSurfer does a decent job of resolving these sulci, but often it doesn't correctly go all the way down into the sulcus, and so you have an incorrect identification of cortical thickness,*" Richard explains. "*We believe, but have to validate, that our method gives a better resolution of these sulci. If it doesn't, we have some ideas by going to* ***super-resolution****, for example, by* ***leveraging high-field MRI*** *to give us some basis for training our models to resolve these sulci better.*"

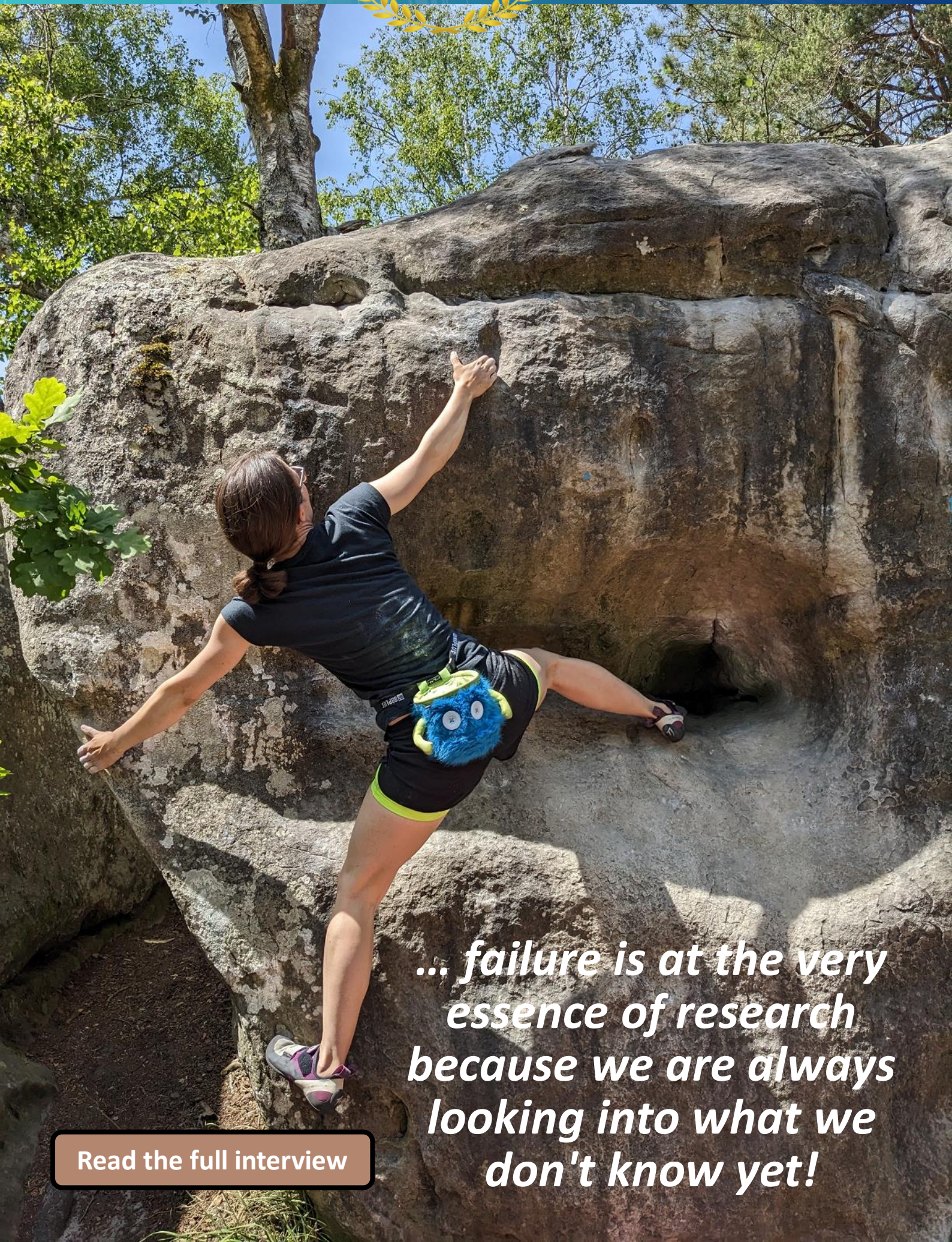
Finally, most other methods give **a point estimate of cortical thickness**, leaving clinicians unaware of the potential margin of error. Another next step is to produce a distribution of **plausible cortical thickness values**, or error bars, using ensemble methods, allowing a better understanding of measurement uncertainty and identifying regions of the brain where measurements may be less reliable.

At **Inselspital**, Richard works in a group called the **Support Center for Advanced Neuroimaging**, a multidisciplinary research group with MDs, physicists, computer scientists, and psychologists interested in interpreting and quantifying imaging of the human brain. Before we finish, he is keen to mention and appreciate the work of his recently graduated PhD student, **Michael Rebsamen**.

"*Michael worked together with me on developing DL+DiReCT, which is the foundation of this work, funded by a grant from the Novartis Research Foundation,*" he tells us. "*Without that foundational work, we wouldn't have been able to do this.*"

BEST OF MICCAI 2023

Lisa Koch



... failure is at the very essence of research because we are always looking into what we don't know yet!

Read the full interview

An experienced and skilled development team is paramount to ensuring efficient development while avoiding potential challenges and pitfalls throughout the project lifecycle. It often takes years to develop such a team, and a significant amount of time and effort is required for the team to transition to a new modality or task. As we will see, RSIP Vision offers the fittest solution to solve this challenge.

Let's briefly see what risks are involved when this challenge presents itself. First and foremost, **recruiting, onboarding, and training individuals with the right skillset and expertise can take very long**, and – more often than not, longer than expected. In fact, it might be overly challenging to attract the best candidates, engaging them and hiring them fast, as well as ensuring a positive candidate experience. That will be easier for a company with **a strong**

employer brand, but very difficult for newcomers in the AI world. In addition, forming a new development team requires creating ex nihilo an **efficient engineering recruitment process**.

Specific to the **Medical AI field**, the recruited engineers will probably have knowledge gaps in specific modalities or tasks: "*This risk may lead to potential wrong decisions in the development process,*" said **Ilya Kovler, CTO at RSIP Vision**. "*Bad decisions may lead a project astray: they generate steps which are a long way from the intended mark (or totally unneeded) as well as dire inefficiencies, resulting in lower quality outcomes and badly missed deadlines!*" This may happen even when task milestones were properly scheduled and planned by the project manager.

Working with RSIP Vision and its development team enables **full mitigation of these risks**. The company has a several decades' long experience in multiple projects, tasks, and modalities, which include an unmatched **expertise in AI for medical imaging**. Whatever the project involved, this rich experience includes very frequently projects with similar tasks, and modalities.

RSIP Vision's senior members of the engineering team constantly care to **transfer knowledge and mentorship** to the new recruits and junior engineers.

One of the crucial advantages of this care, which has always been a regular habit in the company, is that **collaborative tools, platforms, and solutions** are shared along the project team, which is not possible when all team members lack previous relevant experience.

Additional advantages of RSIP Vision's solid team are its **flexible team structure** - stemming out of its efficient size and experience – as well as **continued training and education**. "*We are very well connected to academia along both of our fields of expertise,*" remarks Ilya. "*Both **computer science and clinical experts** are always available to provide valid support during all the development steps.*"

RSIP Vision's project management keeps monitoring and optimizing the teams work, following the highest standards of **Continuous Improvement Process (CIP)**. You can mitigate all the risks and challenges of building a new development team. Talk about your project with our R&D experts.

Alaa Eldin Abdelaal

Stanford University

Alaa Eldin Abdelaal has been selected as "Pioneer of Medical Robotics" to present his work at the Data vs Model in Medical Robotics Workshop at upcoming IROS 2023, where two stellar doctoral / post-doctoral candidates will present their bodies of work. What follows is Alaa's Research Statement, which won him one of the two winning spots, the other being claimed by Shan Lin.

by Alaa Eldin Abdelaal

My research addresses the problem of assisting humans while performing complex tasks in human-robot collaborative environments.

In these environments, a human controls one or more robot manipulators in collaboration with one or more additional controllers (humans or autonomous agents). Humans can benefit from such assistance in complex operations in applications such as surgery and manufacturing. For example, many surgical operations involve several physicians and assistants to help the main surgeon. My research investigates how to assist humans in these environments using the following two approaches:

• Developing autonomous systems to perform the repetitive parts of the task using the additional manipulators, allowing the human to focus on the more demanding ones.

• Designing interfaces that facilitate the human control of his/her robot manipulators.

My research applies the above two approaches in surgical applications, motivated by the worldwide need to augment surgeons. For instance, there is an unmet need of additional 140 million procedures due to the shortage of surgeons worldwide. In addition, even in places with better concertation

of surgical expertise like the United States, medical errors are the third cause of death, right after cancer and heart diseases and well before stroke and car accidents, causing around 250,000 deaths every year. Approximately, 40% of these errors happen in the operating room and 50% of the resulting complications are avoidable.

The mission of my research is to address these problems by building systems and interfaces for human skill augmentation in robot-assisted surgery (RAS). My focus has been on two instances of the skill augmentation spectrum: (i) Surgical skill acquisition, where I design interfaces to improve the motor skill learning aspect of surgery, and (ii) Task execution, where I design autonomous systems to perform repetitive tasks, such as suturing, so that surgeons can focus on the more demanding ones.

The key insight of my research is that leveraging robot's unique capabilities, that are different from humans, can revolutionize human skill augmentation in RAS. This is unlike the related work in this area, that has the implicit assumption that RAS basically replicates open surgery, through smaller incisions, ignoring many of the robot's unique capabilities. For instance, existing RAS platforms only have one camera, imitating the single pair of eyes of the main human surgeon in open surgery. Moreover, autonomous systems in RAS perform tasks following the same steps as humans. In contrast, in my research, I identify some of the robot's unique capabilities and design systems and interfaces to leverage these capabilities to better augment humans in RAS.

One example of these unique capabilities is that robots can have more than one pair of eyes, unlike humans. To leverage this capability, we designed and tested a stereoscopic 6-DoF (Degrees of Freedom) surgical camera to provide an additional view of the surgical scene, in addition to the existing camera in the RAS platform. Moreover, we developed a visual-motor alignment method that enables the surgeon to easily control the surgical tools with respect to the view of the additional camera. We also developed a multi-camera, multi-view system, where the user can see two views of the surgical scene simultaneously, as a picture-in-picture view. One of the views comes from the original surgical camera and the other comes from our additional one. We showed that such system improves surgical training and skill assessment compared with the traditional single-view system.

Another example is that robots can have adjustable "interpupillary distance", unlike humans. To leverage this, our additional stereoscopic camera design is "side-firing". That is, the two vision sensors are placed on the side of the camera body. This is unlike the current "end-firing" surgical

camera design where the view is aligned with the endoscope axis and the vision sensors are placed at the distal end of the camera body. Our novel design allows us to change the (baseline) distance between the two vision sensors, without the need to increase the size of the incision where the camera is inserted into the patient's abdomen. Our results showed that increasing this baseline distance improves the depth perception based on the views from our camera, with a precision of 2.5 mm.

Moreover, robots, unlike humans, can focus on two (or more) different locations at the same time when performing a task autonomously. Such capability allows robots to perform multiple steps of the task in parallel. We leveraged the surgical robot's capability to move multiple arms in parallel to devise autonomous execution models that go beyond the humans' way of performing repetitive surgical tasks. In other words, we applied the "parallelism" concept to automate surgical tasks. This work challenges the dominant idea in autonomous robotic surgery literature that the best execution model to automate a surgical task is the one used by humans. We applied this idea to automate parts of the suturing task and showed that our work leads to an order of magnitude faster execution compared with the state-of-the-art methods.

Moving forward, I will research how automation would fit into (in the short term) and revolutionize (in the long term) the current surgical practice in RAS. For example, I am interested in developing criteria with which one can decide if a surgical task is a good candidate for automation or not. I am also interested in developing the algorithms and systems needed to automate such tasks. My work will leverage other robot's unique capabilities to design and deploy human-centered collaborative systems in RAS.

Shan Lin

University of California San Diego

Shan Lin, the other **Pioneer of Medical Robotics** selected by the workshop.

The J. Crayton Pruitt Family Department of Biomedical Engineering at the University of Florida has received the 2023 Biomedical Engineering Society's (BMES) Diversity Lecture Award. Since 2013, the representation of women faculty has increased from 2 to 15 (now comprising 52% of the faculty), while Black and Hispanic faculty have grown from 1 to 6 (now accounting for 21% of HMC faculty). Notably, 60% of UF BME faculty promoted to Associate Professor and 33% promoted to Full Professor have been women or historically marginalized communities (HMC).

Many thanks to awesome Ruogu Fang of the SMILE LAB for the news

The prestigious Latsis University prize was awarded to awesome Mara Graziani.

"To all future generations of PhD's,"* Mara said, *"my advice is to remain curious, seek innovation where needed, even when others at first may doubt its possibility. Pursue high quality work always, and stay true to your values and beliefs. It will bring you a long way!"

“First of all, as a member of the African research community, I would like to express my sincere gratitude to **MICCAI Society** for bringing **MICCAI 2024** to Africa for the very first time, and to the beautiful city of Marrakesh, Morocco.

“My name is **Noussair Lazrak**, and I’m born and raised in the beautiful country of Morocco where I did study and became a researcher. I have been following MICCAI with great interest for many years, and I have always been impressed and inspired by the quality of the research and the impact that MICCAI has had on the field.

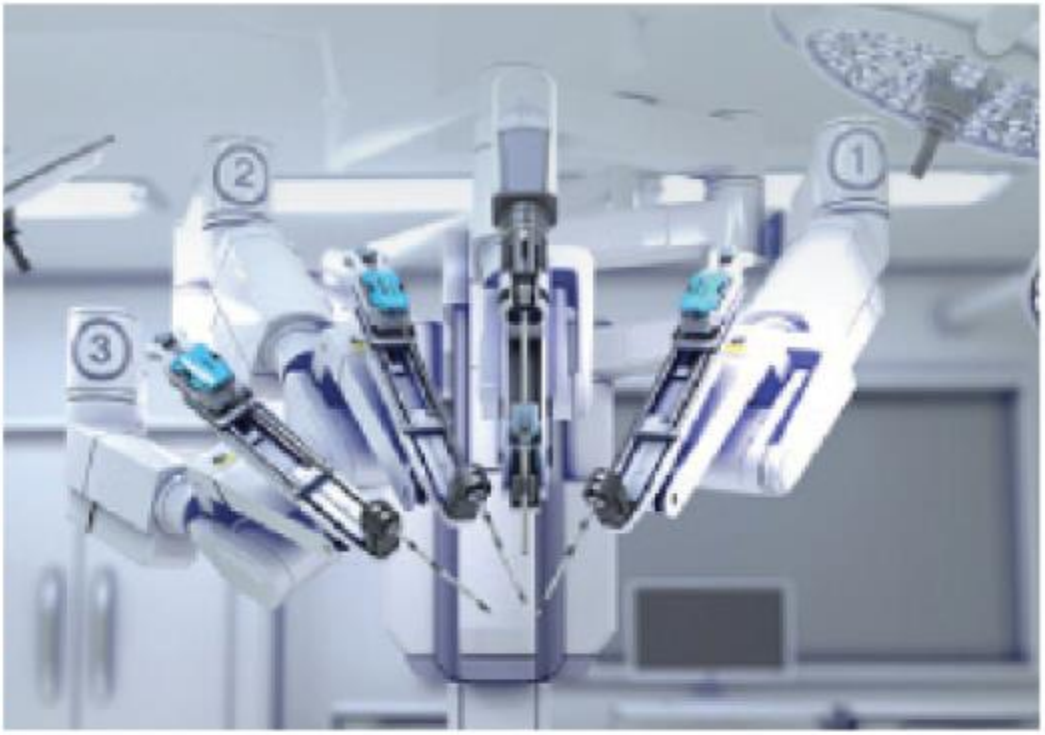
“It is true that (unfortunately) I have not been able to attend in the past, but I am thrilled (as many other African researchers) that MICCAI is finally coming to AFRICA. This is a unique opportunity to learn about the latest R&D in the field, and to network with leading researchers and practitioners from around the world. I am specifically excited about the fact that MICCAI is coming to Africa for the first time as it will help to raise the profile of AI in medical imaging in Africa and to attract more African researchers and practitioners to the field. It will also provide an opportunity for African researchers to showcase their work to the international community and to learn from the experiences of their colleagues from other parts of the world.

“This edition will have a tremendous impact not only on AI in medical imaging but on the whole African AI profile and lead to the development of new and improved medical imaging solutions that are tailored to the unique needs of African populations, and making AI more accessible and impactful for all.”

Welcome to Marrakesh, welcome to Africa!

Noussair Lazrak

MICCAI 2024 Marrakesh MOROCCO



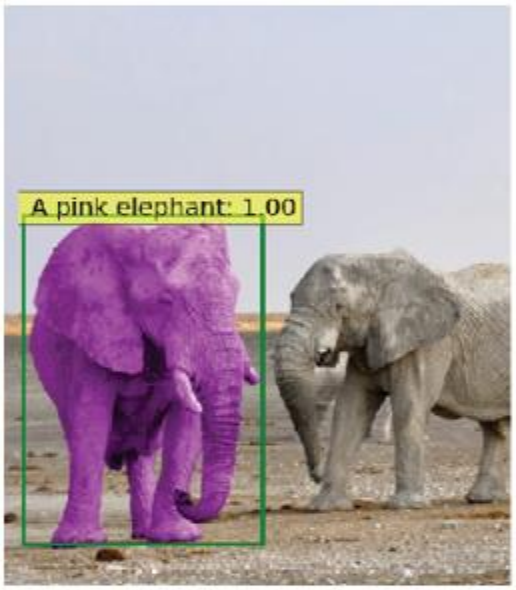
IMPROVE YOUR VISION WITH

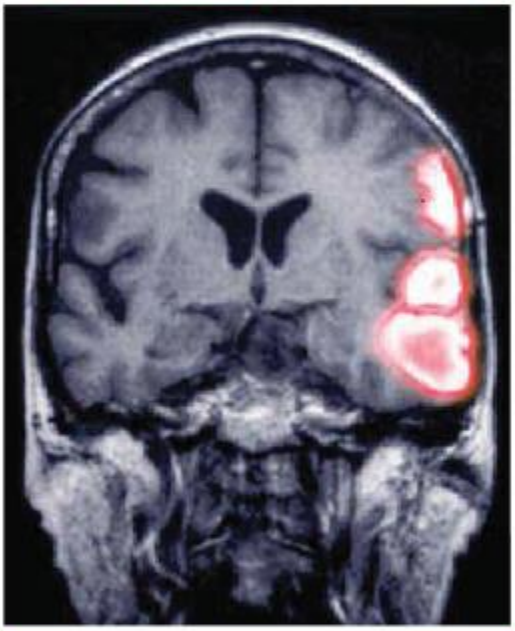
Computer Vision News

SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!





RSIP VISION