April 2024

Computer Vision News

& Medical Imaging News

The Magazine of the Algorithm Community

Woman in Science
Elisa Roccia p.22

Best of SPIE
Medical Imaging p.34

A publication by RSIP VISION

New Trends in Image Restoration and Enhancement

Radu Timofte is a Professor of Computer Science and a Humboldt Professor for AI and Computer Vision, leading the Computer Vision Laboratory at the University of Würzburg.

He is here to tell us about NTIRE2024, the ninth edition of an exciting image restoration and enhancement workshop he co-organizes at CVPR 2024.

The New Trends in Image Restoration and Enhancement (NTIRE) workshop is a meeting place for professionals working with image and video, focusing primarily on **low-level and mid-level vision**. "*We have to remind ourselves that computer vision itself is not possible without light and image sensors,*" Radu begins. "*We're working on the side of the processing pipeline at the heart of computer vision algorithms.*"

The workshop's core themes revolve around **image restoration, enhancement, and manipulation**, aiming to bring more detail to images and videos by restoring degraded content, filling in missing information, and transforming them **to achieve desired targets in terms of quality and performance**. The event spotlights new trends and concepts, such as super-resolution, and creates a platform for academic and industrial participants to

engage and collaborate.

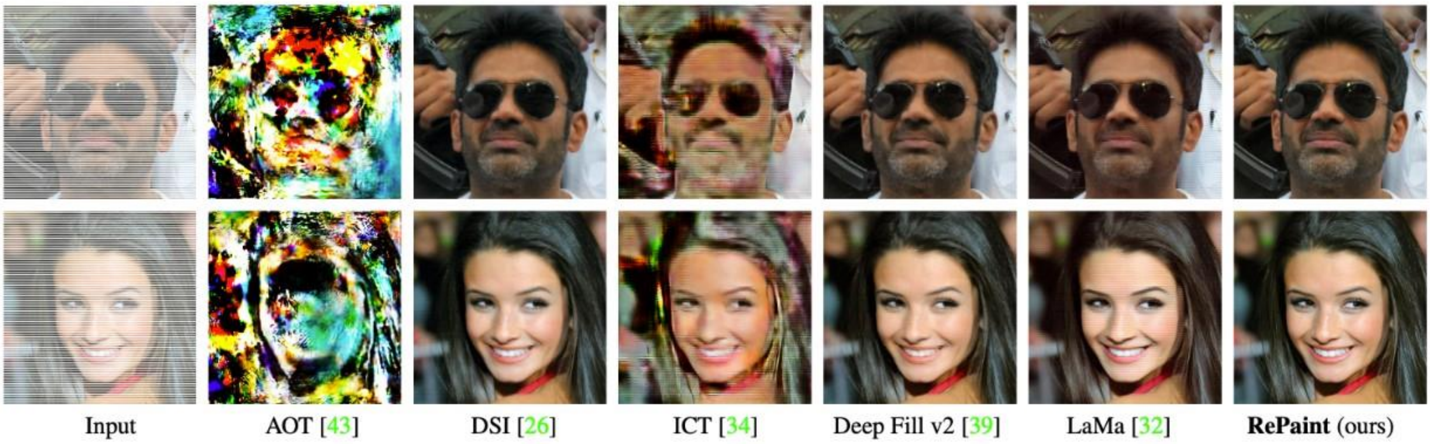
"NTIRE targets quite a large spectrum of users because everybody is working with image and video," Radu points out. *"Our workshop stays closer to topics related to low-level vision, but even researchers who work on mid-level or high-level vision would enjoy seeing what happens after the processing steps are conducted to enhance the quality of an image or video and bring it to a level of usefulness."*

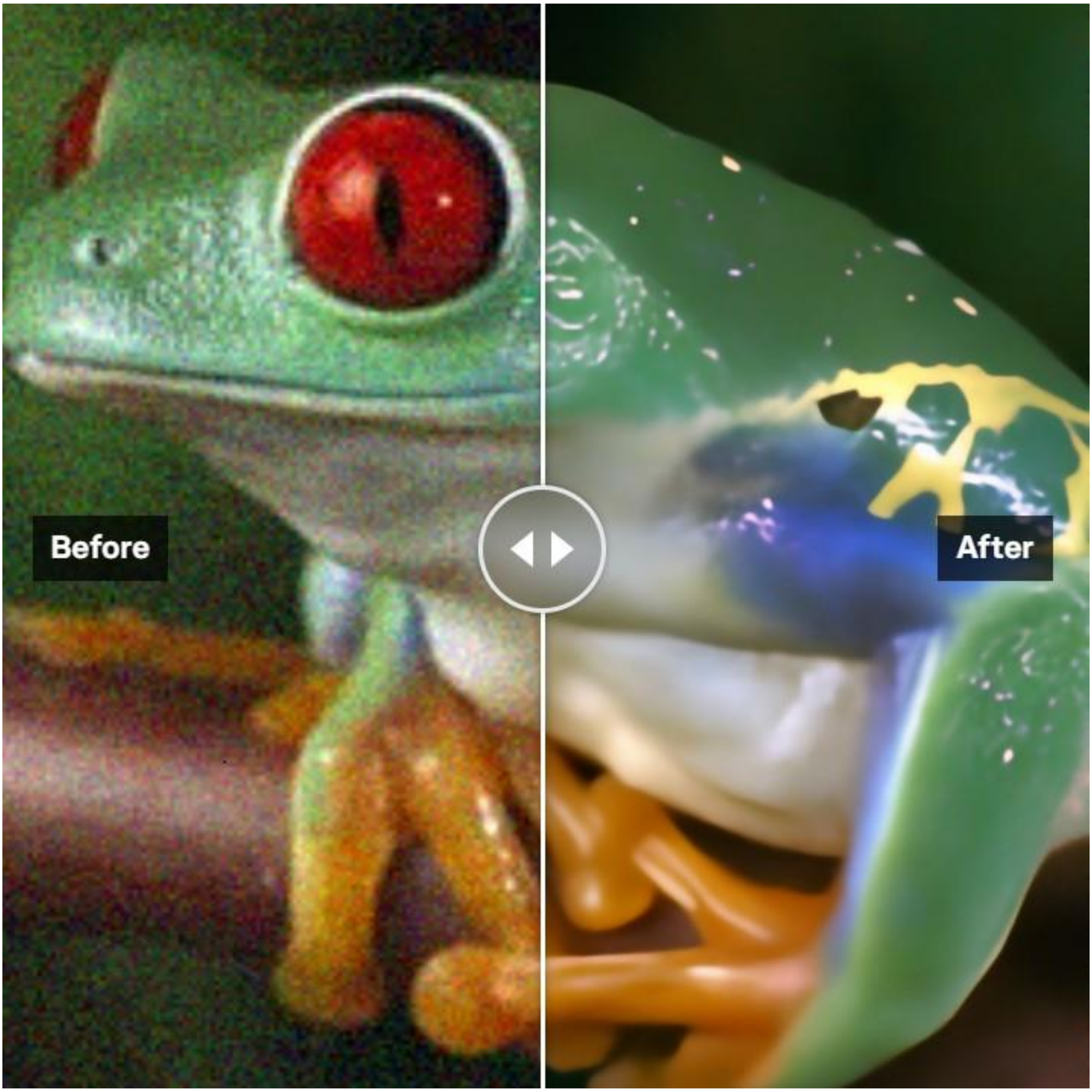
NTIRE's journey began in Taipei at ACCV 2016, but it has been a core part of CVPR since 2017, with notable professors like **Luc Van Gool from ETH Zurich** being part of its organization from the start. Diverse teams from different universities and the industry have been coming together annually ever since to tackle new and challenging topics and problems and compete in the workshop's associated challenges. This year, 17 challenges have been organized, each addressing different facets of low-level vision, restoration, and enhancement, attracting hundreds of solutions that gauge the state-of-the-art.

"It's been a fruitful and interesting journey," Radu recalls. *"I'm very grateful to my co-organizers and PC members. This year, we have around 30 organizers and 80 PC members, and they all have special duties, either in charge of the challenges or the reviewing process or just bringing interested industry or distinguished speakers to the workshop."* The team plans to invite five illustrious speakers, hoping to inspire new research directions, although their identities and topics will be confirmed in due course.

Reflecting on NTIRE's significance within the CVPR landscape, Radu highlights its history as one of the most extensive workshops in terms of participation and paper presentations. Each person attending will find something that aligns with their interests or curiosities. Participants can expect a

Extreme Case 1: Generate every second line



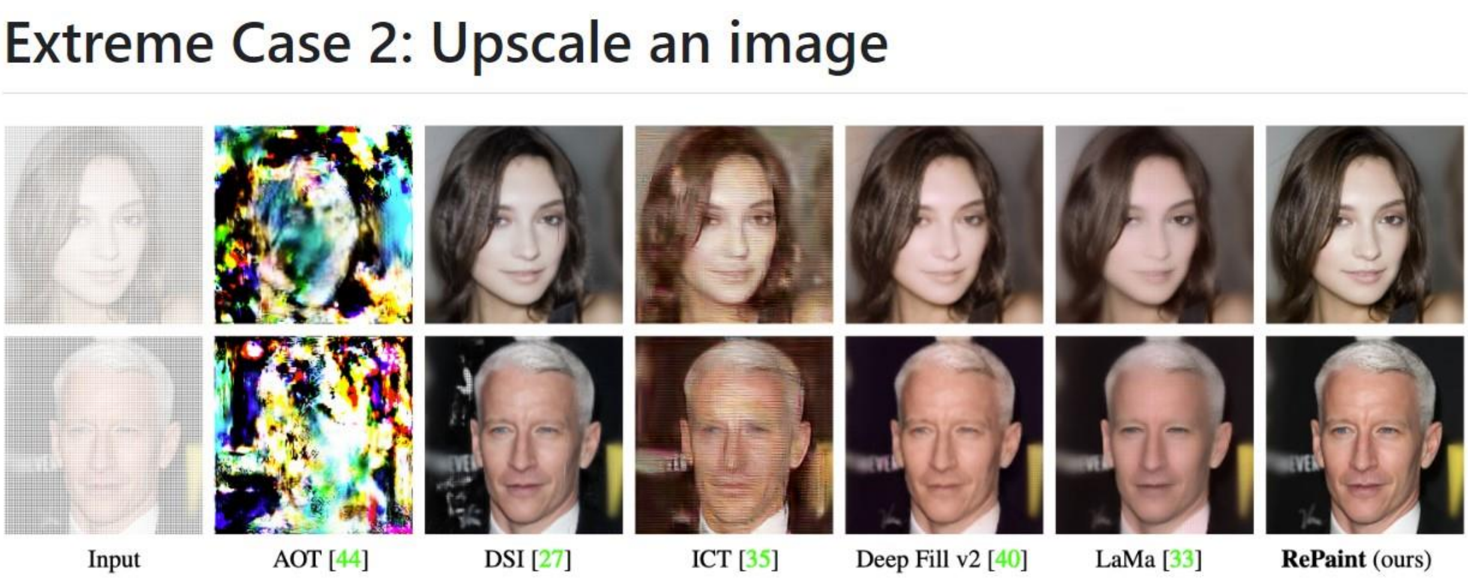


vibrant community eager to interact and contribute to ongoing advancements in the field.

Regarding what differentiates NTIRE from other workshops in this domain, Radu says it goes more into the details of the core problems in restoration and enhancement. However, there is **more cross-pollination than division**. *"When we look at the research, it boils down to components, and those components are typically shared,"* he explains. *"It's a very good thing that we're sharing the components to reach some new combinations and ideas across fields."*

In addition to NTIRE, his team is co-organizing two other workshops at CVPR 2024 in June. **AIS: Vision, Graphics, and AI for Streaming** will be the first workshop on AI for Streaming at CVPR, taking restoration and enhancement problems into the realm of real-time, optimized algorithms for streaming. Meanwhile, the fourth **Mobile AI workshop** will address specific restoration and enhancement challenges for mobile devices like smartphones, where algorithm deployment faces particular time and memory constraints in pursuing the best user experience.

Radu's lab's day-to-day work overlaps with the workshops, focusing on topics typically seen as low-level vision and augmented reality (AR) and tackling the underlying challenges of machine learning and AI that power these algorithms in the background. *"We're handling restoration and enhancement and manipulation of image and video, seeing them in the context of typical streaming scenarios like social media, AR and communications,"* he adds. *"We're very interested in placing these kinds of algorithms on mobile devices that are used everywhere."*



Vision-aided Screw Theory-based Inverse Kinematics Control of a Robot Arm Using Robot Operating System (ROS2) - Part 2


Women in Science

As we work harder, we become luckier!

Mahdieh (Madi) Babaiasl is the founder of the robotics startup, Mecharithm, where she is a researcher and robotics instructor.

More than 100 inspiring stories of Women in Science here!

Madi, can you tell us about your startup?

I started Mecharithm about six months ago. My mission is to revolutionize learning, teaching, and research in robotics. There are lots of news outlets out there, like CNN and the BBC, but there is no news outlet for robotics. I want the researchers to have a great resource, and I also want to revolutionize robotics learning. Now we are working on university-level robotics: how to make it easy to understand for professors and students, because robotics is hard for everybody. Instead of PowerPoints and lectures, we want to make it accessible, understandable, and practical for students.

Do you remember awesome Madi Babaiasl, who two years ago promptly taught me about redundancy in Robotics? A really fascinating moment!

We asked her to prepare a full lesson for our readers. What she did with three of her students (they are all PhD students and one of them is a soon-to-be professor) is a great "educational" lesson – so complete that we had to divide it in 3 parts. Part 1 was published in the March issue of Computer Vision News. Here is part 2.

Part 3 will be published in our May issue. You will find it in our archives during and after May 2024

by Madi Babaiasl, Bryan MacGavin, Daniel Montes Tolon, Namrata Roy

1 Introduction

In part 1 of this series of lessons, you learned how to set up your hardware and software to make everything ready to implement the vision-aided inverse kinematics control of the robot arm. In this lesson, you will learn about some theory behind screw theory-based numerical inverse kinematics using Newton-Raphson iterative method for robotic manipulators and thus you will be able to calculate the necessary parameters and equations to implement your robotic arm's numerical inverse kinematics. The equations for the robotic arm used in these series of lessons are provided for your reference.

2 Screw Theory-based Numerical Inverse Kinematics Using Newton-Raphson Iterative Method for Robotic Manipulators

Inverse kinematics of robotic arms is the problem of finding joint variables that produce a desired end-effector configuration. Formally, the inverse kinematics problem can be stated as follows:

Given an n degree of freedom open chain, with forward kinematics represented by the variable homogeneous transformation $T(q)$, where $q \in R^n$ are the joint variables, it is required to find a solution q that satisfies $T(q) = X$, where X is a desired (and given) homogeneous transformation and $X \in SE(3)$ (IK: given $X \in SE(3)$, find q such that $T(q) = X$). Fig. 1 shows the difference between forward kinematics and inverse kinematics.

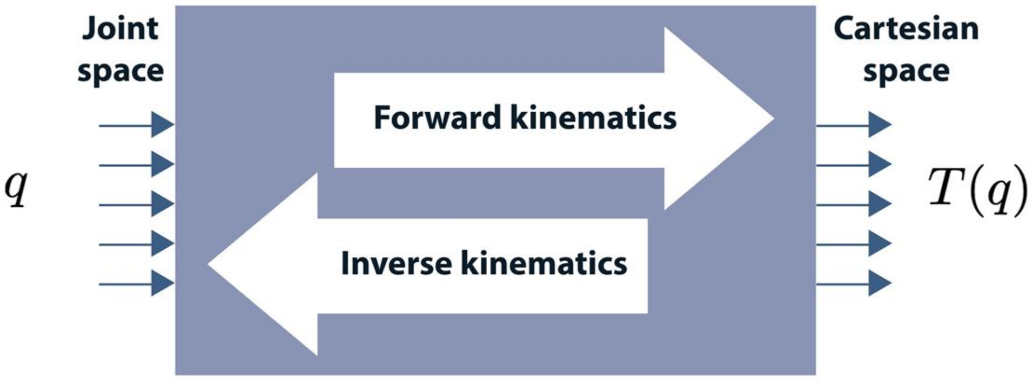


Figure 1: Forward Kinematics vs. Inverse Kinematics in Robotics

Inverse kinematics is a very important problem in the sense that here, we want to control the end-effector's configuration for it to be able to interact with the world. However, it is a more complicated problem to solve than the forward kinematics. In forward kinematics, we will have a unique end-effector configuration for a given set of joint values, but the inverse kinematics problem can have zero, one, or multiple solutions for the joint values q that can produce the given desired end-effector configuration.

There are two approaches to solving the inverse kinematics of an open-chain robot:

- Analytic approach to inverse kinematics in which closed-form solutions for the joint variables can be found. Here, geometric approaches to the problem are utilized. These analytic solutions may not always exist.

➢ Iterative numerical approach to inverse kinematics in which Jacobian is used to iteratively find a solution using the Newton-Raphson method. An initial guess for the solution should be made, and then this method iteratively pushes the initial guess towards a solution. This approach will only give us one solution and not all possible solutions.

If the inverse kinematics equations cannot be solved analytically, iterative numerical techniques may be utilized. Moreover, numerical methods are frequently employed to enhance the precision of analytic solutions, even when such solutions are available. In this case, the analytic solution can be used as an initial guess for the iterative numerical approach. For this purpose, we will use Newton–Raphson method, which is an integral method in nonlinear root finding. In situations where there is no exact solution exists, we will need optimization methods to find the closest approximate solution. If the manipulator is redundant, there will be infinite solutions to the inverse kinematics problem. In this case, we need to find a solution that is optimal with respect to some criterion.

Now, let's start by providing a general algorithm for numerical inverse kinematics using the Newton-Raphson method. Suppose that T_{sb} is the pose of the end-effector frame {b} in the base frame {s} (calculated from the forward kinematics) and the desired end-effector configuration is given by the transformation matrix T_{sd} as depicted in Fig. 2 below. Solving the inverse kinematics means that we need to find the set of joint angles that

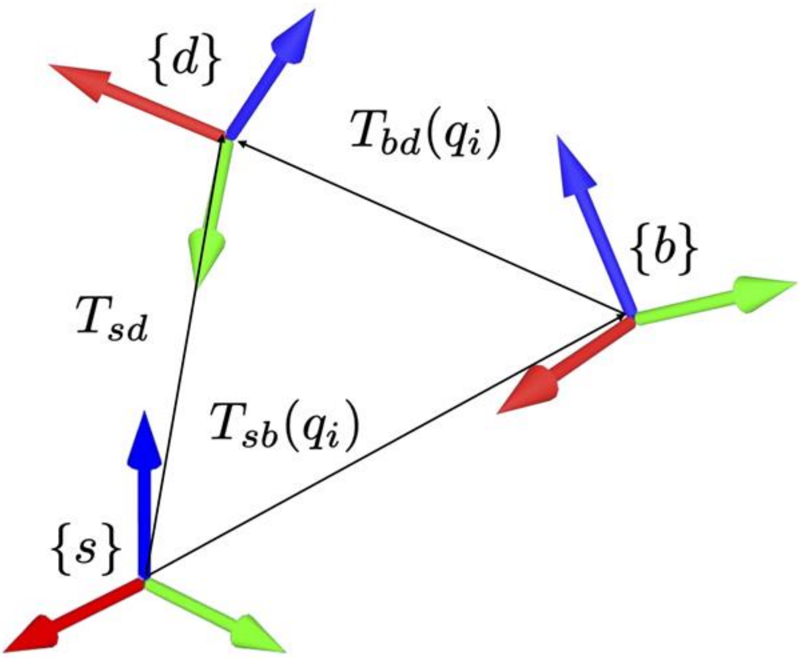


Figure 2: Illustration of the pose of the end-effector frame in the base frame, T_{sb}, and the desired end-effector configuration is given by the transformation matrix T_{sd}. Solving the inverse kinematics means that we need to find the set of joint angles that can take the end-effector frame {b} to the desired frame {d}.

take the end-effector frame {b} to the desired frame {d}. To learn about reference frame assignment in robotics and the homogeneous transformation matrices to represent pose in robotics, you refer to the orientation in robotics and pose in robotics lessons. To have a brief review of the Newton-Raphson method for nonlinear root finding, you can refer to this lesson.

To find the numerical screw theory-based inverse kinematics of open-chain manipulators, follow the following algorithm:

1. Initialization: Given T_{sd} (transformation of the desired frame with respect to space frame) and an initial guess of joint variables $q_o \in \mathbb{R}^n$, set $i = 0$

2. Set $[\mathcal{V}_b] = log(T_{sb}^{-1}(q_i)T_{sd})$. While the algorithm is not converged:

- Set $q_{i+1} = q_i + J_b^{\dagger}(q_i)\mathcal{V}_b$
- Increment i

Where q_i is the current joint angles guess, $\mathcal{V}_b$ is the body twist calculated from the error between the current end-effector configuration and the desired configuration, and $J_b^{\dagger}(q_i)$ is the pseudoinverse of the body Jacobian matrix evaluated at the current joint angles q_i . To learn about how to calculate the forward kinematics of robot manipulators, refer to this lesson, for Twists refer to this lesson, and to learn about screw theory based Jacobian matrix calculation, refer to this lesson.

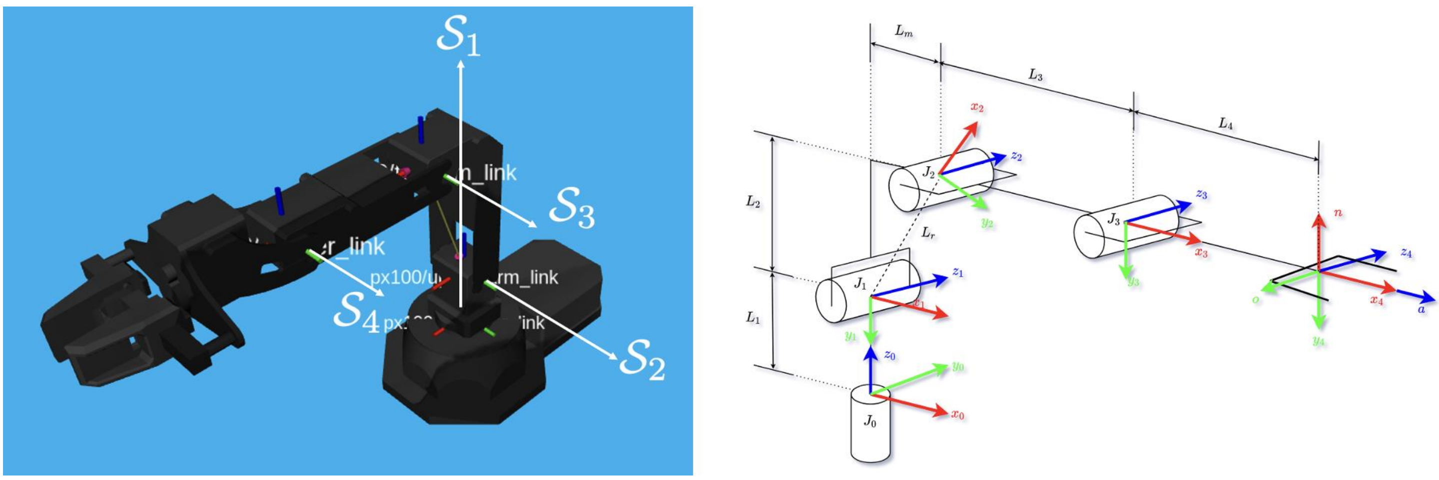


Figure 3: Robot arm depicted in home position with screw axes and link

lengths assignments that are essential to find the forward kinematics. The lengths depicted on the right photo are: L1 = 0.08945, L2 = 0.1, Lm = 0.035, L3 = 0.1, and L4 = 0.08605.

For our robot arm that we use for this lesson, the screw axis of each joint expressed in the base frame in the robot's zero position (see Fig. 3 for link lengths and screw axes assignments) are as follows:

$$\mathcal{S}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathcal{S}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -0.08945 \\ 0 \\ 0 \end{pmatrix}, \mathcal{S}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -0.18945 \\ 0 \\ 0.035 \end{pmatrix}, \mathcal{S}_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -0.18945 \\ 0 \\ 0.135 \end{pmatrix}$$

and thus, $T_{sb}(q)$ can be written as the following equation using the Product of Exponentials PoE formula:

$$T_{sb}(q) = e^{[\mathcal{S}_1]q_1} e^{[\mathcal{S}_2]q_2} e^{[\mathcal{S}_3]q_3} e^{[\mathcal{S}_4]q_4} M$$

Where, q1, q2, q3, and q4 are joint angles, and $M \in SE(3)$ is the end-effector configuration when the robot is at its zero position and for our robot arm it is represented by the following matrix:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0.22105 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.18945 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The Space Jacobian of this robot $J_s(q)$ can be expressed as:

$$J_s(q) = \begin{bmatrix} J_{s1} & J_{s2} & J_{s3} & J_{s4} \end{bmatrix}$$

Where J_{s1} J_{s2} J_{s3} and J_{s4} can be calculated using the following equations:

$$\begin{aligned} J_{s1} &= \mathcal{S}_1 \\ J_{s2} &= [Ad_{e^{[\mathcal{S}_1]q_1}}]\mathcal{S}_2 \\ J_{s3} &= [Ad_{e^{[\mathcal{S}_1]q_1} e^{[\mathcal{S}_2]q_2}}]\mathcal{S}_3 \\ J_{s4} &= [Ad_{e^{[\mathcal{S}_1]q_1} e^{[\mathcal{S}_2]q_2} e^{[\mathcal{S}_3]q_3}}]\mathcal{S}_4 \end{aligned}$$

in which, $\mathcal{S}_i$ is an expression for the screw axis describing the ith joint axis in terms of the fixed frame with the robot in its zero position given above, and $[Ad_{T_{i-1}}]\mathcal{S}_i$ is the screw axis describing the ith joint axis, but after it undergoes the rigid body displacement T_i instead of being at zero position. In other words, it is the Adjoint map of the screw axis for when the robot is no longer in zero position. Note that if $T = (R, p) \in SE(3)$ is a transformation matrix where R is the rotation matrix and p is the position vector, then its adjoint representation $[Ad_T]$ can be calculated by:

$$[Ad_T] = \begin{pmatrix} R & o \\ [p]R & R \end{pmatrix} \in \mathbb{R}^{6\times 6}$$

Note that the bracket notation [p] is the 3 × 3 skew-symmetric matrix representation of the position vector p. To learn more about the skew-symmetric matrix representation of a vector, you can refer to this lesson.
The body Jacobian can then be calculated from the space Jacobian using the adjoint transformation (to learn more about the adjoint transformation in robotics, refer to this lesson):

$$J_b(q) = [Ad_{T_{bs}}]J_s(q)$$

where $T_{bs} = T_{sb}^{-1}$ is the inverse of the homogeneous matrix T_{sb} that was derived from forward kinematics earlier.

3 Summary

Part 2 of the lesson series explored the theory behind screw theory-based numerical inverse kinematics for robotic manipulators, focusing on the Newton-Raphson iterative method. It also aimed to equip learners with the ability to calculate parameters and equations necessary for implementing a robot arm's numerical inverse kinematics. In the next part, we will start working on implementing the vision-aided numerical inverse kinematics control of the robot arm.

Computer Vision News is very grateful to Madi and her team for another awesome lesson in robotics!

References

1. Madi Babaiasl. [Modern Robotics Course]().
2. Peter Corke. *Robotics, Vision and Control: Fundamental Algorithms in Python*. Vol. 146. Springer Nature, 2023.
3. William Hoff. [Intro to Computer Vision]().
4. Peiyuan Jiang et al. "A Review of Yolo algorithm developments". In: *Procedia Computer Science* 199 (2022), pp. 1066–1073.
5. Kevin M Lynch and Frank C Park. *Modern robotics*. Cambridge University Press, 2017.
6. Richard M Murray, Zexiang Li, and S Shankar Sastry. *A mathematical introduction to robotic manipulation*. CRC press, 2017.
7. OpenCV. [OpenCV-Python Tutorials]().
8. Richard Szeliski. [Computer vision: algorithms and applications](). Springer Nature, 2022.
9. [Trossen Robotics]().

Computer Vision Book — Computer Vision: Algorithms and Applications — 17

COMPUTER VISION: ALGORITHMS AND APPLICATIONS

Richard Szeliski is an Affiliate Professor at the University of Washington, where he supervises students and teaches some lectures. He retired from Facebook just over a year ago, after a career spanning 40 years in computer vision research, but that has not stopped him. He speaks to us after publishing the second edition of his all-encompassing computer vision textbook, 'Computer Vision: Algorithms and Applications'.

Richard, how did you come to write the book?

In 2003, Steve Seitz invited me to co-teach a class at the University of Washington, which we called 'Computer Vision for Computer Graphics'. Several techniques based on computer vision and analyzing images could be used to create computer graphics effects or 3D models. In 2005, we taught a more general computer vision class, and I decided to convert all that knowledge we were sharing into a book. It finally came out in 2010. That first edition followed the traditional layering of low-level, mid-level, and high-level vision in those days, starting with basic things like projective geometry, then image processing and 2D and 3D vision, and finally recognition. That whole idea has been upended with deep learning. We still have different layers, but everything is done all at once in a deep neural network.

What made you decide to publish a second edition?

Deep learning has taken over as the preferred technique, and it wasn't mentioned at all in my book because, in 2010, it still hadn't proven its utility. Also, many techniques are so much better now, like structure from motion and recognition. I wanted to update the book to reflect that. I've always tried to track what people are teaching. I've listed classes that teach a curriculum similar to what I cover in the book on my website.

Do you expect in another 15 years you will need to write a third edition?

That's possible! We'll see what the next revolution is. There are some new techniques like transformers that seem to be doing well. It's a very dynamic field, so I'm sure there will be a need for new textbooks. Of course, even as I was finishing up this book, there were lots of recent papers that were very relevant and students should be reading, but I had to stop at some point. Although it's fascinating work, I wanted to get on and do other things!

A free PDF edition of the book is available on your website. How did that come about?

My primary motivation has always been to share information. There was a particular way of teaching the material that I developed with Steve Seitz, and we were both very excited about it. Universities like Carnegie Mellon would take our class notes and slide sets and build their classes around them. I wanted it to be as widely used as possible. I had to negotiate with publishers about the free copy, and most were worried about book sales. Fortunately, Springer said, *"Sure, we can do that!"* I was working with the same great editor, Wayne Wheeler at Springer, on the second edition and it was just a given that we would do it again.

You've had 16,000 downloads already this year!

I never got around to putting a counter up for the first edition, but so many students came up to me at computer vision conferences and said, *"Professor Szeliski, the reason I'm here is I read your book, and I got excited about computer vision!"* That was a wonderful feeling. When the second edition came out, I put the counter up, and it skyrocketed. There are more downloads now than people who attended CVPR!

You acknowledge the help of around 100 people in the book. What part did they all play?

I've posted drafts chapter by chapter for both editions and asked people to send me comments. I worked on the second edition for three years and had hundreds of comments, suggestions, and corrections. All those people listed in the book helped me. I even reached out to a few students to teach me some of the finer points of deep learning.

"There is *absolutely no way in hell* we will ever reach human-level AI without getting machines to learn from high-bandwidth sensory inputs, such as vision.

Yes, humans can get smart without vision, even pretty smart without vision and audition. But not without touch. Touch is pretty high bandwidth, too."

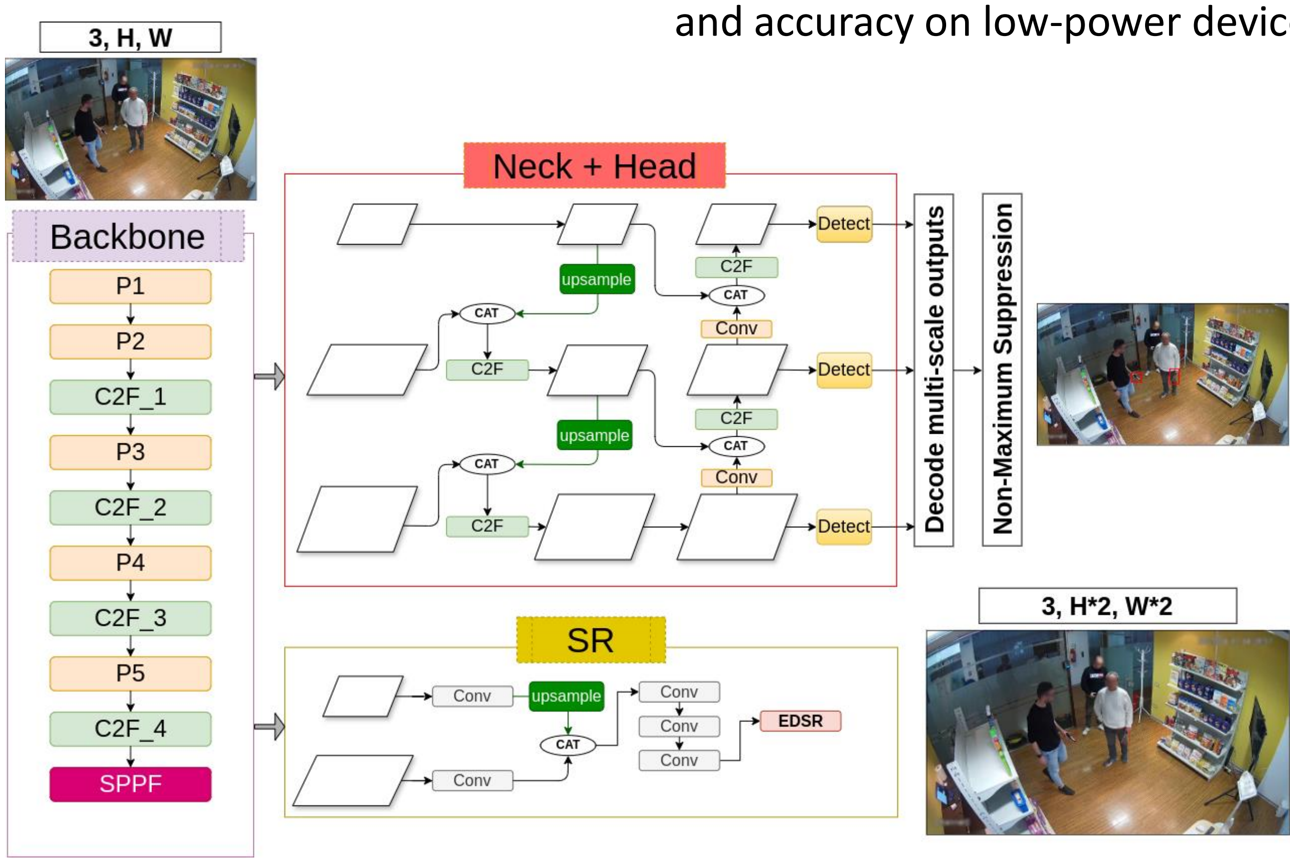
Daniele Berardini recently completed his PhD with the Vision Robotic and Artificial Intelligence (VRAI) group at Università Politecnica delle Marche, receiving his degree cum laude.

His thesis focused on developing Edge AI-enabled monitoring systems.

He is now a Postdoctoral Researcher, still at UNIVPM.

Congrats, Doctor Daniele!

The recent progress in Computer Vision is driven by data growth, computational advancements, and a scientific focus on innovative but increasingly complex AI algorithms like **Large Vision Models**. These algorithms require **huge computational power**, incurring high costs and energy usage, which raise environmental issues and restrict technology access and affordability. **Edge AI addresses these issues by providing decentralized AI solutions** suitable for low-cost edge devices, despite challenges in developing AI models balancing execution speed and accuracy on low-power devices.



My thesis aims to contribute to the **integration of Edge AI in Computer Vision** by focusing on the design and development of lightweight, deep learning-based monitoring systems for the real-time analysis of images and videos. It explores two domains, each with its own challenges: security, focusing on weapon detection in surveillance videos, and healthcare, focusing on segmenting preterm infants' limb poses from depth data.

Regarding the security domain, in collaboration with **INIM Electronics**, an Italian leader in security systems, I tackled weapon detection challenges, notably the difficulty in identifying small-sized weapons and the need for real-time weapon localization. Current solutions, like **image Super Resolution (SR)** methods or complex detection architectures, are inapplicable on edge devices due to computational constraints. To address this, after the creation of a **surveillance dataset for weapon detection (WeaponSense)**, I proposed the first Edge AI framework for real-time weapon detection in surveillance videos through the use of two cascaded CNNs, optimized for edge devices. Despite the results improving the state of the art, the framework shows limitations on efficiency in crowded environments. Thus, in collaboration with the University of Córdoba (Spain), I proposed **a novel method that integrates during training an Enhanced Deep Super Resolution network into an edge-oriented CNN for weapon detection**, discarding the former during inference. The proposed approach overcomes the previous limitations, enabling accurate and real-time on-device localization of weapons (Figure 1).

In the medical domain, my research was motivated by the need for automated technology to continuously **monitor the movement of preterm infants**, which is essential for early assessment of potential long-term complications. Current methods are effective but require very high operational costs, hindering their implementation in budget-limited facilities. Driven by these premises, I initially proposed a CNN that incorporates lightweight computational blocks from a segmentation network (EDANet) into the bi-branch structure of a preterm infants' pose segmentation network (BabyPoseNet). Subsequently, by conducting a per-layer complexity analysis of the proposed CNN, I redistributed computation throughout the network to reduce complexity. This approach yielded **a real-time framework capable of running on edge devices, achieving optimal accuracy in segmenting infants' limbs**.

Ultimately, my thesis aims to push research toward more sustainable and affordable AI solutions in different domains, valuing not only high accuracy, but also the **efficiency and adaptability across various needs and contexts**.

Simone Mentasti, (standing left) a researcher at Politecnico di Milano, has recently gained recognition as part of the winning AIRLab POLIMI team at the Robotic Drone Contest during the European Robotics Forum (ERF) held in Rimini last month.

He speaks to us about the competition and the innovative drone robotics that helped his team (with Matteo Molinari and Mattia Giurato) scoop the top prize.

The Robotic Drone Contest is a competition supported by **Leonardo**, an Italian aerospace, defense, and security company. Initially involving six universities in Italy, it has recently expanded its reach to include institutions from other countries.

Simone tells us that for the first three years, the contest focused on participants **building aerial drones and developing algorithms for navigation in new environments**. Last year, this focus expanded to **multi-agent cooperation**. *"In this scenario, we still have drones, but we also have a ground robot,"* he says. *"The goal is to develop algorithms that allow cooperation between these agents to explore a partially unknown environment and identify and track targets of interest."*

Winning the contest requires a strategic approach. Participants go through multiple rounds, aiming to locate specific targets within the environment accurately and efficiently. Success is measured by the time taken to find targets and the precision of their positioning on a map, leading to a decisive formula for determining the winner.

However, securing the victory comes at the end of a challenging process. Simone highlights the **delicate balance between computational power and flight time** in design considerations. *"For our task, we required a lot of computational power because we had to identify targets and obstacles fully autonomously,"* he explains. *"We had a trade-off between how long the drone could fly and the payload computationally on board. We added more computational power, boards, and sensors, so the drone's flight time constantly reduced with increased power on board."*

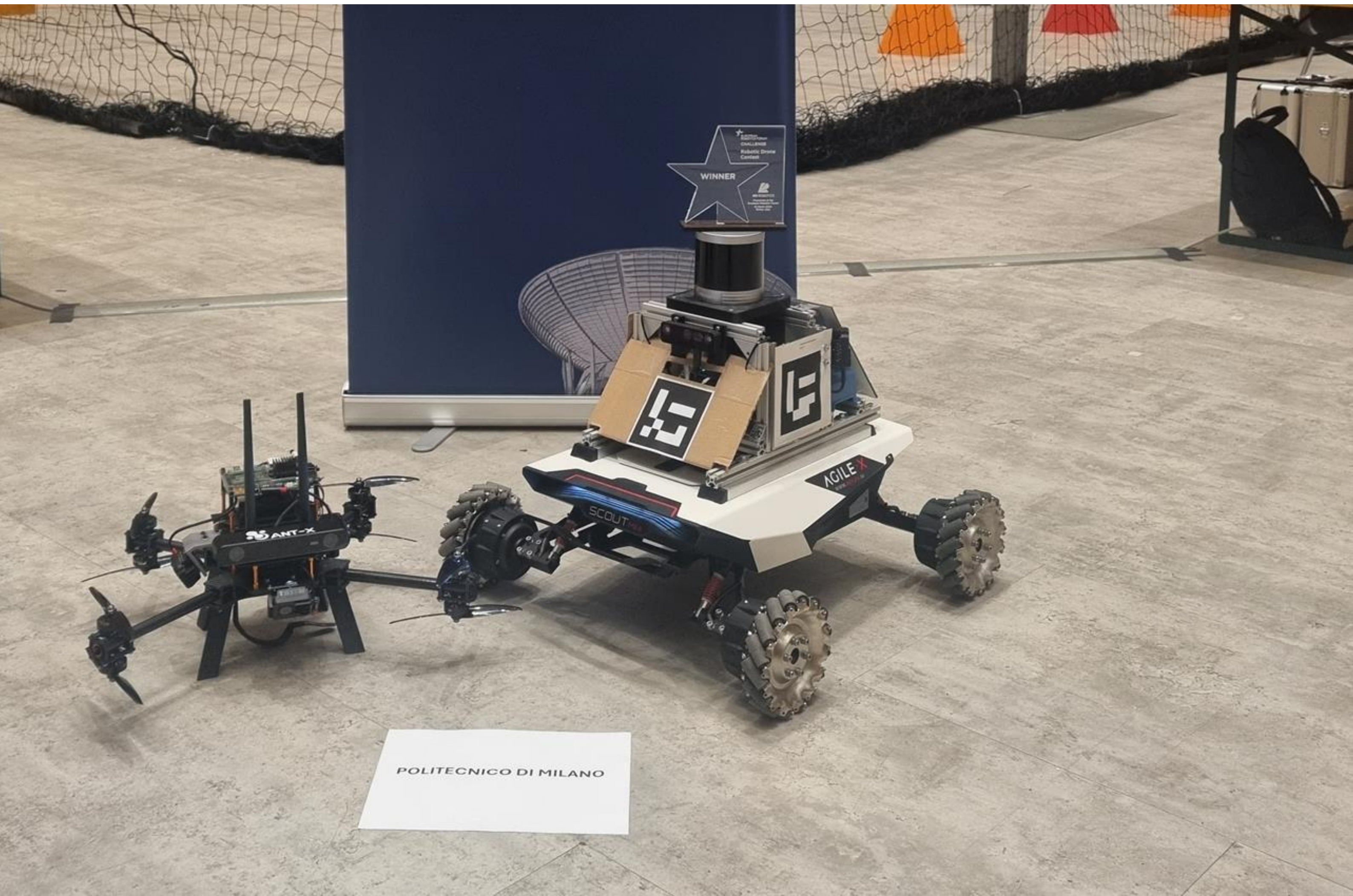
"We had a trade-off between how long the drone could fly and the payload computationally on board!"

All the computation is performed onboard the drone, with only limited information provided to the ground control stations. Simone points out that part of the computation could have been offloaded to agents, but not doing this helped to optimize performance in a scenario where bandwidth was limited and the Wi-Fi connection was unstable.

Reflecting on their winning strategy, Simone credits the team's success to **a robust and efficient solution rather than advanced algorithms**. *"We wrote most of the communication on layers and most* *of the libraries to be really lightweight,"* he reveals. *"We transmitted only the data we needed, and most of the computation was done on the different agents. Each one was independent and able to carry out its own task without having to rely on information that could be disrupted by the Wi-Fi signal, which was not very good, and stop the movement of the robots."*

Despite their achievements, Simone recalls initial hurdles in **developing navigation algorithms from scratch**, which made it challenging to design functions like **obstacle avoidance and path planning**. They switched to **ROS-integrated control algorithms** to solve this, which worked well.

Away from competitions, Simone's core research focuses on **autonomous vehicles and vehicle-infrastructure cooperation**, and he is part of a team that performs **agricultural robotics**. Looking ahead, he is already gearing up for the next round of the competition in October. *"Now, we're starting our new round of development,"* he declares. *"We still have some ideas on how to improve the robots on the detection part, the navigation, and the controller, particularly on the drone side. We have a new team starting to work on all those aspects to improve the solution for the new round in October!"*



DON'T MISS THE

BEST OF
SPIE 2024
Medical Imaging

On pages 34 to 43!

*"It's been a privilege to witness the range of innovation and excellence displayed at **SPIE Medical Imaging** this year,"* said **RSIP Vision's CTO Ilya Kovler**. *"Across over 800 papers and a broad range of topics, the conference has been a showcase of cutting-edge technology and visionary research. My congratulations to the award winners. Their work sets a high standard for the entire community. I was particularly inspired by efforts to **integrate classical or more traditional imaging approaches with modern AI advancements**, drawing on the benefits of both to improve the state of the art. The conscious focus from participants on clinical applications and improving patient care has left me with a newfound excitement about the possibilities that lie ahead **to transform healthcare through computer vision in medical imaging!**"*

Emlyn Roy Davies

(29.11.1940 - 29.2.2024)

I (Ralph) interviewed Prof. E. Roy Davies exactly two years ago, when he co-edited a new book with Matthew Turk: *Advanced Methods and Deep Learning in Computer Vision*. Roy was Emeritus Professor of Computer Vision at Royal Holloway, University of London. I have the memory of a very kind and humble man. Our sympathies to daughter Sarah and family.

“As an editor, I found it stimulating finding how best to guide eminent authors to produce chapters that were even better than their initial drafts – and of course more didactic and beneficial for the eventual readers!”

“Knowing the fundamentals of computer vision and image processing has been crucial in making all this tremendous progress, and many people are worried we’ll have a whole new generation coming along who don’t know the basics. Part of the answer is that the field will probably start to break up more than it has in the past. It’s hard for one person to be an expert in everything!”

“Readers would do well to […] maybe wonder whether further developments in these directions will take us into totally new worlds that reflect the parts of our brains that are not so narrowly focussed on vision: after all, vision is only one aspect of our cognitive evolution!”

*“Dear Ralph and All,
Thank you for getting our feature up so quickly, and in such a good format/arrangement! It looks very pleasing - indeed, so appealing that I feel inclined to send it to all my non-vision friends as well as relevant colleagues!
Thanks for all your help with this.”*

28 Computer Vision Book

Advanced Methods and Deep Le

Roy Davies and Matthew Turk are speaking to us about a new book they have co-edited: *Advanced Methods and Deep Learning in Computer Vision.* Roy Davies is Emeritus Professor of Computer Vision at Royal Holloway, University of London, UK. Matthew Turk is the president of the Toyota Technological Institute at Chicago (TTIC), an independent philanthropically endowed academic computer science institute.

What can you tell us about your new book?

MT: The book covers advanced computer vision methods, emphasizing machine and deep learning techniques applied to computer vision and imaging, particularly those that have become more frequent, common, interesting, and essential in the last five to ten years. There’s been a tremendous amount of progress in the field, and we felt it was the right time to have a book that was neither an introductory text nor a collection of very detailed research, but something in the middle that provides some depth across several important topics.

RD: Indeed, we thought that it was important to emphasise principles and new methodologies rather than dwell excessively on details of ongoing research: in that way we should be able to help readers obtain a significantly deeper and longer term understanding of the subject.

Who is the target audience for the book?

The Editors: The book is aimed at people studying computer vision – graduates and undergraduates – and current researchers working in the field who want to look in more detail at an area they haven’t focused on much. Practitioners who wish to gain a better insight into specific areas of computer vision will find it helpful too.

If a newcomer to the field has your book in their hand, what advice would you give them to best take advantage of it?

MT: A great place to start is with the first chapter, written by Roy, which is an overview of the basic concepts of computer vision and sets the stage for the rest. After that, stand back and look at all the other topics. Decide which ones are most interesting to you. You may end up with two or three chapters. For each, start with the introductory sections and if you need more background on the topic follow the references that those bring up.

Roy Davies

That will give you an understanding of the bigger picture, and then you can delve bit by bit into the details and explore those references in depth if you want to go even further.

RD: While following this overall plan, it is also relevant to think carefully not only of the approach to be adopted but also how the input data (often in the form of many millions of images or image patches) should optimally be managed so as to most effectively train the final system.

My aim in writing the rather long first chapter was to ensure first that readers were brought up to speed on legacy computer vision work; second, to open the doors to deep learning and to show how it can successfully be applied to computer vision; and third, to demonstrate that even when applied to the familiar rather basic subject of texture analysis that substantial changes are needed in the old ways of thinking.

Elisa Roccia is the Global Clinical Marketing Manager for MRI Oncology at Siemens Healthineers.

Elisa, can you tell us about your work?

In Siemens Healthineers, I'm responsible for the MRI Oncology global marketing strategy within the magnetic resonance imaging business line. The main topics I cover are breast cancer, liver cancer, and prostate cancer, but there are also a few other minor topics I cover as well. My position is similar to a product management position, but facing more towards the outside of the company. I act as a bridge between the clinical world, what's happening in the field, in the community, and what happens inside Siemens Healthineers.

Last time we spoke, you had just completed your PhD. You have strong research training but seem to be doing less research and more product management now. How did that happen?

Yeah, that's right. I do come from a research background. I studied biomedical engineering as an undergrad in Italy. Then, I did my PhD

at King's College London. The PhD was very technical indeed. I developed MR sequences to acquire MR images. I programmed a lot on my laptop and had to study many new things, from the physics of the MR system to C++ programming, for instance.

During the PhD, because we were in a very nice, stimulating environment at King's College, I had the chance to do many other activities alongside my PhD. For instance, I was very much involved in public engagement. We were going to different schools, talking with kids on what we were doing as part of our PhD, what the research was about, and what it means to be a scientist. Other events that come to my mind are the Royal Summer Science Exhibition, where we talked with the general public about our work and research. All these activities were super engaging to me. I was always so happy to talk in these more lay terms about science.

I also enjoyed organizing events and conferences, so I understood that what I was missing as part of my PhD was more of a communication and human aspect. The position I have right now in Siemens Healthineers is a really good combination of these two aspects, because I need to use my technical knowledge to act as a translator between what happens from the technical point of view into more easily understandable concepts.

I imagine speaking to kids and the public is very different from being a marketing manager competing in the global market with all the sharks in the ocean.

Yeah, no, absolutely. At Siemens Healthineers, I've been very lucky to be working in the Headquarter team, which is based in Germany in Erlangen. There, I met many brilliant colleagues, and it's extremely stimulating to have all these people around you with so much knowledge about MRI in particular. It's an environment that fosters being very competitive and at the top of what's happening. In MRI in particular, there's always something going on from the research perspective, it's an ever-evolving field. This stimulates advancements in the field and challenges us to remain on top.

With access to MRI remaining very competitive, do you ever worry that people will work more on cheaper and more accessible technologies like X-ray and ultrasound instead?

The access aspect is definitely a real issue, but nowadays we have the right technologies and innovations that enable us to make MRI more accessible. One of these is the fact that the MRI acquisition is much faster than what used to be. If you have a shorter examination time, you can scan more patients. This means that you have a higher patient throughput, which balances the costs of the MRI systems.

As we're talking in a computer vision journal, maybe we can also mention that these technologies are often based on deep learning – how we can speed up the scan and the reconstruction of the images. This technology has been really disruptive in the past few years, and this is just one of the examples of how we can improve access to MRI, in addition to having a broad portfolio of systems that can meet different needs in different countries.

What computer vision you work on and how algorithms are used in your office?

In the past few years, new deep learning-based reconstruction algorithms have been introduced, not only by Siemens Healthineers but also by other companies and universities. It's a topic a lot of academics and PhD students are working on, on how we can make the acquisition and the reconstruction faster, which means acquiring less data and still obtaining a comparable result thanks to the neural networks that lie behind it. For MRI, in particular, this means not having to compromise between signal-to-noise ratio, acquisition time, and image resolution. Now, you can find a nice balance between the three.

More signal, less noise.

Exactly! [*Elisa laughs*]

I have a devil's advocate question – why does Siemens Healthineers need you in its workforce?

Everyone brings their own personal contribution to the role thanks to our past experience. What I think is valuable in my particular case is the technical background, because that means I can take a technical publication and translate into a clinically relevant presentation, with a catchy title, or come up with a topic for an event that reflects what's happening in the field, so that it becomes attractive to someone not necessarily as technical. Then, as I mentioned earlier, the communication skills, engaging with people and with the community, helps a lot in this position.

What are you most proud of in the first years at Siemens Healthineers?

If I look back at the past three years, the difference I see in the confidence I have in the job I do day to day has changed massively. Starting as a fresh PhD student, getting into this huge company where there are so many smart people, and covering many more aspects and topics than I was used to. During the PhD, you look at a very narrow topic, and you're the expert on that topic. Whereas, at this level, you have to cover much more at a higher level, so you're not that deeply into everything, but still, it's a lot. I remember that at the beginning, I was thinking, how will I be able to know all these things? Then, with time and experience, you learn the right balance between knowing enough to cover the topic without necessarily having a PhD in everything, which, of course, is not possible! I think what I'm most proud of now is feeling the confidence I have in handling my position and going out there and talking about my topics.

If I told you that 10 years from now, you will be a medical imaging researcher who writes code, would you believe me?

If I look back at the past three years, the difference I see in the confidence I have in the job I do day to day has changed massively. Starting as a fresh PhD student, getting into this huge company where there are so many smart people, and covering many more aspects and topics than I was used to. During the PhD, you look at a very

narrow topic, and you're the expert on that topic. Whereas, at this level, you have to cover much more at a higher level, so you're not that deeply into everything, but still, it's a lot. I remember that at the beginning, I was thinking, how will I be able to know all these things? Then, with time and experience, you learn the right balance between knowing enough to cover the topic without necessarily having a PhD in everything, which, of course, is not possible! I think what I'm most proud of now is feeling the confidence I have in handling my position and going out there and talking about my topics.

If I told you that 10 years from now, you will be a medical imaging researcher who writes code, would you believe me?

I think that will not happen! [*she laughs*] I tried that already, and it was a really fun experience. I especially liked seeing how a single line of code could change the outcome of what I could see in a patient's scan. That was invaluable. Changing something that seemed to be a small thing on my laptop and then going downstairs at the scanner, and with that small change, I could see a tumor that I couldn't see before. That was really exciting and satisfying.

Does it ever make you feel sad working with people's diseases all the time?

No, that's not a feeling I've ever had. On the contrary, I'm happy to, in my small way, contribute and do something about it. Of course, there's always a huge team behind this. My contribution may be small, but I still feel like I'm doing something to go in that direction, to improve the technology so that people can receive an earlier diagnosis. Yeah, to have an impact and be involved, and also, especially for me, what is very important is to communicate what I do to the community, to my network, so that more people know what's happening in the community. If MRI is going to be used for screening prostate cancer, for instance, which is something that might happen in the future, it's good that people know about this, because not everyone reads scientific publications.

What is the best thing that you have learned or received from Siemens Healthineers?

That's a tricky one.

I'm sure that you are very grateful to your company.

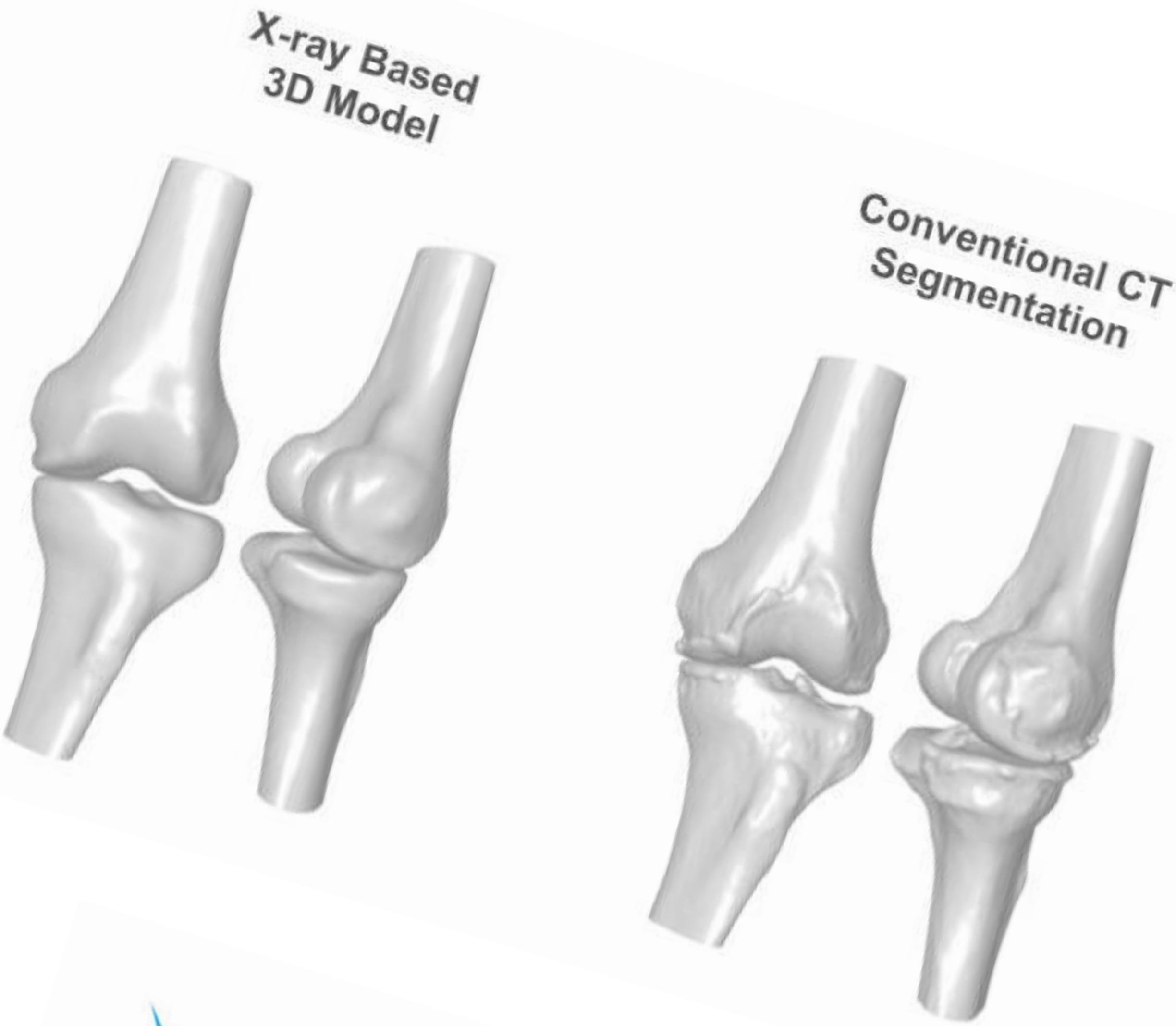
Indeed, it's tricky because there are so many things! [*she laughs*] I would say, the fact that I now cover many more topics than I used to before. The possibility to learn something every single day. Because it's such a big company, you have a lot of touchpoints with many different people. Some of them might be working in MRI, and others might be working in ultrasound or CT, so I get to expand my knowledge by being with all these people.

Something that I also like very much is the travelling, which I've been doing quite often in the past couple of years. That gives me the opportunity to see how things are done in different countries and what's really happening in the field – even just comparing Italy to Germany or to the UK or the US, every country has its own trends or preferences. The nice thing about these global roles is that you get to see a little bit of everything, and you get to think of what there is in common and what there is that's different, how you can adapt your talk so that you include everyone and make sure that what you say is relevant for everyone. That's something invaluable.

Is there anything else our readers should know about your work?

Maybe not necessarily related to my job, but the advice I'd give students is to try things out during your PhD. That's the time when, yes, you're super busy because, of course, you need to work on your thesis and projects, but that's also the chance you have to experience new things. I was working mainly on the technical side. I tried something more toward the communication aspect, and that helped me find the direction I wanted to go. There might be someone who's working in communication and could try out what programming means, and they might find out that's exactly what they want or exactly what they don't want to do. Yeah, just explore things as much as possible!

Read 100 FASCINATING interviews with Women in Computer Vision!


X-ray Based
3D Model
Conventional CT
Segmentation

XPlan.ai Confirms Premier Precision in Peer-Reviewed Clinical Study of its 2D-to-3D Knee Reconstruction Solution.

The study, featured in the prestigious Journal of Clinical Medicine, found sub-millimeter accuracy on real-world patient imaging, enabling widespread access to precise, image-based computer-assisted surgery without the need for a CT scan.

TEL AVIV, Israel – Mar 21, 2024

XPlan.ai, a spinoff of **RSIP Vision** that uses AI to democratize precision orthopedics, reaches a new milestone for its **X-ray based 3D bone modeling system** with the publication of a peer-reviewed clinical study confirming unprecedented, sub-mm accuracy in a variety of clinically relevant measurements. The study, led by a consortium of orthopedic surgeons and published in the **Journal of Clinical Medicine**, found that XPlan.ai offers a promising alternative to conventional CT scans, opening up the market for image-based computer-assisted surgery such as robotics, AR, and navigation while increasing efficiency, saving costs and avoiding unnecessary radiation exposure.

XPlan.ai uses advanced artificial intelligence (AI) to produce accurate 3-dimensional bone models from two standard X-ray images. Together with XPlan's automated planning technologies, this model can be used for surgical planning and navigation during orthopedic procedures such as total knee replacement, potentially providing incredibly large patient populations with the most advanced care while avoiding the cost, time, administrative overhead, reimbursement issues, and added radiation involved in a conventional CT scan.

"Replacing CT scans with standard, universally available X-rays has long been considered a 'holy grail' of

Journal of Clinical Medicine MDPI

Article

Validating a Novel 2D to 3D Knee Reconstruction Method on Preoperative Total Knee Arthroplasty Patient Anatomies

Shai Factor [1,*], Ron Gurel [1], Dor Dan [2], Guy Benkovich [3], Amit Sagi [4,5,6], Artsiom Abialevich [5,7,8] and Vadim Benkovich [5,7,8]

1 Division of Orthopedic Surgery, Tel Aviv Medical Center, Faculty of Medicine, Tel Aviv University, Tel Aviv 6423906, Israel
2 Orthopedic Department, Meir Medical Center, Faculty of Medicine, Tel Aviv University, Tel Aviv 4428164, Israel
3 Orthopedic Department, Sheba Medical Center, Faculty of Medicine, Tel Aviv University, Tel Aviv 5262000, Israel
4 Orthopedic Department, Barzilai Medical Center, Ashkelon 78278, Israel
5 Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva 8499000, Israel
6 South West London Elective Orthopaedic Centre, Epsom KT18 7EG, UK
7 Department of Orthopedic Surgery, Soroka Medical Center, Beer Sheva 84101, Israel
8 Israeli Joint Health Center, Tel Aviv 69710, Israel
* Correspondence: factor310@gmail.com; Tel.: +972-527360753; Fax: +972-74-7219810

Abstract: Background: As advanced technology continues to evolve, incorporating robotics into surgical procedures has become imperative for precision and accuracy in preoperative planning. Nevertheless, the integration of three-dimensional (3D) imaging into these processes presents both financial considerations and potential patient safety concerns. This study aims to assess the accuracy of a novel 2D-to-3D knee reconstruction solution, RSIP XPlan.ai™ (RSIP Vision, Jerusalem, Israel), on preoperative total knee arthroplasty (TKA) patient anatomies. Methods: Accuracy was calculated by measuring the Root Mean Square Error (RMSE) between X-ray-based 3D bone models generated by the algorithm and corresponding CT bone segmentations (distances of each mesh vertex to the closest vertex in the second mesh). The RMSE was computed globally for each bone, locally for eight clinically relevant bony landmark regions, and along simulated bone cut contours. In addition, the accuracies of three anatomical axes were assessed by comparing angular deviations to inter- and intra-observer baseline values. Results: The global RMSE was 0.93 ± 0.25 mm for the femur and 0.88 ± 0.14 mm for the tibia. Local RMSE values for bony landmark regions were 0.51 ± 0.33 mm for the five femoral landmarks and 0.47 ± 0.17 mm for the three tibial landmarks. The RMSE along simulated cut contours was 0.75 ± 0.35 mm for the distal femur cut and 0.63 ± 0.27 mm for the proximal tibial cut. Anatomical axial average angular deviations were 1.89° for the trans epicondylar axis (with an inter- and intra-observer baseline of 1.43°), 1.78° for the posterior condylar axis (with a baseline of 1.71°), and 2.82° (with a baseline of 2.56°) for the medial–lateral transverse axis. Conclusions: The study findings demonstrate promising results regarding the accuracy of XPlan.ai™ in reconstructing 3D bone models from plain-film X-rays. The observed accuracy on real-world TKA patient anatomies in anatomically relevant regions, including bony landmarks, cut contours, and axes, suggests the potential utility of this method in various clinical scenarios. Further validation studies on larger cohorts are warranted to fully assess the reliability and generalizability of our results. Nonetheless, our findings lay the groundwork for potential advancements in future robotic arthroplasty technologies, with XPlan.ai™ offering a promising alternative to conventional CT scans in certain clinical contexts.

Keywords: knee reconstruction; accuracy assessment; 2D-to-3D; total knee arthroplasty (TKA); registration; artificial intelligence; 3D reconstruction; RSIP XPlan.ai

check for updates

Citation: Factor, S.; Gurel, R.; Dan, D.; Benkovich, G.; Sagi, A.; Abialevich, A.; Benkovich, V. Validating a Novel 2D to 3D Knee Reconstruction Method on Preoperative Total Knee Arthroplasty Patient Anatomies. *J. Clin. Med.* **2024**, 13, 1255. https://doi.org/10.3390/jcm13051255

Academic Editors: Tilman Walker, Tobias Gotterbarm and Babak Moradi

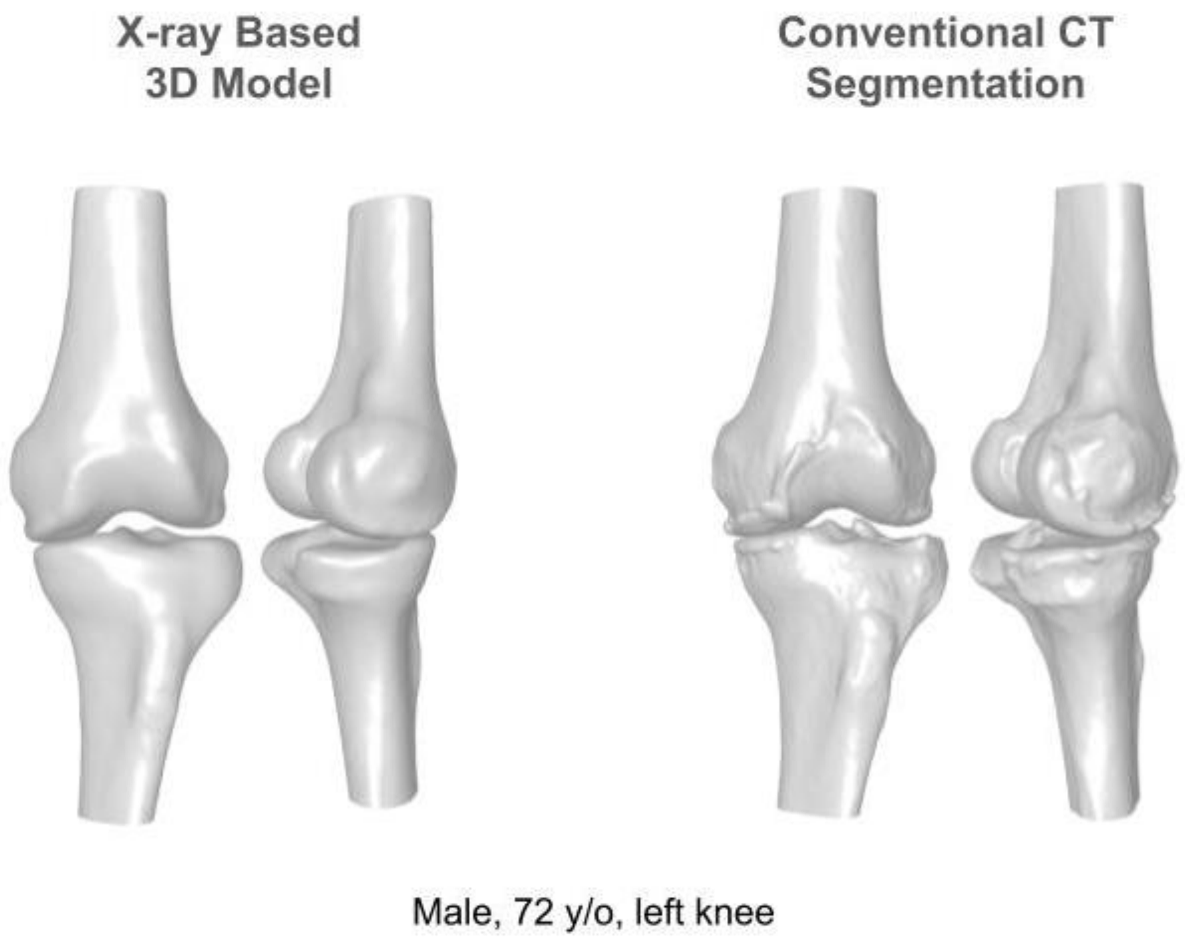
Received: 2 January 2024
Revised: 18 February 2024
Accepted: 20 February 2024
Published: 22 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

J. Clin. Med. **2024**, 13, 1255. https://doi.org/10.3390/jcm13051255 https://www.mdpi.com/journal/jcm *computer-assisted orthopedic surgery,"* said **Moshe Safran, CEO at XPlan.ai**. *"Accuracy and robustness have been the key challenges, and our technology provides unique capabilities in this regard, paving the way for universal access to image-based computer assisted surgery."*

XPlan.ai's groundbreaking technology provides two main benefits: clinical and operational. The clinical benefit is lower levels of radiation exposure from X-ray imaging compared to a complete knee CT scan. Operationally, all relevant patients routinely undergo X-ray imaging, offering much better accessibility than CT. Additionally, X-rays are more widely reimbursed in the U.S. healthcare system, are often lower in cost, and offer a quicker and more streamlined patient journey from diagnosis to the OR.

The clinical evaluation of this

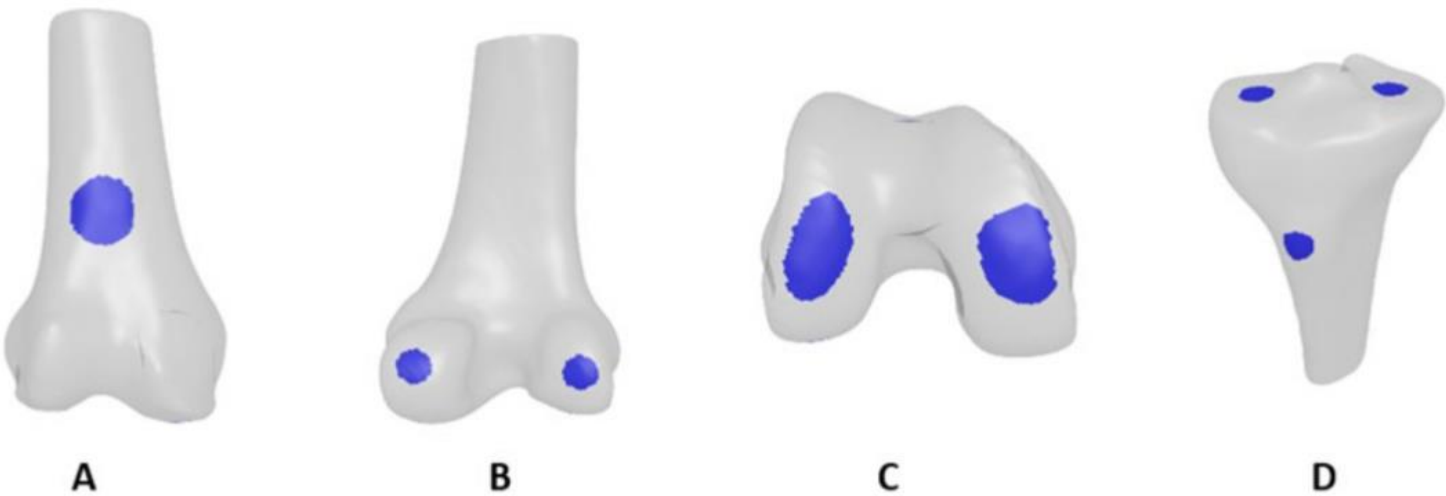


XPlan's high-fidelity X-ray based 3D knee model, compared to a conventional CT based model of the same patient

Tool was conducted using imaging from total knee replacement patients from **Assuta Medical Center in Tel-Aviv**, a leading medical center in Israel.

Unlike cadaver-based studies often used in the orthopedic technology space, the patients enrolled in this study had pathological anatomies typical of real-world clinical cases. The accuracy of the tool was proven by comparing the resulting 3D models to the ground truth patient anatomy given in a corresponding CT scan. The accuracy was measured in multiple areas that are used for actual surgical planning, including bony landmarks and anatomical axes, and was found to be equivalent to CT-based measurements at a sub-mm level across the board.

"Today's orthopedic patients demand precise and personalized care, incorporating technologies such as AR and surgical robotics. Using a 3D image-based preoperative model to plan the case is the best approach, enabling surgeons to be better prepared and saving precious time in the OR," said **Dr. Vadim Benkovich, Head of Orthopedic Department at University Medical Center Soroka and Founder & Medical Director of the Israeli Joint Health Center at Assuta Medical Center**. *"This is a win for both patients and providers – both reducing the chances of a complication or infection, and at the same time improving efficiency and*



**Bony landmark regions. (A) Femur anterior cortex.
(B) Femur posterior condyles. (C) Femur distal condyles.
(D) Tibial tuberosity, medial and lateral plateaus.**

providing care to more patients with the same amount of resources. I am encouraged to see that the accuracy of XPlan's solution has passed the most stringent tests conducted in our study."

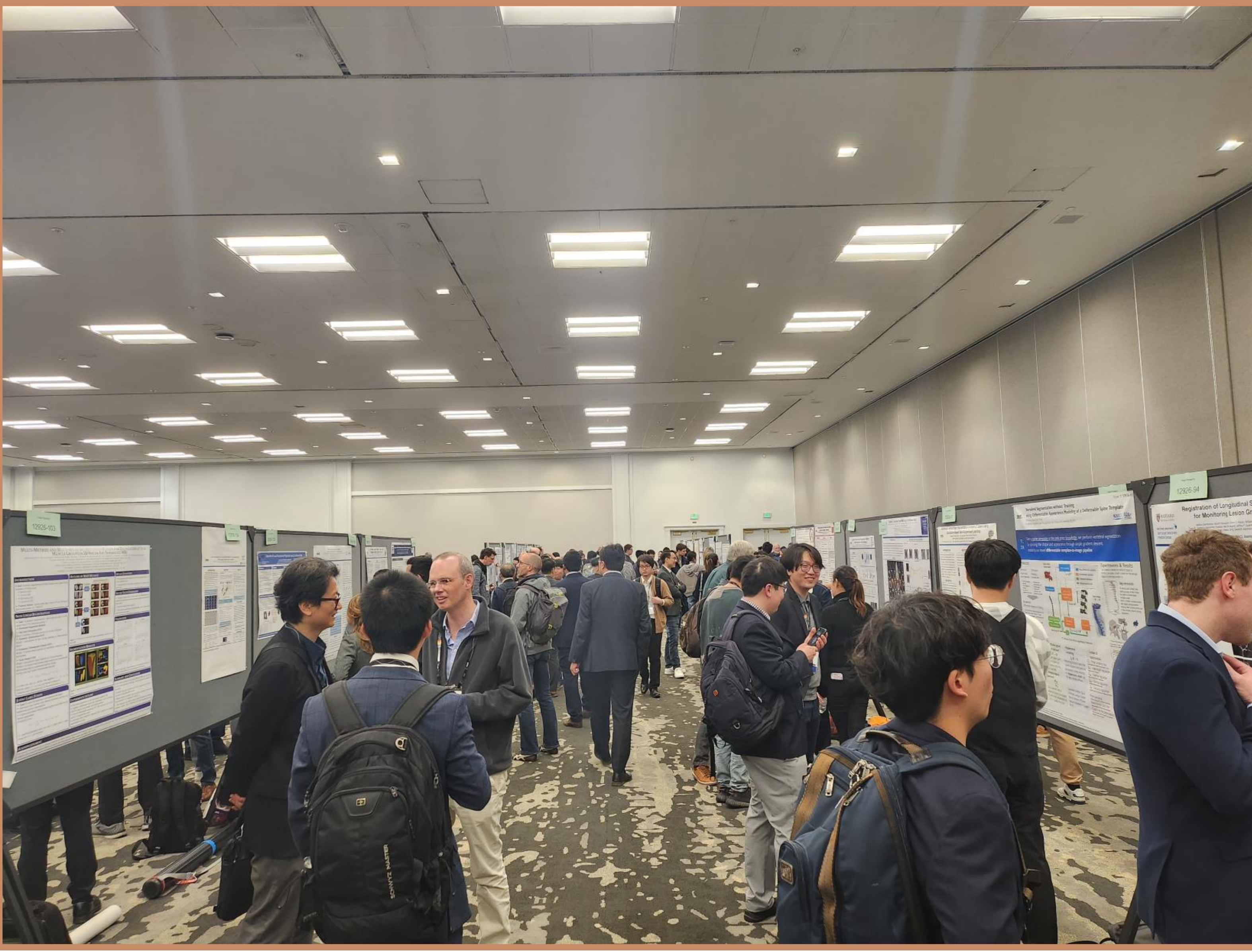
Going forward, XPlan.ai plans to **apply for FDA clearance** of its knee reconstruction solution. In parallel, further applications are under development for additional anatomies, with promising initial results indicating wide applicability of XPlan's unique technology.

About Xplan.ai

XPlan.ai, a spinoff of RSIP Vision, is building an advanced AI-based solution that provides 3D bone reconstructions from standard 2D X-rays. Our platform technology can potentially improve the efficiency and safety of computer assisted surgery, including eliminating the need for preoperative CT scan and reducing time in the OR. XPlan.ai is headquartered in Jerusalem, Israel. More information is available on the company website at xplan.ai and by contacting us at info@xplan.ai.

Moshe Safran, CEO at XPlan.ai

Photos courtesy of Nadieh Khalili, an AI scientist from Iran. She poses in the photo on the right with Khrystyna Faryna, a PhD student from Ukraine. Both women are affiliated with Radboud University Medical Center. Nadieh is interested in multimodal medical data ranging from Pathology, Radiology and genomics.



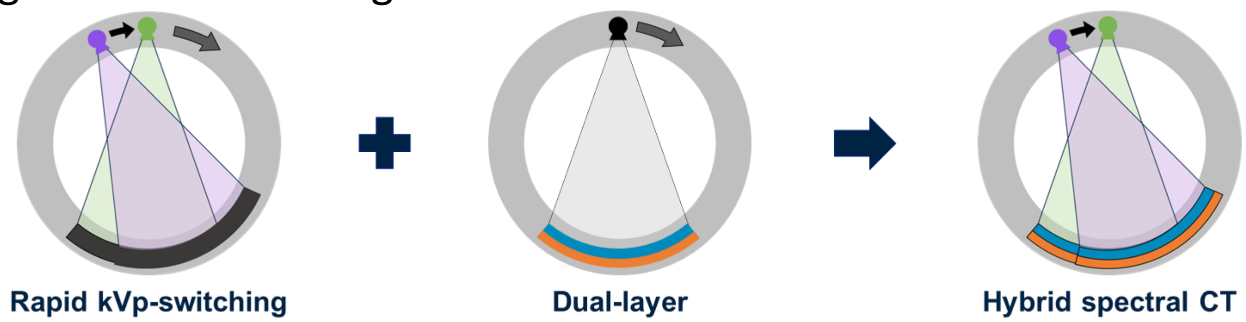

SPIE.
MEDICAL
IMAGING
BEST OF
SPIE 2024
Medical Imaging

Hybrid spectral CT system with clinical rapid kVp-switching X-ray tube and dual-layer detector for improved iodine quantification

Olivia Sandvold is a fourth-year bioengineering PhD student at the University of Pennsylvania. Her paper, demonstrating a novel hybrid spectral CT system to reduce error in iodine quantification, has just won the Physics of Medical Imaging Best Paper Award at SPIE Medical Imaging. This impressive young woman is here to tell us all about it.

In this paper, Olivia introduces a new **multi-energy hybrid CT system**. It combines a traditional kVp switching X-ray tube source, which alternates between various energy levels emitted from the X-ray tube, with a dual-layer or sandwich detector, a spectral detector that discriminates energy into upper and lower domains.

The work demonstrates that joining these technologies and leveraging the properties of the multiple channels acquired can **reduce errors in iodine quantification**, which is important in clinical diagnoses. Physicians use iodine contrast to measure increased uptake from the bloodstream or the oral pathway into different body cavities. If accurate, these measurements can act as **quantitative biomarkers for disease**.



Top-down view of bench

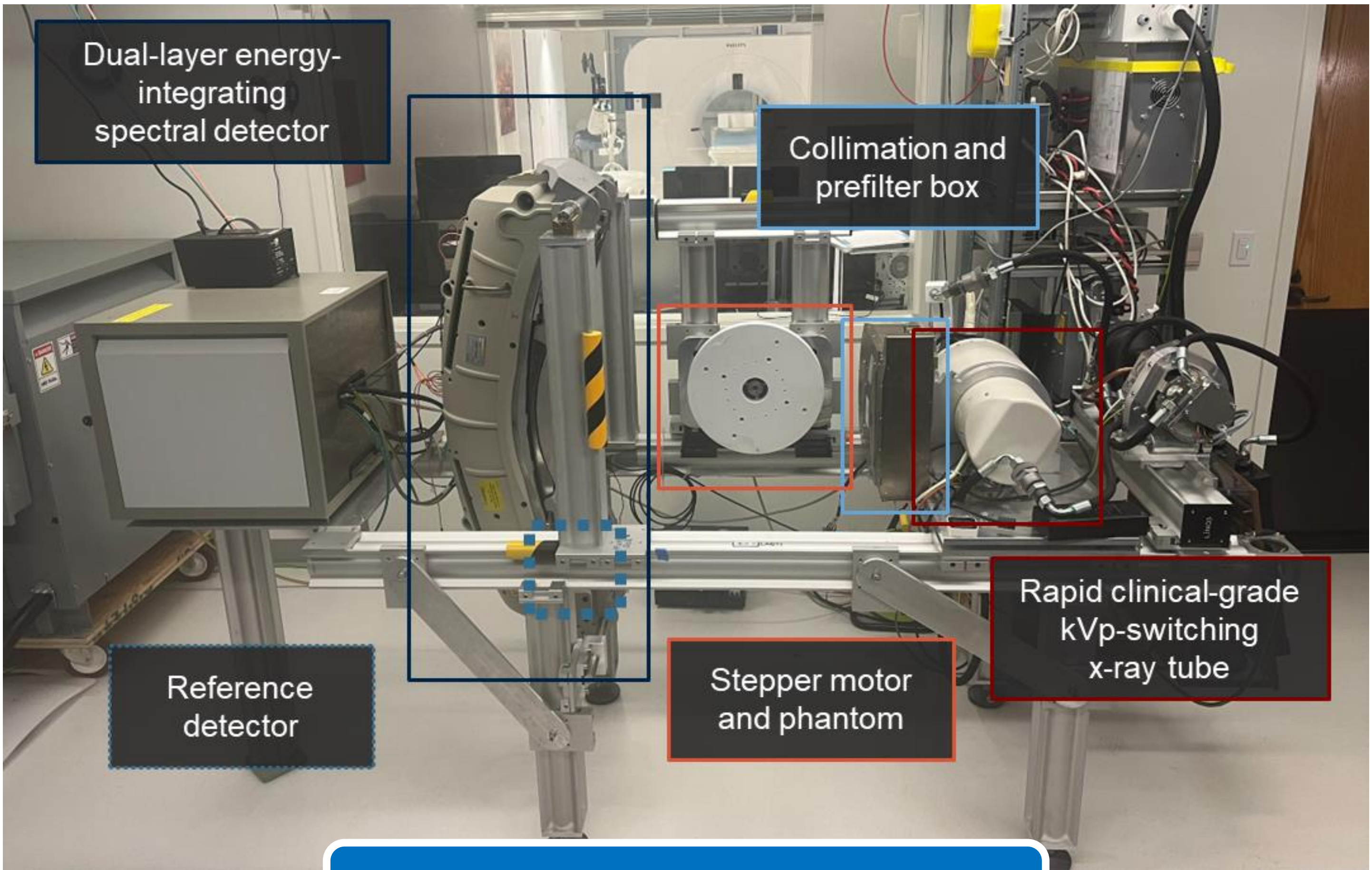
"We want to reduce the number of scans that patients have to take over a period of time," Olivia tells us. *"As we know, CT involves X-rays, and X-rays involve radiation, and we always want to decrease the amount of radiation we supply to the patient. If we have an **increased sensitivity to iodine to be able to acquire very high-quality images for diagnosis**, then the patient should not have to undergo many additional scans."*

From a clinician's perspective, this represents an improvement over conventional CT or even single instrumentation spectral CT systems, offering **greater confidence in quantitative measurements**. This newfound precision improves pre- and post-treatment assessments and contributes to establishing standardized medical practices, ultimately benefiting the entire medical community through increased knowledge and accuracy in diagnostics.

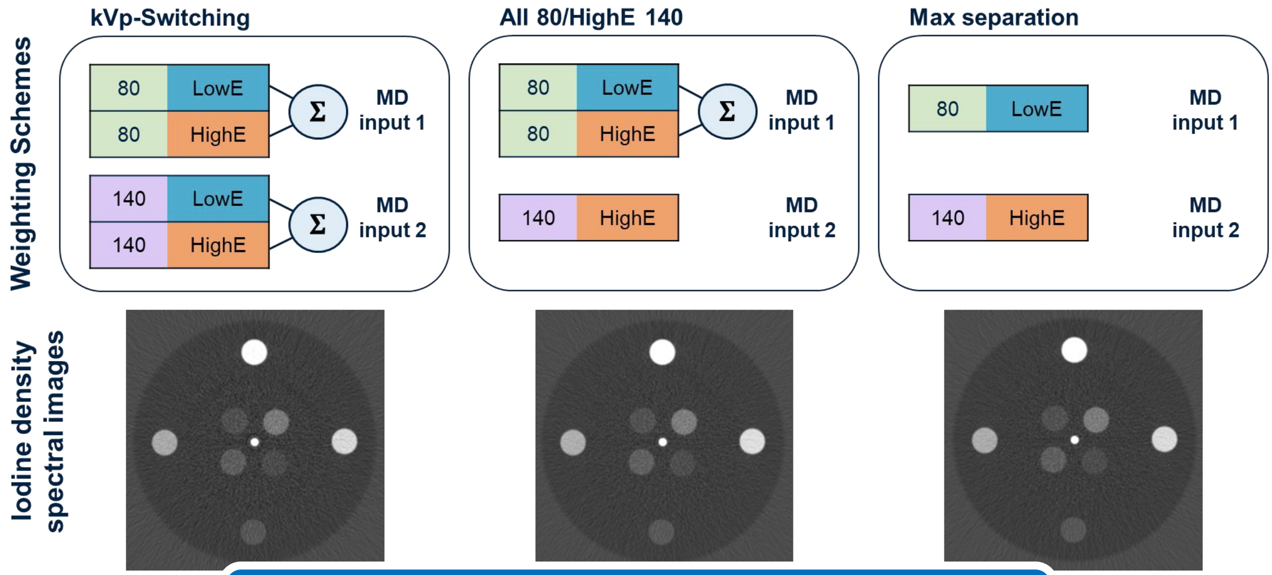
Olivia says the system encompasses **hardware and software advancements over previous setups**. She points out that thinking about hardware in combination with software, AI, and machine learning is key. *"If we have better underlying measurements from our hardware, we'll be better at applying different computer vision, downstream image processing, and algorithmic development,"* she attests. *"The problem is now we have four channels of data, so what do we do with all this data? We're proposing different software instrumentations and the inclusion of these four channels* *to utilize each one of the spectral properties specifically."*

Getting the different systems to talk to each other presented a challenge. With a **rapid kVp-switching tube**, which uses both high-energy and low-energy X-rays, it was vital to precisely coordinate the switching time from high to low so that the measurements did not get mixed up. Ensuring a good hardware setup was critical to accurately separating things. *"On the software side, we wanted to make sure that when we segregate our four channels of data and then combine them to have the least bias in our measurements, we have appropriate weighting based on the spectra,"* Olivia explains. *"It was a challenge to take all of the information our benchtop system gave us and then have a new pipeline in place that fits into the existing clinical pipeline. We have a starting point and an endpoint, and we want to make sure that what we build falls into that scheme so that we don't have to change a lot if we push this to the clinical level."*

While the current focus remains on iodine quantification, Olivia hints at potential future applications, including **the integration of computer vision for automated segmentation of different regions of high iodine contrast**. Higher sensitivity to iodine allows the generation of highly accurate iodine maps, which could revolutionize **tumor analysis, grading, staging,** and



Labeled components of hybrid system



Three defined weighting schemes to combine multi-energy channel data

metastases detection.

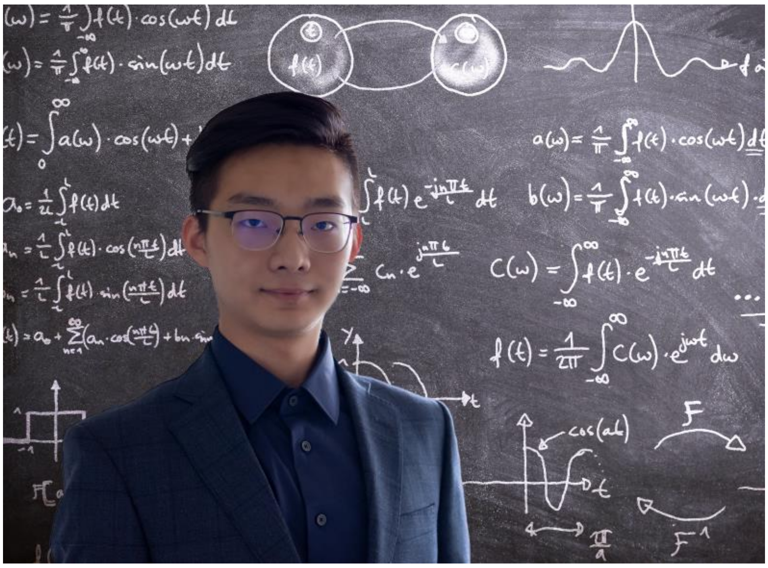
Did we just get a sneak preview of next year's winning paper? *"Yes, hopefully!"* she laughs. *"In this paper, we've shown improved quantification and decreased error by using and leveraging these two technologies. Now that we're more confident, what else can we do that has a downstream effect on specific patient populations?"*

Having scooped the **Best Paper award**, Olivia tells us she is very proud of the work and honored that it has been recognized in this way. She set out to investigate something people are curious about but may not have the resources to work on in their labs. Could this have been the secret to its success in the eyes of the judges? *"I think one of the compelling reasons this paper was important to the committee was that we're doing our best* ***to ensure translatability from the physical work to the clinic****,"* she remarks. *"Being at the University of Pennsylvania, we were so fortunate to be able to build a system that uses clinical components and is not just a benchtop system throwing together parts acquired from an X-ray source or detector."*

Olivia is currently in the fourth year of her PhD and hopes to graduate in the next couple of years. *"I'm passionate about continuing research and development with medical imaging,"* she adds. *"I'd like to continue working in this field, looking at* ***CT physics and the development of new devices****. That's my goal. I could see myself continuing in academic research, but also teaching, down the line, and potentially working in the industry, so it's still a little bit open."*

We certainly hope this award will help open the right doors. *"Thank you so much,"* she smiles. *"It definitely helps to have a little recognition!"*

Is registering raw tagged-MR enough for strain estimation in the era of deep learning?



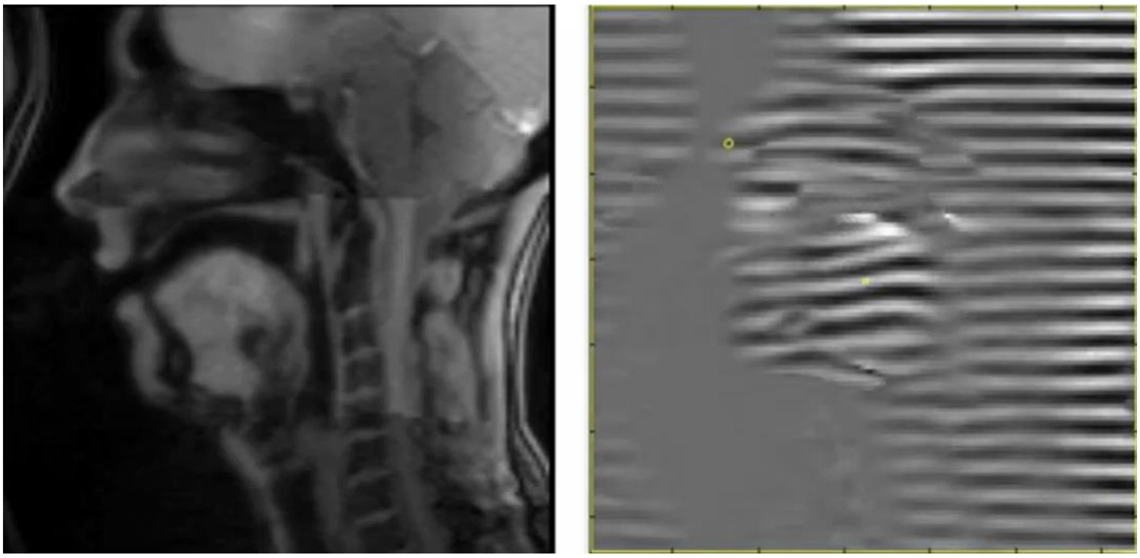
Zhangxing Bian is a third-year PhD candidate at Johns Hopkins University. Fresh from winning the 2024 Image Processing Best Student Paper Award at the SPIE Medical Imaging conference, he is here to tell us more about his work on tag fading, a post-processing complication that affects tagged MRI.

Tagged MRI (tMRI) is a specialized technique that adds specific patterns to tissues, **similar to temporary tattoos**. When the tissue moves, the tag moves with it, and clinicians can track these movements to better understand cardiac, muscular, and speech-related functions post-injury or in a disease context.

However, a challenge within this domain is the phenomenon of **tag fading**, where the visibility of tags diminishes over time, complicating accurate motion tracking and analysis. In this paper, Zhangxing wants to understand **what causes tag fading and what can be done post-processing to estimate tissue motion better**.

"Two decades ago, researchers proposed some classic signal processing methods, which extract the material's phase information through a Fourier transform for tracking the motion of the tissue," he explains. *"It can be seen as a special type of phase-based optical flow approach. The benefit is it circumvents the tag fading problem."*

Recent advancements in deep learning have sparked a revolution across various fields, presenting alternative methods that do not use those classic techniques to pre-process the image but **directly process raw tMRI inputs to estimate the motion or strain fields of the moving tissues**.



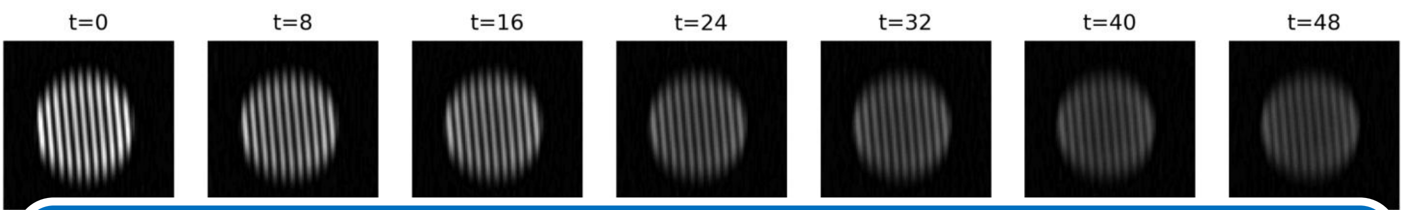
The left image is a sagittal view of a head. The video on the right shows the tagged-MRI acquired during speech when the tongue is moving.

Zhangxing revisits both methodologies in this work, highlighting **the current limitations of deep learning in biomedical applications** and emphasizing that it is not a universal solution to this problem. Instead, he advocates for a balanced approach that integrates **classic signal processing**.

“*People’s understanding of tag fading is currently not very complete,*” he advises. “*Our first contribution is to model the tag fading by considering factors that previous research ignored. The interplay between the* T_1 *relaxation and the repeated application of radio frequency pulses during the imaging sequences was overlooked in previous research on tMRI post-processing.* ***We build a mathematical model to factor that interplay into the equation****.*”

The findings of this work are derived from **both simulated images and an actual phantom scan**. Experiments on synthetic and real tMRI reveal the limitations of widely used similarity losses in raw tMRI and emphasize **caution in registration tasks where image intensity changes over time**.

While not proposing a new algorithm, this multidimensional work encompasses **a thorough comparative analysis between deep learning and**



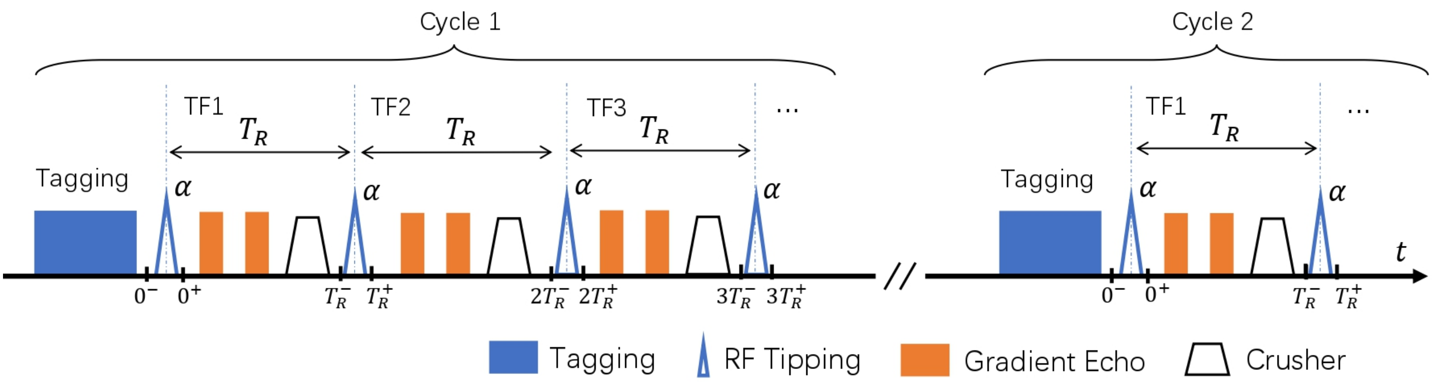
The tagged-MRI has a significant phenomenon called, tag fading, which a gradual decrease in tag visibility over time. The brightness constancy assumption used in optical flow or image registration does not hold, which leads to inaccurate motion estimation.

"The aim of this research is to better understand the tag fading phenomenon itself and evaluate the effectiveness of both traditional harmonic phase-based methods and deep learning-based registration methods (trained with diverse similarity objectives) when estimating motion with the presence of tag fading."

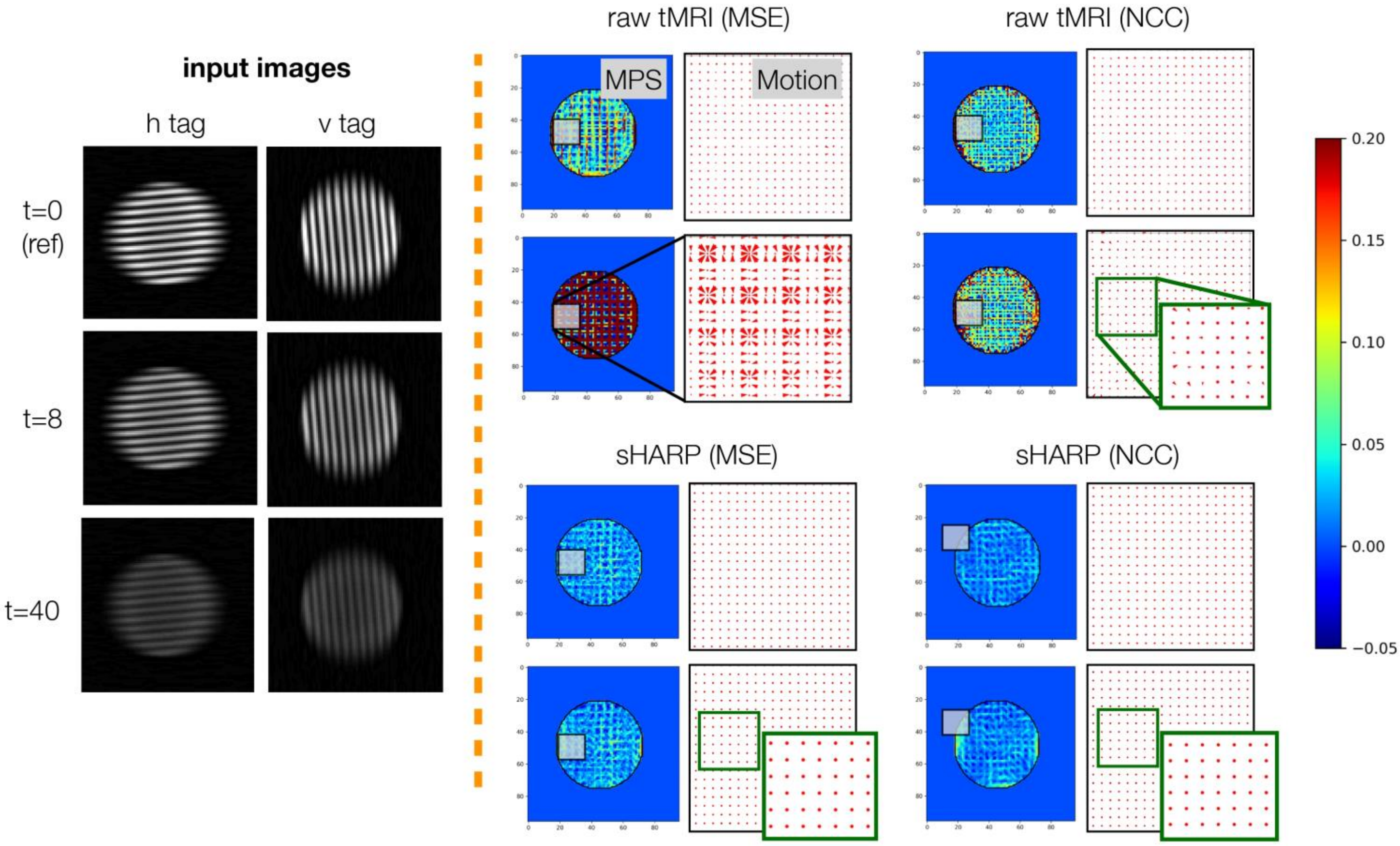
traditional methods to try to solve the tag fading issue. Zhangxing thinks this is why it piqued the interest of the **SPIE Medical Imaging Best Paper Awards judges**.

"The conference typically has different specialization tracks," he tells us. *"The two major tracks are called Medical Imaging and Image Processing. I think one reason this paper drew the committee's attention is that it provides both sides with an understanding of the tag fading issue. Also, our work is like a whistleblower to remind people that* ***there is something that deep learning currently can't handle well*** *and that, two decades ago,* ***the traditional signal processing method did pretty well and elegantly****. This observation provides some promising directions for the field going forward."*

Looking ahead, Zhangxing is working on extending the research by delving deep similarity metric learning. This technique has shown some promise in the inter-modality image registration task and is presumed to be robust against the challenges posed by tag fading. Although this work is limited to using the classic 1:1 SPAMM tagging sequence, the potential of using Complementary-SPAMM (CSPAMM) sequences to



The classic 1:1 SPAMM sequence is used for acquiring tagged-MRI. Each tagging step is followed by a series of imaging sequences. "TF" stands for timeframe. During each TR interval, the spin system is tipped by alpha degree and multiple line segments in k-space are captured using gradient echoes. The "tagging-imaging" cycle is repeated until sufficient k-space coverage is achieved. Based on this imaging sequence, a mathematical model has been built in this research for better understanding the tag fading process.



Different methods are evaluated on a tagged-MRI of a motionless gel phantom, where the ground truth motion is zero. Left panel shows the acquired horizontal (h-tag) and vertical (v-tag) tagging images for three timeframes. Right panel shows the color-coded maximum principal strain map and zoomed-in motion field of four methods. The first and second row show the registration results of ref-to-8th and ref-to-40th frame.

enhance tag contrast and alleviate into tag fading is also something that he intends to explore in the future. Could that constitute a sneak peek at the winning paper for next year? *"Yeah, hopefully!"* he laughs. *"I'm just trying to do my best."*

Zhangxing, originally from China, has been in the US for around five years. He studied for his master's in computer vision at the University of Michigan before moving into medical imaging. *"Before doing any research or publication, I like to have a deeper understanding of the problem at hand instead of just applying any other approach,"* he says. *"I like to explore. I do a lot of sports, like snowboarding and climbing. I consider myself an adventurous person."*

He is still considering his next move after his PhD. While he has not made any firm decisions regarding a choice between academia and industry, he is leaning in one direction. *"Right now, I prefer industry,"* he reveals. *"Probably going to a pharmaceutical or medical imaging device company to do real applications and solve real-life problems."*

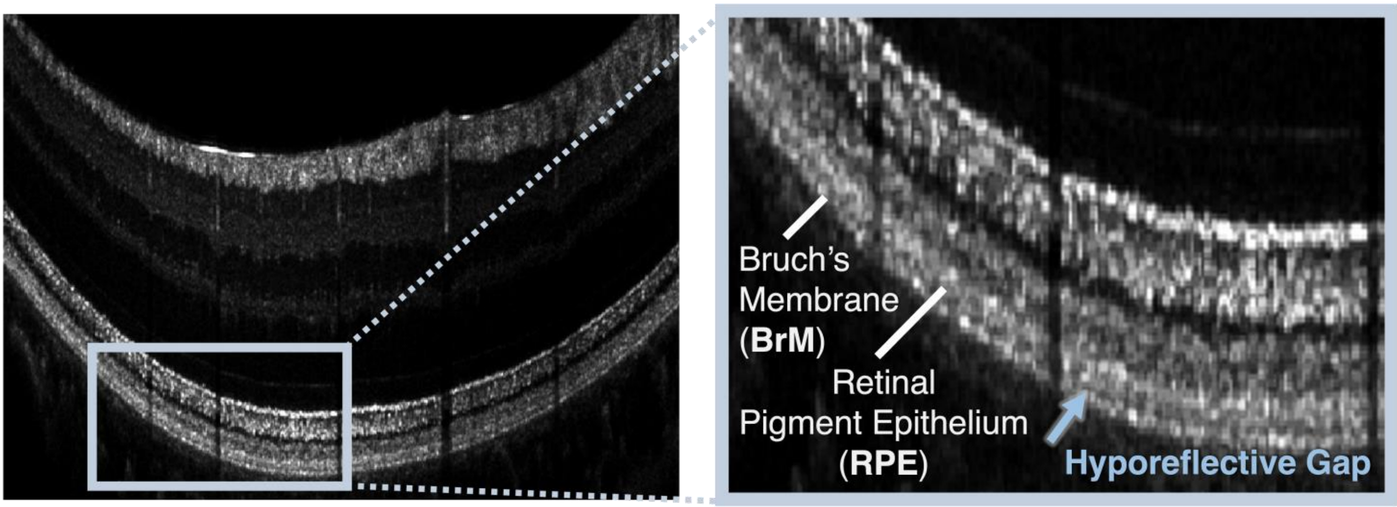
In search for ocular imaging biomarkers with inconsistent labels

by Christina Bornberg
@datascEYEnce

Hello everyone! We are back for another datascEYEnce story focusing on researchers who work in deep learning applied to ophthalmology! I am Christina and this time I want to introduce you to Wenke, whom I met at the German Conference on Medical Image Computing (BVM workshop) last month. Her presented poster "**3D Deep Learning-based Boundary Regression of an Age-Related Retinal Biomarker in High-Resolution OCT,**" caught my attention, and I knew it would be a perfect fit for RSIP Vision’s Computer Vision News magazine. That's why I reached out to her for an interview!

featuring Wenke Karbole

Wenke started her journey in deep learning for ophthalmology with a Bachelor’s degree in medical engineering at the TU Ilmenau. She found herself fascinated by a deep learning module and, therefore decided to enroll in as many courses as possible in this direction. During her Bachelor's Thesis at Siemens Healthineers, Wenke worked on deep learning applications for CT images. Transitioning to ophthalmology was a natural step forward while pursuing a double degree in medical engineering and AI at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), merging her passion for deep learning with her interest in photography and optical systems.



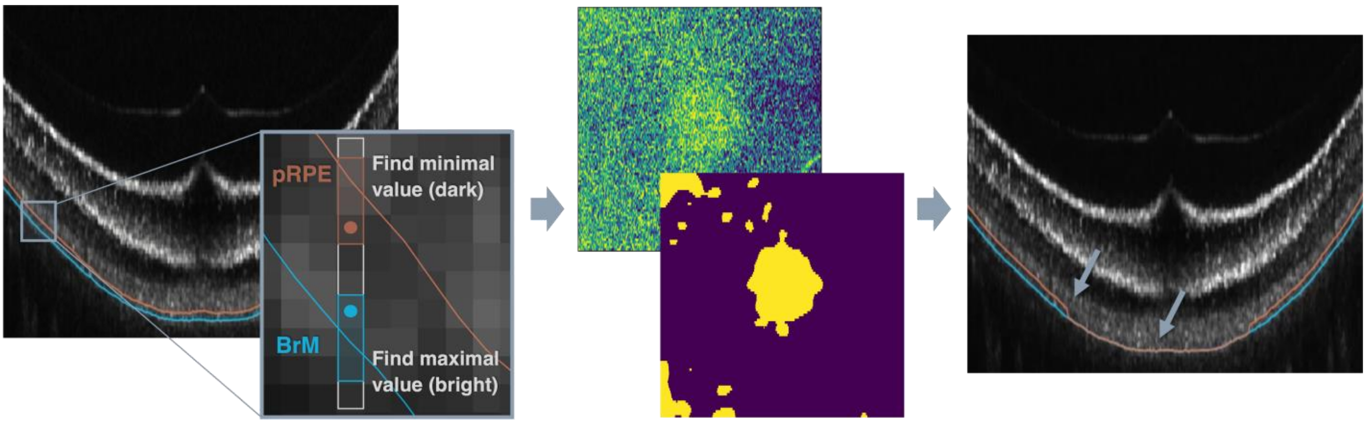
"Pictorial representation and vision are important characteristics in life, without which you are strongly impaired!"

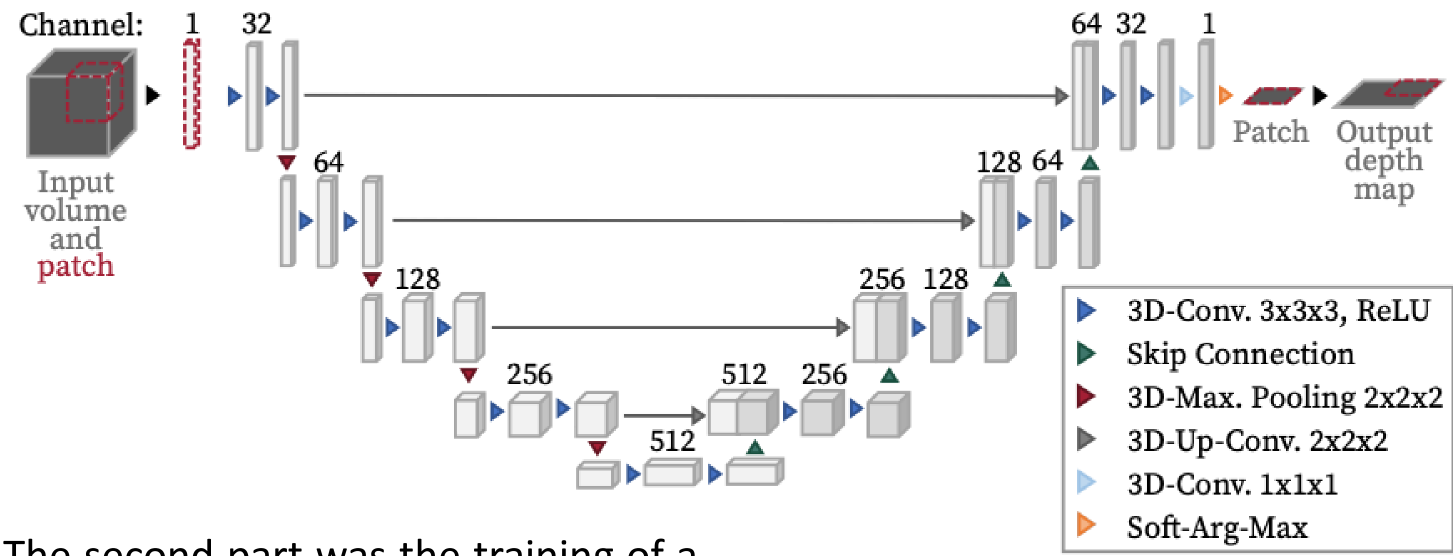
Moving on, Wenke did her master's thesis at the Pattern Recognition Lab at FAU which has a close collaboration with the Biomedical Optical Imaging and Biophotonics Group at MIT and the New England Eye Center in Boston. The collaboration made it possible for Wenke to stay in Boston during her thesis and gave her the perfect opportunity to work with experts in deep learning, optical device engineers as well as clinicians.

So, what did Wenke do during her time in Boston? It was everything from getting to know OCT imaging devices and understanding the image acquisition process, experiencing the imaging procedure, as well as reviewing, analyzing and refining data and labels - and obviously lots of coding.

Wenke's goal was to analyze a thin feature in OCT B-scans called the hyporeflective gap in the outer retina, a proposed biomarker for the common eye disease "age-related macular degeneration (AMD)". Since this gap is very thin, the data needed to be of high resolution and was pre-processed by her supervisor Stefan Ploner. Especially the merging of 6 volumes, motion correction, and layer flattening were important steps for removing inconsistency between the B-scans as well as enhancing the visibility of the hyporeflective gap.

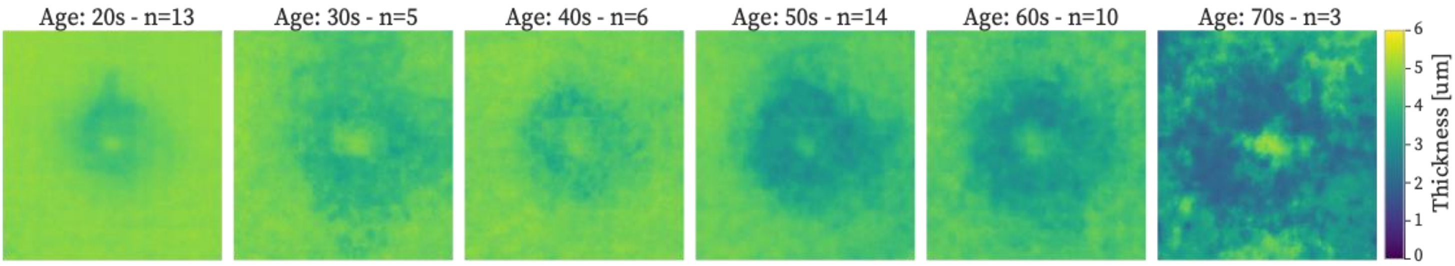
In order to analyze the hyporeflective gap, she first needed to derive a precise and reliable ground truth - one boundary for the posterior Retinal Pigment Epithelium (pRPE) and one for the Bruch's membrane (BrM), which together enclose the gap. A semi-automatic intensity-based label refinement approach was applied using initial labels from a 2D nnU-net approach. In the next step, the darkest (for pRPE) and brightest (for BrM) point in each A-scan from the initial boundaries were searched for. After smoothing and manually thresholding the binarised enface image, regions with no visible hyporeflective gap were found. With the increasing age of healthy subjects, this gap was found to first become invisible in the center of the retina, the parafovea.





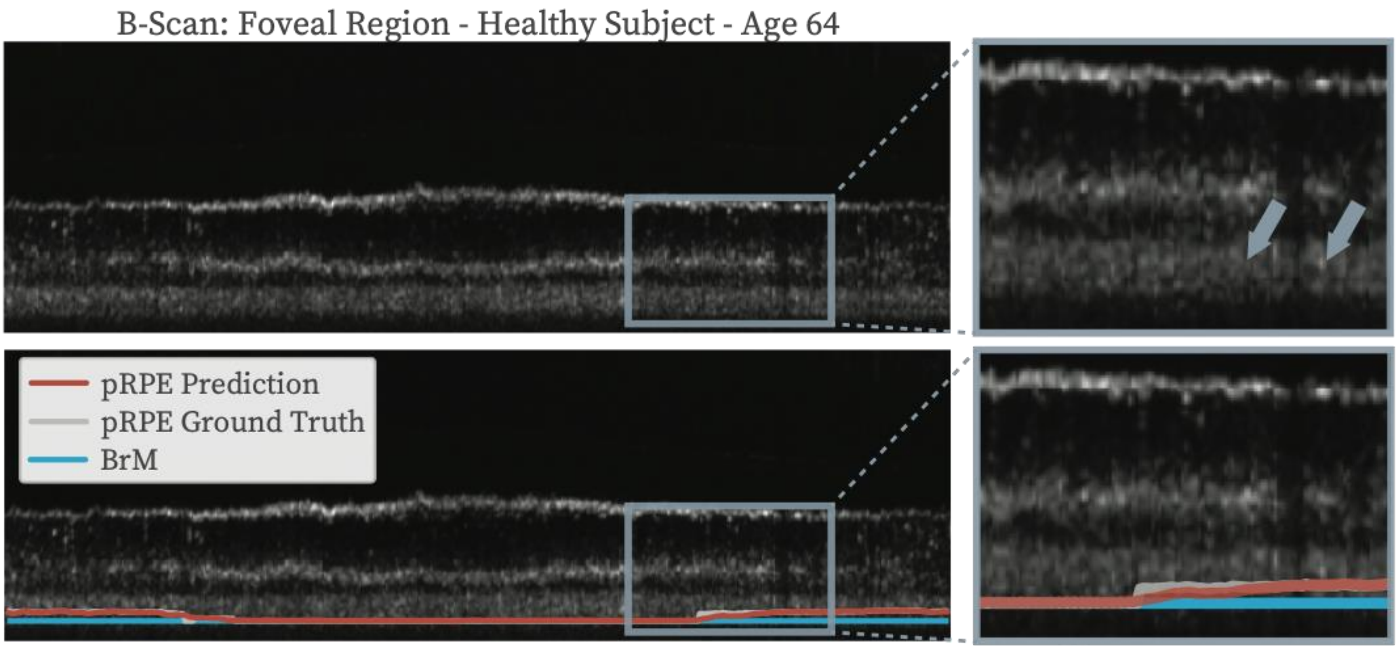
The second part was the training of a neural network for depth map prediction. Since previous work on Bruch's membrane segmentation is available, this part is assumed to be given. The novelty of her work lay in automatically detecting the boundary of the posterior Retinal Pigment Epithelium. She trained her depth map regression network (DMR-Net) with 3D OCT volumes as an input and a 2D enface depth map (estimation of the RPE depth) as an output. Due to the BrM layer flattening pre-processing step, the depth map can be interpreted as the thickness map of the hyporeflective gap.

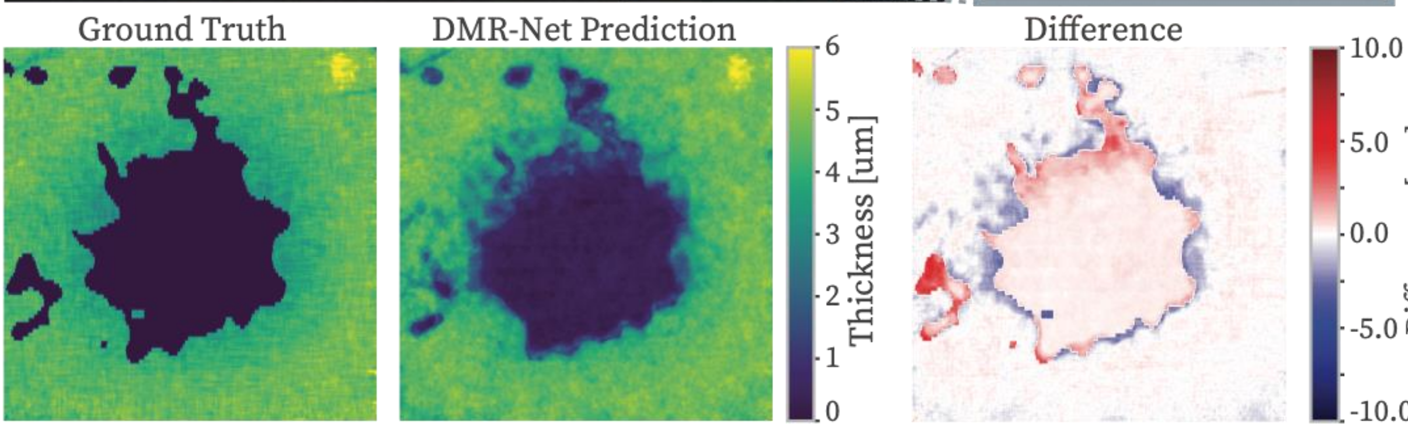
For the final evaluation, Wenke generated average thickness maps. Having access to a range of age groups, she **successfully showed the declining trend of thickness with increasing age and its spatial dependencies**. Future work will set a strong focus on the diseased eyes, hopefully proving an increased hyporeflective gap in patients with age-related macular degeneration.



When asking her for her personal lessons learned, Wenke spoke mostly about the labels having a strong inter- and intra-reader variability. It's a problem we all have experienced before, and in Wenke's case, it is even more problematic, as the feature to be annotated is so small and unclear. She needed to ensure that the network was as robust and consistent as possible. To achieve this, she focused on creating a consistent and meaningful ground truth and applied common techniques such as augmentation and regularisation to prevent overfitting.

I wish Wenke all the best for the future and hope to see her again with some more work in deep learning for ophthalmology at the next conference!



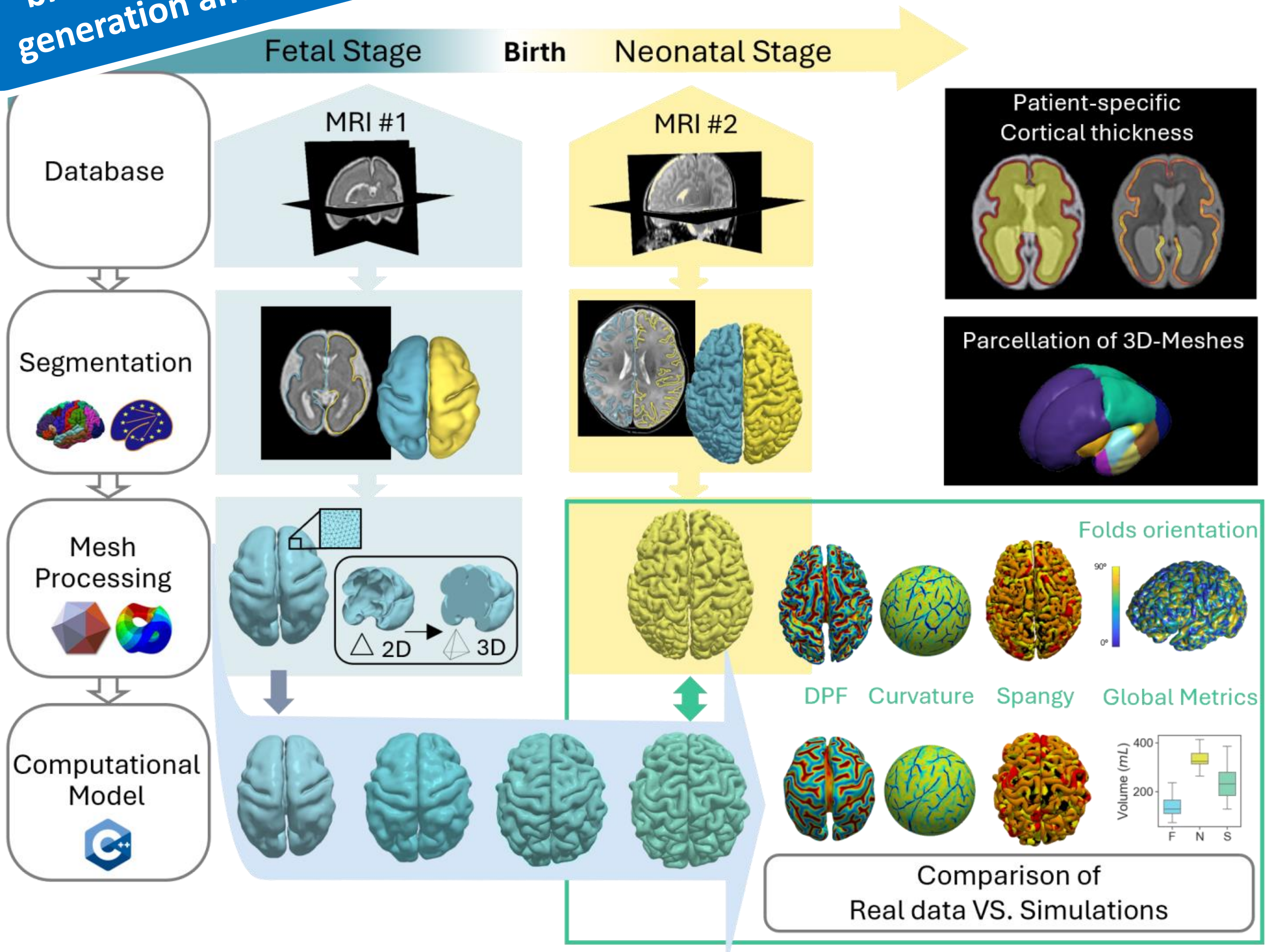


Mireia Alenyà has recently defended her PhD at Universitat Pompeu Fabra in Barcelona, under the supervision of Oscar Camara. Her thesis focused on fetal brain development analysis using computational modelling and deep learning techniques. She fostered collaborations with national and international partners, including clinicians, and adeptly utilized diverse computational techniques to construct a novel pipeline using unique fetal and neonatal data.

Part of her work was in collaboration with Gonzalo Maso from the Auckland Bioengineering Institute (ABI), New Zealand, where she undertook a research stay from December 2022 to April 2023.

Congrats, Doctor Mireia!

Scheme of the computational pipeline for patient-specific brain mechanical model generation and validation.



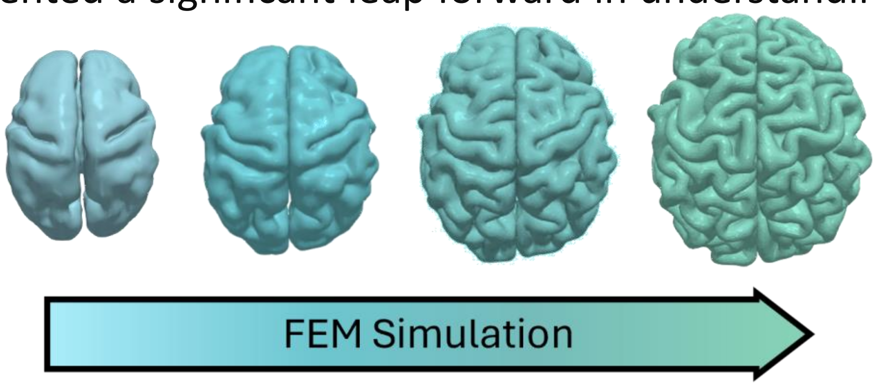
The human brain develops from a smooth cortical surface in early stages of fetal life to a convoluted one postnatally, creating an organized ensemble of folds. This is clinically relevant since abnormal folding patterns are linked to neurodevelopmental disorders such as autism, schizophrenia, or epilepsy. Recent advancements in MRI technology, that allow to collect fetal data at these early stages of pregnancy, have enabled researchers like Mireia to explore how essential **brain folds** form during gestation.

During her PhD, Mireia and colleagues constructed a **3D patient-specific computational pipeline** for analysing fetal brain development (see figure in the left page), beginning with personalised fetal and neonatal MRI scans. The pipeline consisted of segmentation of the grey matter tissue, construction and processing of 3D meshes, and incorporated a **Finite Element Method (FEM)** model that allowed for virtual simulations of brain growth and folding (see figure below), providing insights into developmental trajectories and potential outcomes. Additionally, several evaluation metrics were developed to quantitatively compare these simulations with real brain development data collected over time.

However, the computational complexity of this pipeline, including significant computational time requirements for some stages, led Mireia to the second part of her research. Here, she focused on accelerating these computationally expensive aspects of the pipeline using **deep learning** methods. This involved two main components:

Firstly, Mireia participated in the **Fetal Tissue Annotation Challenge (FeTA)** (2021 & 2022) organised within the MICCAI conference, to develop algorithms that automatically segmented fetal brain tissues within MRI scans. These algorithms were designed to adapt to various data types and even included techniques to artificially expand the datasets used for training (data augmentation).

Secondly, Mireia developed a **"surrogate" deep learning model** to accelerate complex FEM simulations. This model acted as a shortcut, learning to predict the outcomes of FEM simulations in a fraction of the time, significantly reducing computation time from 48 hours to a few seconds. Overall, Mireia's groundbreaking work represented a significant leap forward in understanding fetal brain development.



Simulation example of brain growth and folding.

"Datasets through the Looking-Glass" is a webinar series focused on reflecting in the data-related facets of Machine Learning (ML) methods. Our goal is to build a community of enthusiastic researchers interested who care about understanding the impact that data and ML methods could have in our society.

The webinar is part of "Making MetaDataCount" project and is organized by Veronika Cheplygina (left in the picture) and Amelia Jiménez-Sánchez (on the right) at IT University of Copenhagen. We had five successful editions so far between 2023 and 2024 with 19 speakers in total, the videos are available on our YouTube playlist.

In our last webinar one week ago (March 25), we covered several topics on dataset design and ML evaluation, including the importance of choosing the right metric for your analysis, which factors affect the creation of medical imaging datasets, and strategies to evaluate bias in skin lesion models. In this webinar, we had 5 excellent researchers providing insights about their work.

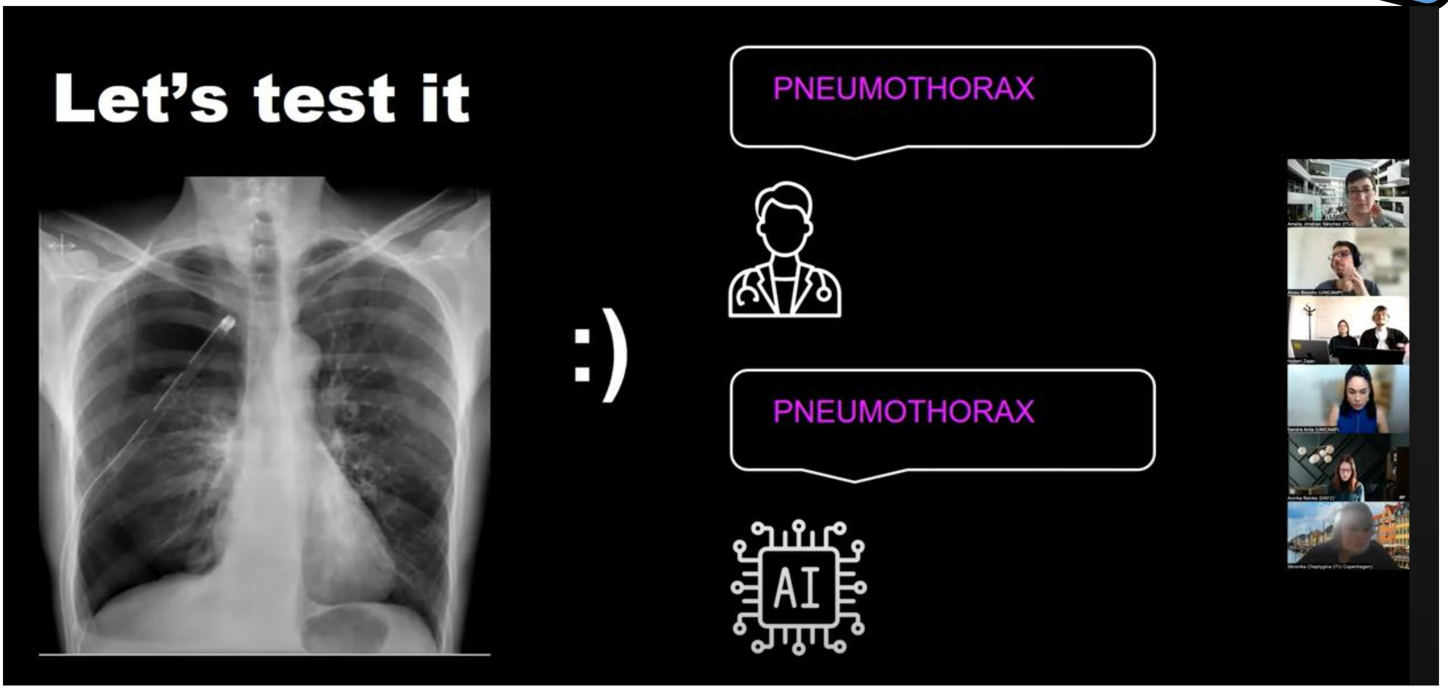
①

Hubert Dariusz Zając and **Natalia-Rozalia Avlona** were the speakers of the first talk. Their research is a critical inspection of **"ground truth" labels** when working with medical imaging datasets.

Hubert is a PhD student at the University of Copenhagen and a member of the Confronting Data Co-Lab, working on a healthcare AI project in Denmark and Kenya. Natalia is a lawyer and a Marie Curie PhD fellow (DCODE) at the Computer Science Department of the University of Copenhagen.

Their curiosity in this topic was sparked by works highlighting that algorithms with high reported performances have been shown to suffer from shortcuts, i.e. **spurious correlations between artifacts in images and diagnostic labels**. Their aim was to address questions regarding who

Check out the video!



decides what should be labelled, what definitions are used for the labels, and what is the relation between the data that is possible to be collected and the labels.

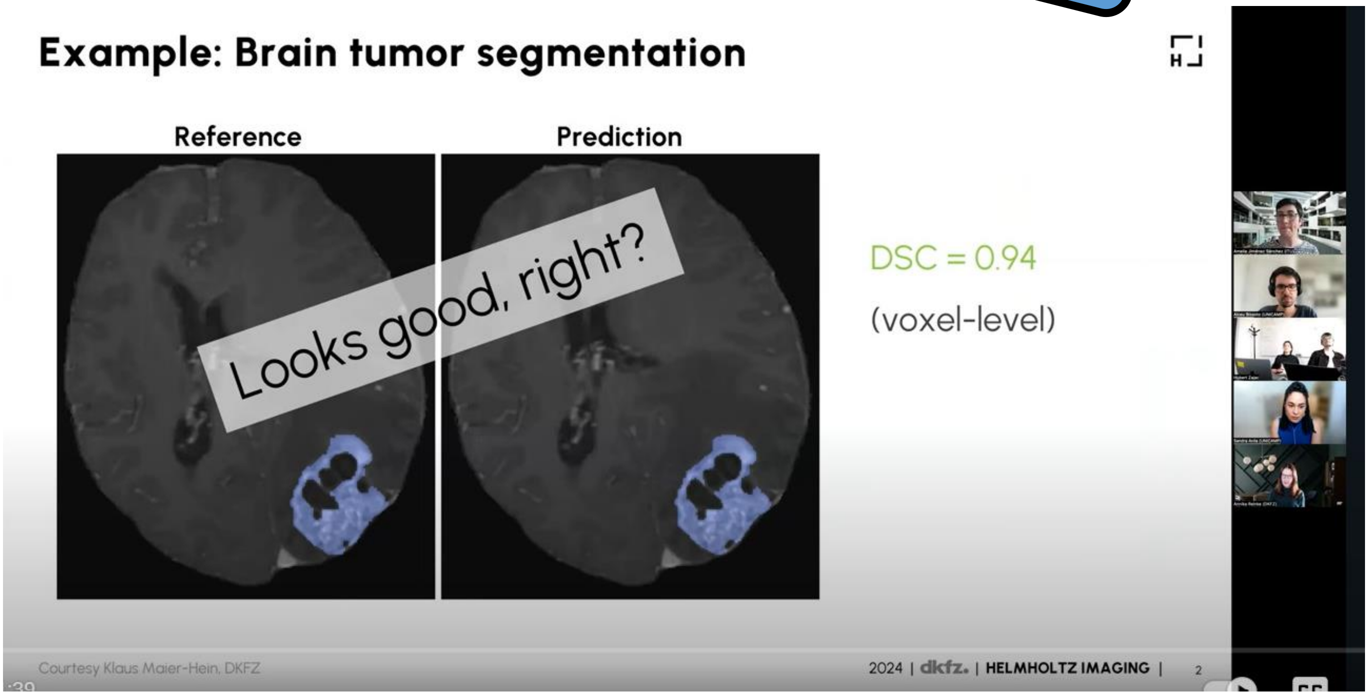
They employed **ethnographic methods** involving three health-tech organizations in Northern and Western Europe, and interviewed 46 participants from East Africa, Northern and Western Europe, including medical professionals, data scientists, designers, management, and other relevant stakeholders.

Through this process, they identified **five key external and internal factors affecting the creation of medical imaging datasets**. Three external factors include regulatory constraints, the context of creation and use, and commercial and operational pressures. Two internal factors comprise epistemic differences and the limits of labeling. They concluded emphasizing the importance of considering these factors when discussing what constitutes high-quality data. This consideration is crucial for ensuring **responsible AI design**.

②

Annika Reinke was the speaker of the second talk. Annika is the Deputy Head and Group Lead Validation of Intelligent Systems at DKFZ, Germany.

Check out the video!

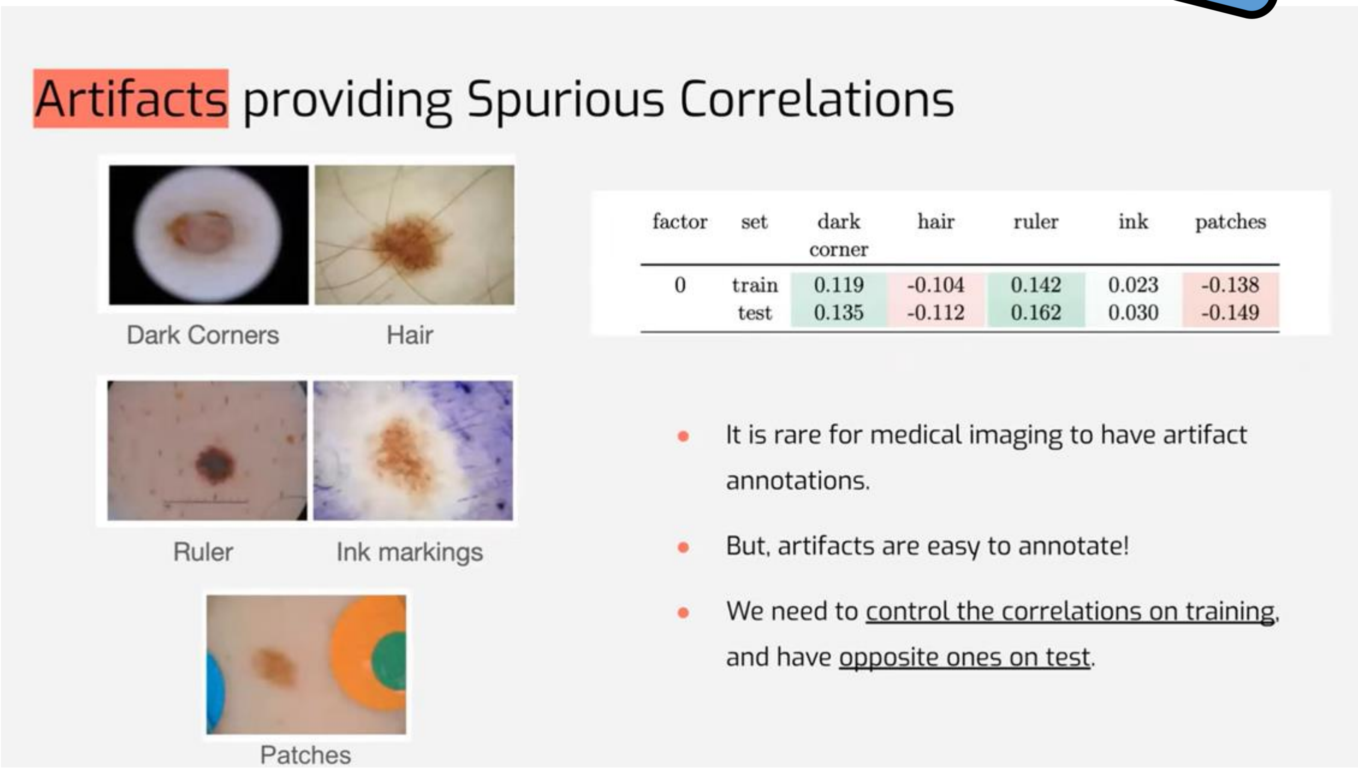


She presented **"Metrics Reloaded"**, a problem-aware metric recommendation framework resulting from 2.5 years of research involving a total of 73 international experts. The talk emphasizes the notion of **"problem fingerprinting"** - a structured representation of the biomedical problem capturing all aspects that are relevant for **metric selection**. Problem fingerprinting consider aspects such as tiny lesions, overlapping targets, uncertainty in the annotations, spatial outliers, etc. that are relevant to your research question. Based on these considerations, the framework recommends suitable metrics for your biomedical problem. There is now available **a python library implementation and an interactive online tool**; check it out!

③

The speakers of our last talk were **Alceu Bissoto** and **Sandra Avila**. Alceu Bissoto is a PhD student at the University of Campinas (Unicamp) and Dr. Sandra Avila is a Professor and Research Scientist at the Institute of Computing, Unicamp, Brazil. First, Sandra presented an overview of their group work on **skin lesions** over the last 10 years, covering areas such as **data augmentation, transfer learning and generative AI**. At this point, they found that their skin lesions were so good that they needed to start thinking about bias. Then, Alceu presented a strategy to evaluate bias on

Check out the video!

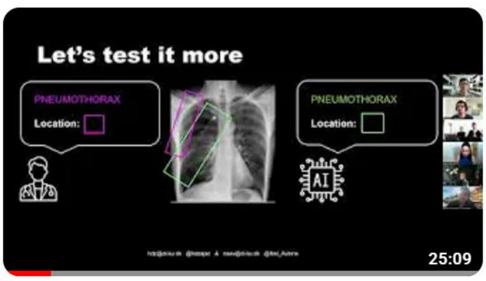


factor	set	dark corner	hair	ruler	ink	patches
0	train	0.119	-0.104	0.142	0.023	-0.138
	test	0.135	-0.112	0.162	0.030	-0.149

skin lesion models, in particular **“trap sets”**. Trap sets are artificially created biased sets with the training and test sets biased in opposite directions. The idea is that biased models “fall into the trap” and have poor performances on test data. Finally, as an example of debiasing solution, he presented their latest work: **Test-Time Selection (TTS)**. The idea is to introduce a human-in-the-loop that will provide positive/negative keypoints in the image to **encourage models to rely on robust features**. As a closing remark, he suggested a promising approach of test-time debiasing to restrict the model's decision-making without altering its pretraining.

If you want to stay updated about upcoming webinars and other events, sign up for our newsletter!

As a final note, good luck to Hubert and Alceu in their upcoming PhD defenses :)



Datasets through the Looking Glass - S05E01 - Hubert Zajac and Natalia Avlona

Datasets through the Looking Glass - S05E02 - Annika Reinke

Datasets through the Looking Glass - S05E03 - Alceu Bissoto and Sandra Avila

formerly Advanced Scanners

Doug Fairbanks
CEO

Aaron Bernstein
Founder and CTO

Doug Fairbanks is the CEO, and Aaron Bernstein is the Founder and CTO of VISIE Inc., a 3D computer vision start-up in Austin, Texas, working on spatial computing for surgical enabling technologies.

They are here to tell us more about its new 3D scanners set to revolutionize vision in the operating room.

VISIE (formerly Advanced Scanners) has a bold mission: to make the unseen seen in surgery. Recognizing surgeons' evolving needs, it has spotted a huge gap in the market for enabling technologies. *"We saw that there's a transition to surgeons making decisions about procedures, not on the approach or the features of the implants, but on the technology they use to do the procedure,"* Doug tells us.

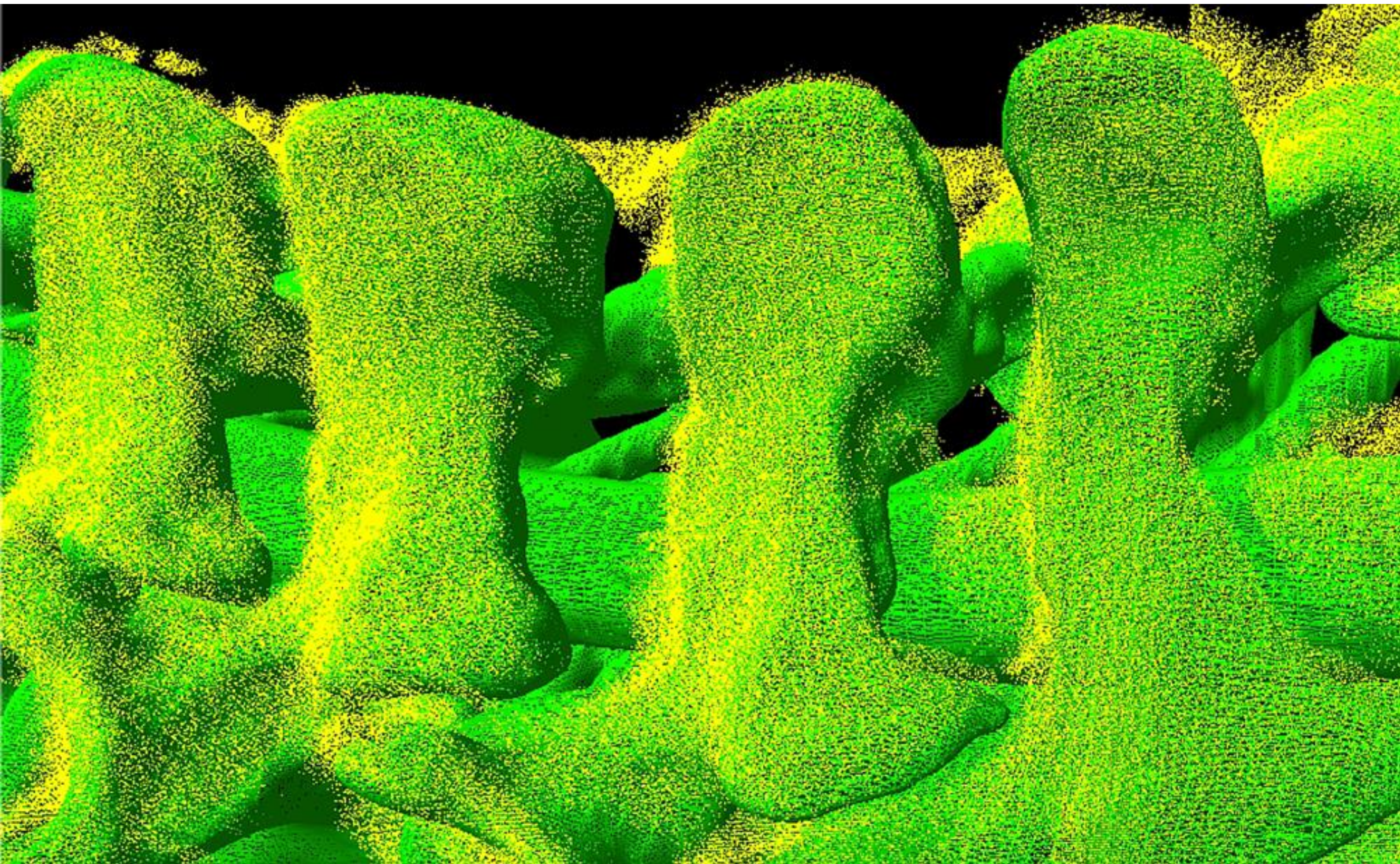
Current technologies in the market today, such as **robotic navigation and AR/VR solutions**, face inherent challenges due to traditional patient registration methods that use arrays and pins, impacting both the patient and surgeon experience. *"Surgeons are under greater pressure to do procedures more quickly and for less money,"* Doug continues. *"With **3D computer vision**, we reduce the procedure time and add some precision to these devices previously unrealized in the OR."*

Doug credits these improvements to Aaron's hardware innovation, offering **a transformative 3D scanning solution** with a very high signal-to-noise ratio that improves the accuracy and efficiency of surgical procedures. Having had the idea for the device and brought it to an advanced prototype state, Aaron is now helping to transition the technology into a product, explore new applications, and increase its capabilities. With patents in the US and worldwide, VISIE looks set to realize its dream.

"*One of the difficulties is figuring out exactly what we help first!*" Aaron laughs. "*We need to go forward in a way that allows us to maximize the device's potential while maintaining the kind of focus that a company needs. The surgeons we speak to in different specialties and doing different procedures see the value in what we do when they see the results of our scans. They want their machines to see in real-time the way they do. That's a technical challenge and one we're taking on.*"

The challenge is a considerable one. "*It's one thing to take an image; it's another thing to go fast and track in real-time,*" Doug explains. "*There's a very delicate balance because we gather loads of data. Where a procedure might take 350 data points on the end of a femur, **we can take 376,000 data points in a fraction of a second**. Everything is registered and tracked without the effort of adding extra arrays, putting fiducials in the patient, or going through a long and complicated process.*"

VISIE has bolstered its ability to quickly process large quantities of data to support this. Having joined **NVIDIA's Inception Program**, it has gained

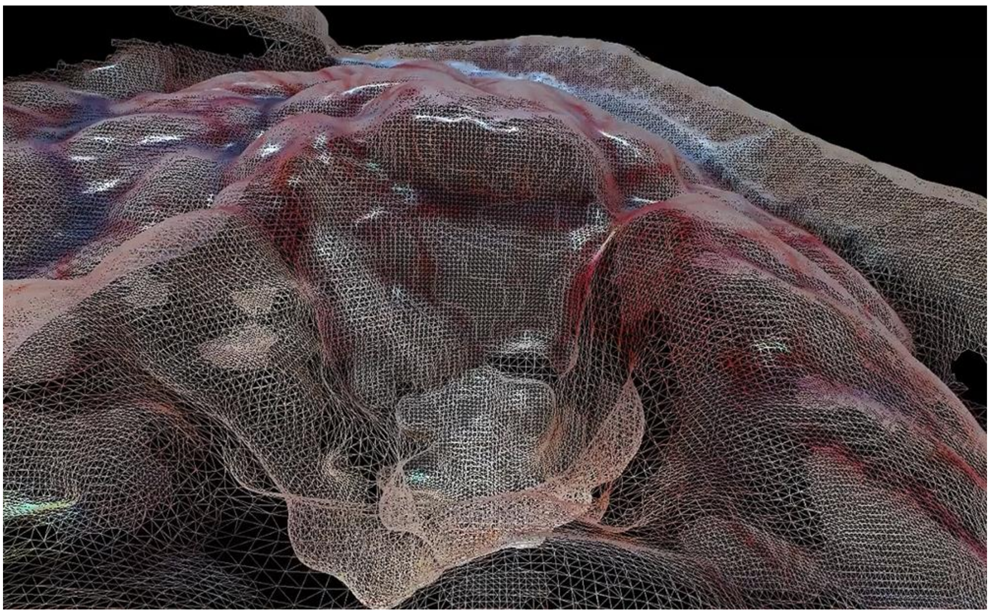


access to **cutting-edge GPU technology**. It can drop data through these processors at a very high rate, moving ever closer to its real-time scanning goal. *"We show our scans to people in the device industry, and they say these are spectacular images, and the quality is really compelling,"* Doug tells us. *"We show our scans to surgeons, and they're instantly convinced. Seeing is believing. Our belief as a company is there is more to see."*

The real-time visualization of surgical procedures has long been a difficult task. VISIE sheds some of these historical problems, with its scanner dropping some of the weight and complexity of a traditional projector and computer vision taking center stage. *"We more fully illuminate the scene,"* Aaron reveals. *"**We have an excellent signal-to-noise ratio in our scans**, and we've been doing them fairly quickly already. We excel in **short-distance scans**, in particular – a place where projectors often fall short. As a structured white light 3D company, we're moving forward well with that. We're marrying the technology with a grand vision of seeing the scene and seeing the patient, to connect those things with the navigation going on so that it is as aware as the surgeon."*

VISIE is investing a growing amount of time and effort into domain knowledge to understand what is happening in the scene. For example, what is the curvature of a femur? What does the ridge of a brain look like when exposed? What does the side of a lamina look like? *"Part of that is taking the high amount of data and distilling it into actionable intelligence,"* Aaron explains. *"This is doing what you need to at a bare minimum, so you don't have a huge amount of overhead in your offerings to provide the solution to the surgeon. We're talking about as lightweight a touch as we can bring to the OR, but to do the most value with what we bring in."*

Like any small start-up in the current climate, VISIE is already entering a

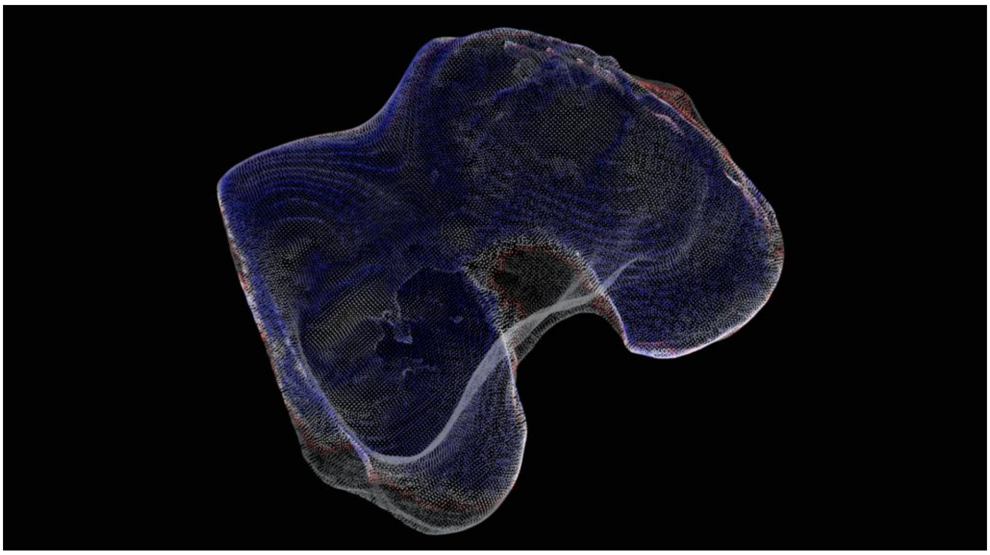


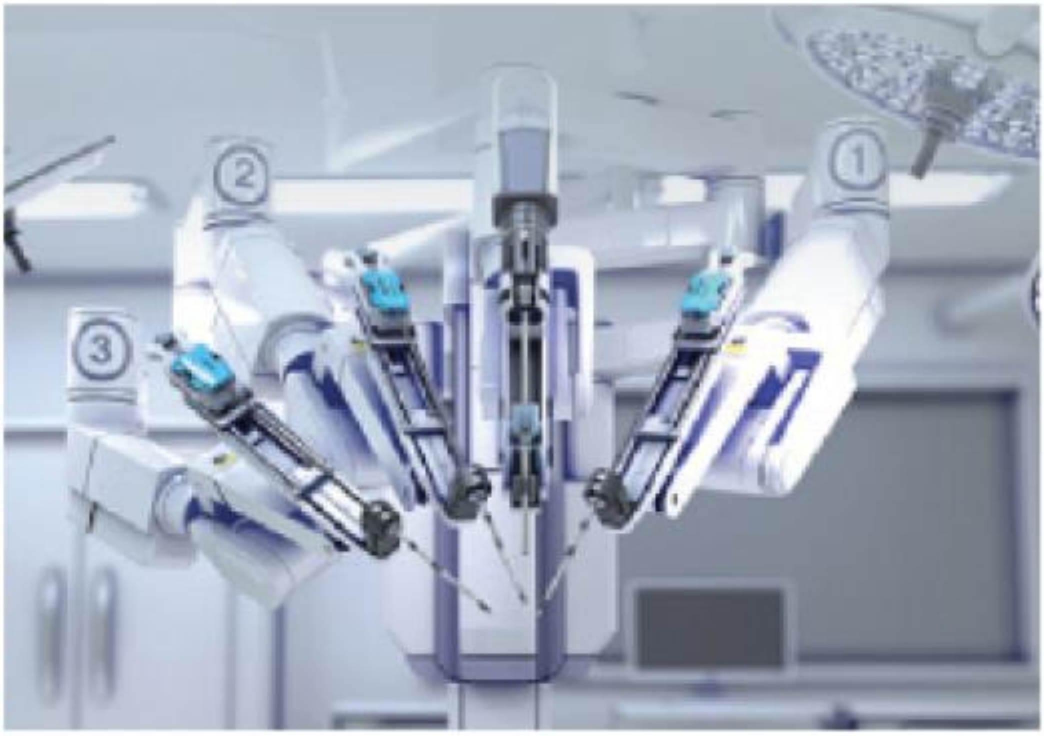
challenging environment, but being in the medical device and healthcare domains adds a host of other regulatory and market obstacles. "*We refer to our technology as **a deep tech solution**,*" Doug says. "*One that's on the bleeding edge of what's capable in science. Add medicine to that, and you add another level of complexity.*"

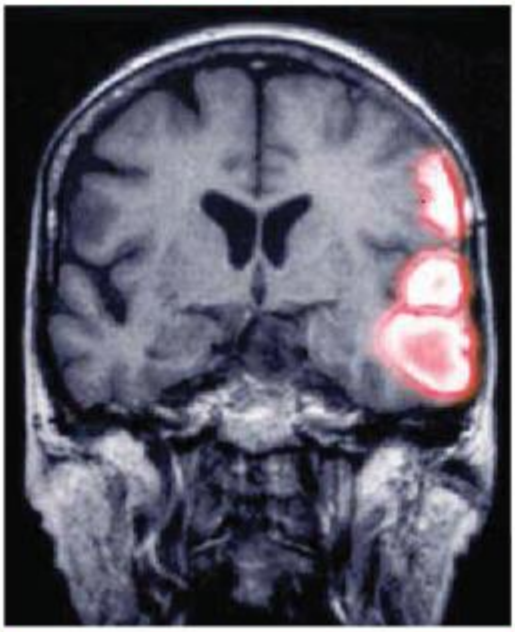
As a navigational medical device, VISIE's 3D scanner is considered lower risk as it does not touch the patient. Still, robust safeguards and validation processes are critical at every stage to ensure it correctly interprets a scene. The hostile environment of the operating room itself poses many challenges, including a high level of light intensity and the necessity of maintaining sterility. "*When you're operating any cameras, there are multiple complex elements that you're dealing with,*" Doug points out. "*It's simply untenable to turn the lights off to use our device. That's not in the patient's or surgeon's best interest. We've been really adept at finding meaningful solutions, and it's exciting to keep moving forward.*"

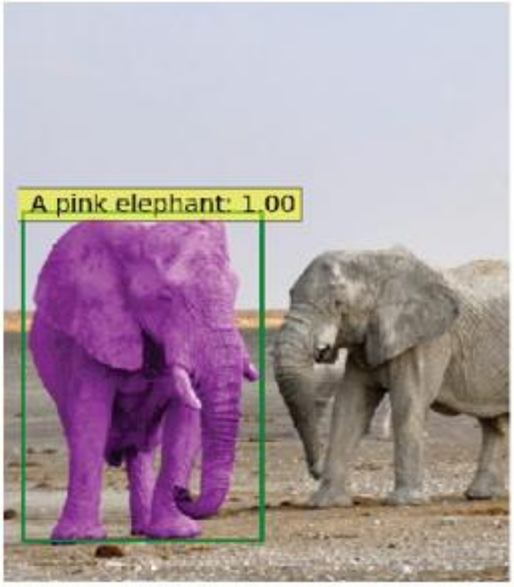
Creating a new device is one thing; convincing surgeons to use it and change their way of working is quite another. Aaron says it helps that **VISIE's technology seamlessly integrates into existing systems**. "*We're not developing this as another console to wheel into the OR,*" he attests. "*We'd like to augment existing systems for the registration of the patient, the reregistration of the patient, and eventually real-time tracking of the operation in 3D. As such, the surgeon will be looking at tools they're used to looking at. They will just be working better!*"

As VISIE continues to work toward its mission, there is one last question I am sure you are all dying for us to ask: Are they hiring? "*Always!*" Doug confirms. "*We're always looking for really talented computer vision and hardware and software engineers.*" The invitation is clear to prospective candidates: you still have a chance to join VISIE on its exciting journey...








A pink elephant: 1.00