

March 2024

# Computer Vision News & Medical Imaging News

The Magazine of the Algorithm Community

**New! Lessons in Robotics with Madi Babaiasl**

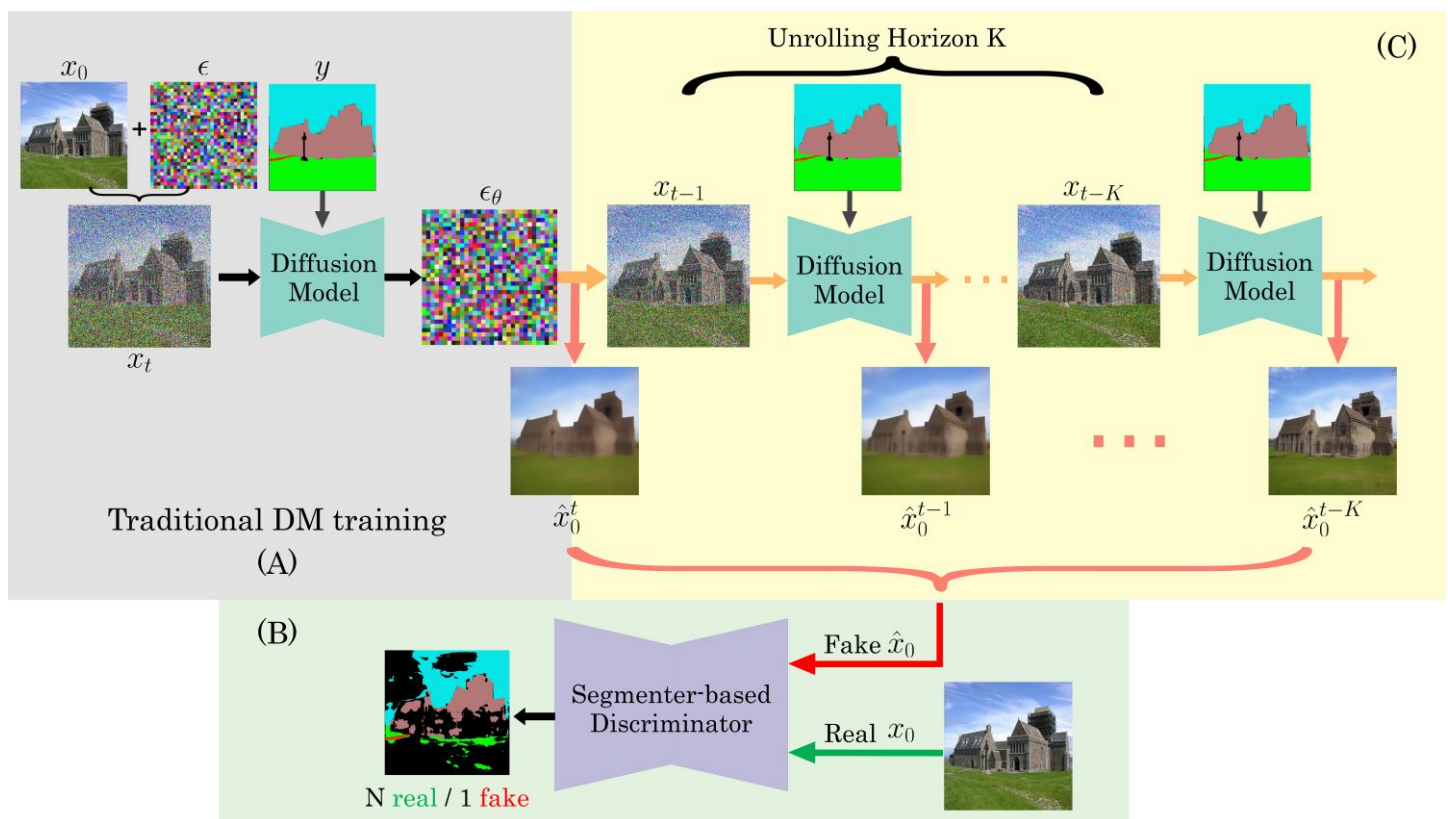
Hue circle showing the chroma component of the measured light source through different brands of polarization filters compared to the ground-truth (no pol).

## Adversarial Supervision Makes Layout-to-Image Diffusion Models Thrive



Yumeng Li is a PhD student at the Bosch Center for Artificial Intelligence (BCAI) under the supervision of Anna Khoreva and Dan Zhang from Bosch and Margret Keuper from the University of Mannheim.

Yumeng is here to tell us about the paper's novel approach to training for layout-to-image (L2I) diffusion models, which has been accepted as a paper at May's ICLR 2024 conference in Vienna



In this paper, Yumeng and the rest of the all-female team behind this work explore image generation given layouts such as **semantic label maps**. The idea is based on diffusion models. Previous training approaches for diffusion model pipelines relied on mean squared error (MSE) reconstruction loss, overlooking the explicit consideration of layout conditions.

Furthermore, when employing diffusion models for image generation, it is necessary to undergo an iterative denoising process. However, previously, people only considered one single denoising step, disregarding the significance of this iterative approach.

*“Before, people tried to improve the architecture of the networks,”* Yumeng tells us. *“They didn’t pay too much attention to the training pipeline, like the training loss and objectives. They introduced **ControlNet**. Basically, you add another branch for incorporating the conditions, but they used exactly the same training objective as the previous diffusion models.”*

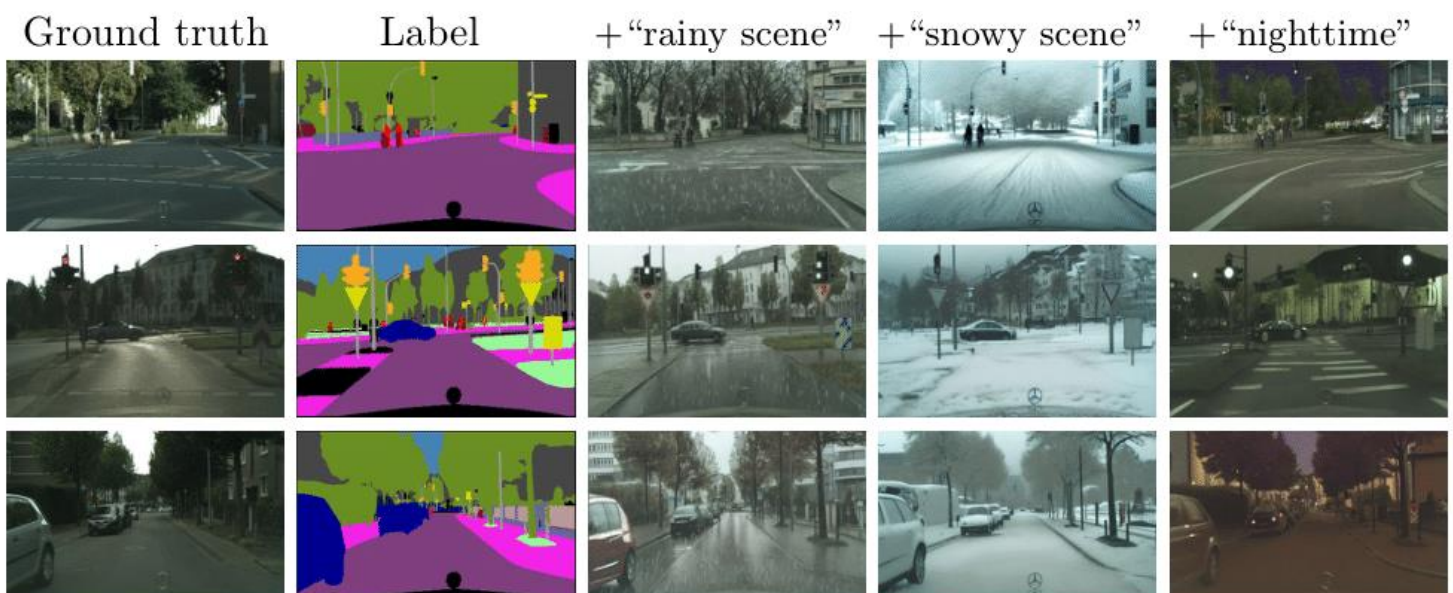
To mitigate these issues, the team proposes to **integrate adversarial supervision into the training pipeline of L2I diffusion models**, leveraging a layout condition to encourage conditional alignment. A segmentation network-based discriminator guides the diffusion model training using semantic label maps as a supervision signal. The diffusion model generator is encouraged to follow the label map explicitly.

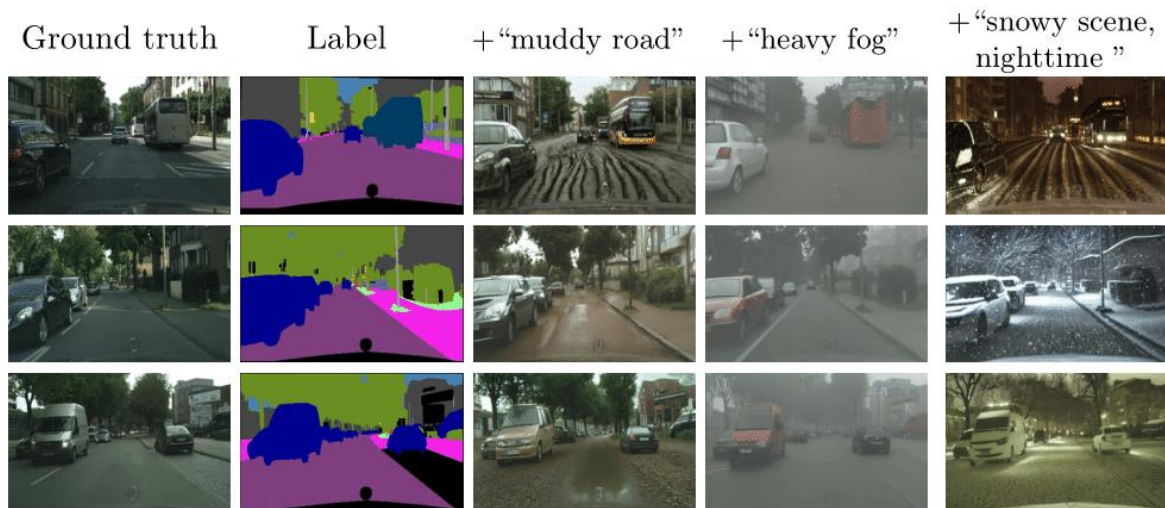


“We also propose a *multistep unrolling strategy*,” Yumeng explains. “Instead of just applying the loss at one single timestep, we unroll according to the iterative denoising process at the inference time. We mimic this process during training to encourage alignment across different timesteps.”

**Instability during training** posed a significant challenge due to introducing a discriminator and varying noise levels at different timesteps. To address this, the team proposed **sparse unrolling**, applying the strategy intermittently to reduce training costs while maintaining stability.

Looking ahead, Yumeng emphasizes the broader applicability of this approach beyond L2I generation. “*Multistep unrolling is not specific to layout-to-image generation*,” the researcher points out. “It can be generally applied to improve diffusion model training. As diffusion models apply an





*denoising process during the inference time, single-timestep supervision may not be sufficient in many cases, such as text-to-3D or text-to-video diffusion models. It’s crucial to apply multiple timesteps during training.”*

Originally from Tianjin, China, Yumeng has so far explored generative models, from the first PhD project on generative adversarial networks (GANs) to the current focus on diffusion models.

In addition to the technical advancements, this work aims to bridge the gap between generative models and real-world applications. Leveraging synthetic data generated through its methodology demonstrates notable improvements in tasks like semantic segmentation.

***“ControlNet can be controlled by text information; for example, we can generate snowy scenes and rainy scenes, and it’s quite flexible,”*** Yumeng adds. *“We generated diverse images and then applied this to the semantic segmentation task, showing that it can significantly boost the generalization of the segmenter. People put a lot of focus on generative models, but quite often, they just show beautiful images without mentioning how to use this synthetic data for real-life applications. You can have some fancy images just for fun, but eventually, we want to use them to improve downstream models. That’s quite important.”*



# Vision-aided Screw Theory-based Inverse Kinematics Control of a Robot Arm Using Robot Operating System (ROS2) - Part 1



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. Here is part 1.

Parts 2 and 3 will be published in the April and May 2024 issues of Computer Vision News. You will find them [in our archives](#) after April and May 2024.

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

## 1 Introduction

This lesson is the first part of a series of four lessons that integrate the principles of screw theory-based numerical inverse kinematics with vision-based control of robotic arms using the Robot Operating System (ROS2). It is designed to provide an understanding of how to leverage the mathematical foundations of screw theory-based numerical inverse kinematics in robotics alongside the functionalities of ROS2, to control a robot arm through visual feedback. In this lesson, you will learn 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 the following lessons. Note that you do not need to have the exact hardware and the proposed method can work on any ROS2-controlled robot arm and it is not contingent on the specific hardware used.

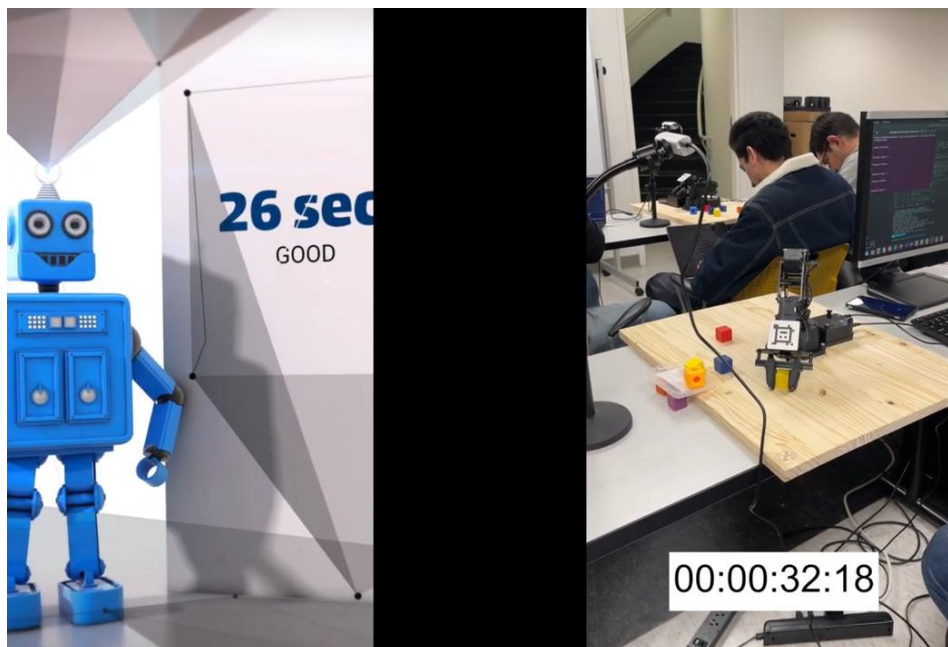
## 1.1 Objectives of the Lesson Series

By the end of these lesson series, participants will be able to:

- Understand the fundamental concepts of screw theory and its application in robotic numerical inverse kinematics.
- Set up and configure the robot arm and its vision kit for ROS2 and RViz.
- Implement numerical inverse kinematics solutions using Python and the ROS2 framework to control the robotic arm.
- Integrate and utilize a vision system with ROS2 to enable object detection and manipulation tasks.
- Troubleshoot common issues related to vision-aided robotic control and apply best practices for system optimization.
- Recognize the limitations of traditional perception methods and the potential of deep learning techniques for enhancing robotic vision.

## 1.2 Implementations by Two Groups of Students

Two student teams have successfully applied the concepts outlined in these lesson series, with their implementations showcased in the video below.



This practical demonstration serves as a testament to the adaptability and applicability of the theoretical framework provided. Readers are encouraged to customize the provided code to suit the specifics of their experimental designs.

## 2 Setting up the Hardware and Software: Robot Arm, Vision Kit, ROS2, RViz, and Python-ROS API

The robotic arm used for this lesson is the serial robot arm, [the PincherX 100](#) from Trossen Robotics with [the vision kit](#) for vision-based control. This robot arm is controlled by Robot Operating System (ROS) and we will use [ROS2 Humble](#) which is the eighth release of ROS and it is a long-term support (LTS) release meaning that it will be updated and the bugs will be fixed until May 2027. The operating system that we run ROS2 humble on is Ubuntu 22.04 and you can find the [guide to its installation here](#). We will use [RViz](#) as the visualization tool to visualize the state of the robot. Python is the preferred language and we will use the Python-ROS API which sits above the ROS so that even if you are not proficient with ROS, you can still control and program the robot arm. The complete instructions on how to get started with this robot arm and the vision kit including assembling, installing the required packages, ROS2 and RViz are in the following links: [assembling the robot and the vision kit](#), [ROS2 and RViz installation](#), and [installing Python-ROS API](#). If you are not proficient with Python, [a quick introduction to Python programming for robotics](#) tutorial can get you started.

## 3 Summary

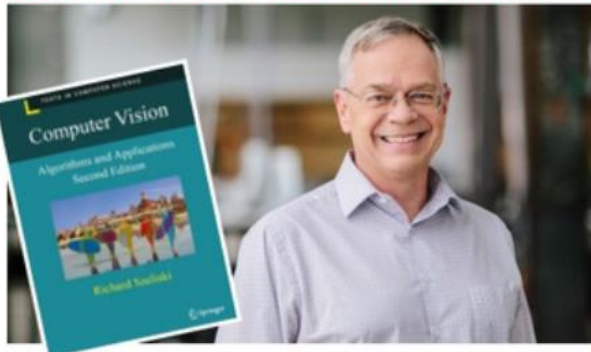
This lesson served as the introductory part of a lesson series focused on integrating screw theory-based numerical inverse kinematics with vision-based control of robotic arms using ROS2. It guided learners through setting up their hardware and software, emphasizing the applicability of the teachings to any ROS2-controlled robot arm. Instructions for assembling the arm, installing necessary software, and links for further guidance are provided, ensuring learners are well-prepared for the coming lessons ahead.



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

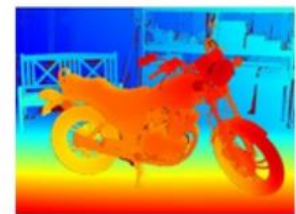
book sales. Fortunately, Springer said, "Sure, we can do that!" I was working with the same great editor, Wayne Wheeler at Springer, on the second edition and it was just a given that we would do it again.

**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.



## 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](#).



Tarek A. Haila has just completed and defended successfully his PhD dissertation at the Technical University of Darmstadt in Germany.

His research work was in collaboration with Fraunhofer IGD where he worked on the CultArm3D scanning technology during the past 3 years (June 2020 - May 2023).

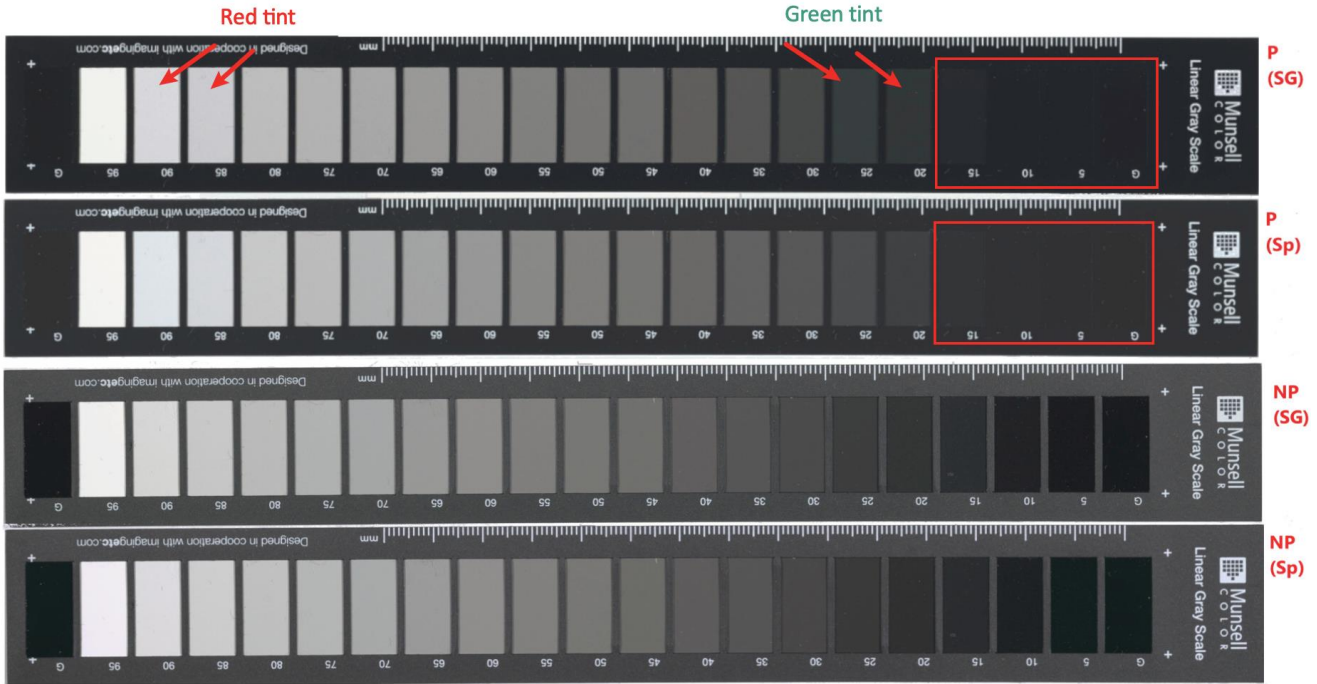
Tarek is a recipient of the prestigious Marie Skłodowska-Curie scholarship. He is very enthusiastic about color perception and reproduction as one can tell from the set of 3D printed color targets he holds in the photo!

The *CultArm3D* is a fully automated scanning robot equipped with a high-end camera (PhaseOne iXG 100MP) and a D50 light source mounted directly on the camera lens. Due to this geometry, the *CultArm3D* uses cross-polarization filters (CPF) to eliminate potential and undesirable reflections. It turned out that the literature lacks detailed coverage of the impact of CPF on color imaging and reproduction accuracy, for which Tarek found the incentive and the opportunity to conduct systematic research in this area helping *CultArm3D* in its mission of capturing accurately color information of virtually any 2D/3D scanned object.

Tarek formulated and addressed 3 research questions (RQs) in his publications that were presented at different conferences of the society of *Imaging Science and Technology*

(*IS&T*). The first question focused on examining the impact of cross-polarization and RGB imaging on color reproduction. It has been found that cross-polarization leads to an unavoidable loss of dark shades (deep blacks; CIELAB(L\*) < 20.0 on a Munsell Linear Grayscale), as well as causing an increase in the perceived lightness (CIELAB(L\*)) for matte materials, resulting in the perception of washed-out colors.

In his second research question, Tarek investigated the influence of CPF on illumination, for any changes in the characteristics of the light source would inevitably impact the reproduction of surface colors of objects. The findings revealed that different CPF brands affect the chroma components of a light source, also known as the white point, in different ways,



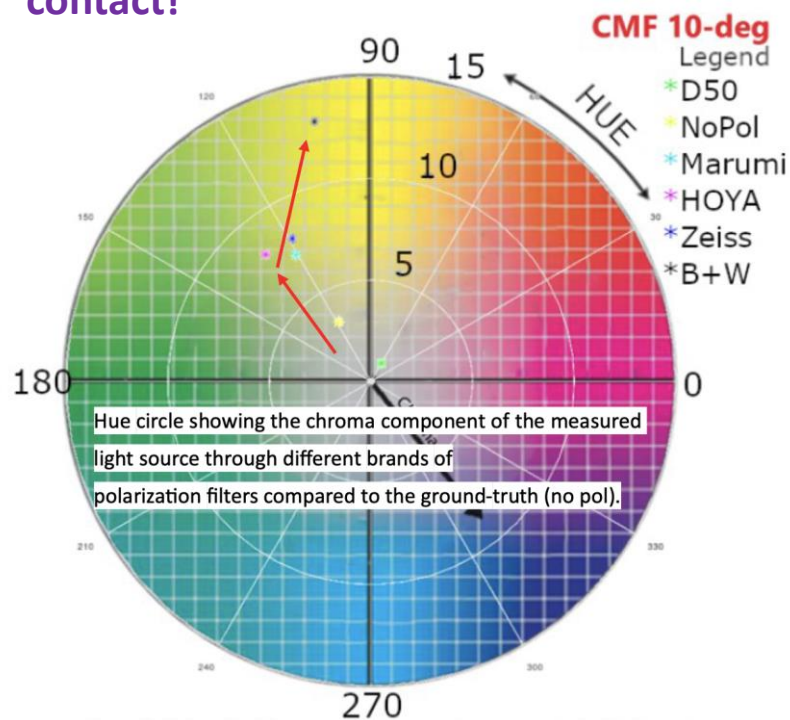
Munsell Linear Grayscale (MLG) color corrected using either Spyder24 (Sp) and X-Rite SG (SG) under two conditions, with (P) and without (NP) cross-polarization, our pipeline IQA.

consequently leading to color-correlated temperature (CCT) variations and resulting in color difference  $3.5 < DE00 < 8.0$  and a worst-case scenario shift of 360 Kelvin in CCT.

For the third research question, Tarek faced two challenges: dealing with highly reflective materials despite using CPF, as well as addressing the issue that comes with small field of view (FOV) of the camera when using a Macro lens and a short focal distance while aspiring for high-precision color calibration model that requires a large color target that does not fit in such a small FOV. The novel mosaic stitching approach, he came up with, overcomes the constraint and size limitation for matching a color target to a specific lens FOV and/or focal distance, enabling a more accurate color calibration model. Furthermore, this method resolves the issue of handling highly

reflective materials by consistently extracting and selecting only the clean, clear, and sharp parts of any scanned surface, similar to how texture atlases and UV maps work in 3D, paving the way for 3D color calibration.

**Congrats, Doctor Tarek! Many thanks to Arjan Kuijper for making the contact!**



## A History of Fake Things on the Internet



Author Walter Scheirer is the Dennis O. Doughty Collegiate Associate Professor of Computer Science and Engineering at the University of Notre Dame and a dear friend of Computer Vision News.

His new book introduces readers to a creative side of computer vision that users love but one which has, he says, undeservedly received a bad reputation in the tech press.

Walter is here to tell us more and set the record straight

**What is your new book, *A History of Fake Things on the Internet*, about?**

It's a really fun history that goes back to the Internet's early days about all of these fictions we see on the Internet. What I think is most interesting about this book is that it's not the standard technology ethics tale about all this fake stuff is bad. We need to get it off the Internet. We can't have anything that is fictional on the Internet. This is really a story of human creativity. Why do people like telling stories, especially on the Internet? What's the deal with all that creative software that it feels like computer vision has been developing over the years? Now, it's mainstream, people are using it. Why do they use it in the ways they do? How do we

understand that media landscape?

**How were you able to find things that others missed before you?**

I was able to have a different perspective on this. Number one, I have a background in computer hacking [*he laughs*]. I knew about underground communities that were using the Internet at an early period. I knew of some interesting people to talk to who were there at the very beginning. They led me to interesting files that were still in the Internet Archive. They led me to other people that I did not know that had interesting things to tell me. I was able to become a historian, based on that background, fairly easily compared to a mainstream academic or journalist.

## **Why is this particularly interesting for the computer vision community?**

When we look at the news in terms of what's being covered with these new technologies, there's a lot of concern about generative AI. If you look at the writing around tools like Midjourney, DALL·E, and even the emerging stuff coming out of the research community that's published at CVPR, it's very negative. It's like these things are being used to create photorealistic scenes that will fool people, and that's the primary reason people are creating. That's not true at all. I think all of us in computer vision know this. We're working on these tools because we know people really enjoy creating art with them, telling stories with them, creating games with them, virtual environments. The use cases, as far as I can tell, are basically positive. Sure, there are a few bad actors out there, but that's not the mainstream.

## **Can you tell us a bit more about the history of all this?**

I think what will really intrigue the computer vision audience is the book's coverage of the history of photography. Obviously, the whole field is based around digital images – those can be still images or video – but I think what's surprising to many people is that we don't really think about where the camera came from or who was using the camera. The book, for instance, looks at how photo editing came to be.

A really interesting anecdote: the camera was invented in the mid-19th century. What's interesting is you think, okay, this is very crude technology. It's very basic. People were just basically capturing images, and that was it. That's not true at all. As soon as the camera is invented, one of the early inventors of the camera process fakes the first photograph, and it just spirals from there. Darkroom photographers get really good at editing photos, like editing the negatives. They're able to create effects that are very similar to facial filters on the Internet. If you're using Instagram, you can clean up your complexion and add all sorts of fantastic effects. You do that in the 19th century, too. This was well known. It was a manual process. Some technician in the darkroom had to modify the negative, but it was pretty standard. In fact, it was so standard that when you'd sit for a portrait in the 19th century, just as an ordinary person, you would often be presented with a catalog of special effects. You could order it as an add-on to your photos. It just goes from there. None of this is new. People love the modifications. That's part of the fun of it.

## **Does this mean the ancient photos we see today may not be entirely loyal to reality?**

Correct. In fact, the camera was really never intended to capture some objective reality. Again, there are so many ways to modify a photo. The book looks at a lot of

thinking around this. These debates are not new. They go all the way back to the invention of the camera. For instance, the filmmaker Errol Morris, who's a really famous documentarian here in the United States, he was asking these questions about 15 years ago in essays he published in the New York Times. He concludes that photographs are really a social object. They're only useful so far as you can ask questions. You can interrogate them. Those answers may be different, depending on the context. You're never going to get some kind of objective truth out of it, which I think is a really important thing to understand.

***"It's a very fun book!"***

**With so much content on the Internet, how are you able to ascertain what is fake and to what degree it is fake?**

A lot of my research in computer vision image processing has been about media forensics. It's like, what's the integrity of this image? Has it been altered in some way? Can you determine if it's from a generative model? Developing algorithms like that. Now, what I think is interesting is that when we started to actually look at the Internet in more recent years using these algorithms basically everything is coming back as fake. I think there's some interesting reasons. Computer vision people

sort of know why I think right away.

Think computational photography. The smartphone, my iPhone, has a really sophisticated AI pipeline. When I take a photo, I'm not really getting the raw pixels right off the sensor. What I'm getting is an AI reconstruction of what the camera thinks the scene should look like. It's usually cleaned up. The lighting is being adjusted. There are lots of tricks you can play with super-resolution to get a much higher-quality image from a low-quality sensor. The lenses are cheap, so you're going to correct for that. There are all of these things that are sort of textbook built into the cameras, but those are all modifications to the image. A lot of the classical techniques for detecting forgeries are flagging these images like they're altered.

The other thing we found is there's a ton of fake stuff that's obviously fake, like memes. Memes are edited images, but we love that. It's not like it's a secret that it's edited. That's why they're effective. That's so much more prevalent than anything that can be considered a photorealistic fake. That's really interesting because why are we developing all these algorithms for detection when, at the end of the day, we know everything has been altered in some way? That's the Internet.

**As an author, do you like writing about things connected to but separate in some way from your day-to-day research work?**



Yeah, exactly. Why did I write a history instead of another technical paper? I think what's interesting about computer vision is that we're having a lot of discussions about its impact on society, and you can't really talk about that in a CVPR paper [Walter laughs]. CVPR reviewers, they're not going to accept that. You just don't have the room to tell that story. I thought it would be a fun project to become a historian of technology myself because I know how the technology works, so I can have an advantage when I'm writing about it from a social perspective because I'm not going to misunderstand its capabilities. I have deeper connections to the technologists who created the technologies in many cases, and I just wanted to

explore the social aspect in a non-conventional way. The book was the way to do that. Luckily, I found a publisher that was very supportive of that. Notre Dame was very supportive of this project, too, knowing it would be important. It's very different from what ordinary computer vision researchers do, but I think that's important.

### **Do you have plans to delve further into this topic in the coming years?**

Yeah, absolutely. This is not just a one-and-done book. Given especially the very positive reception and also the enormous interest in this project, I think I'm going to be working more in this direction in the future. There's a huge opening. I feel like computer vision is just developing the algorithms and publishing the paper. The more we interface with the public, the more we're going to need this sort of work to communicate what is actually happening.

### **Is there anything else our readers should know?**

I think the computer vision community is going to love the book. I've already received a lot of comments from people who have been reading it, and they really enjoy it. There's a lot of new stuff, a lot of new storytelling. There are many original interviews. This is not just a superficial look at this topic; it's very deep with a lot of new stuff. I think people should just have fun with it. It's a very fun book!



SPIE

@SPIEtweets



“I will tell you the secret of SPIE: they save the best talk for last!”

The final keynote speaker of [#SPIEMedicalImaging](#), [@ShuoLi49764724](#) of [@CWRUBME](#), takes the stage to discuss the impact of vision-language AI models in medical imaging. ✨

*“I will tell you the secret of SPIE:  
they save the best talk for last!”*

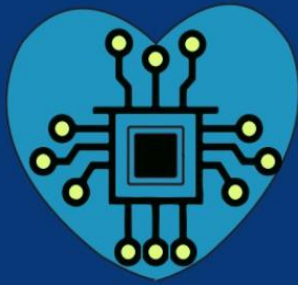


Shuo Li is AI-in-imaging Professor at Case Western Reserve University and a dear friend of our magazine Computer Vision News. He was Keynote Speaker a few days ago at SPIE Medical Imaging in San Diego, CA.

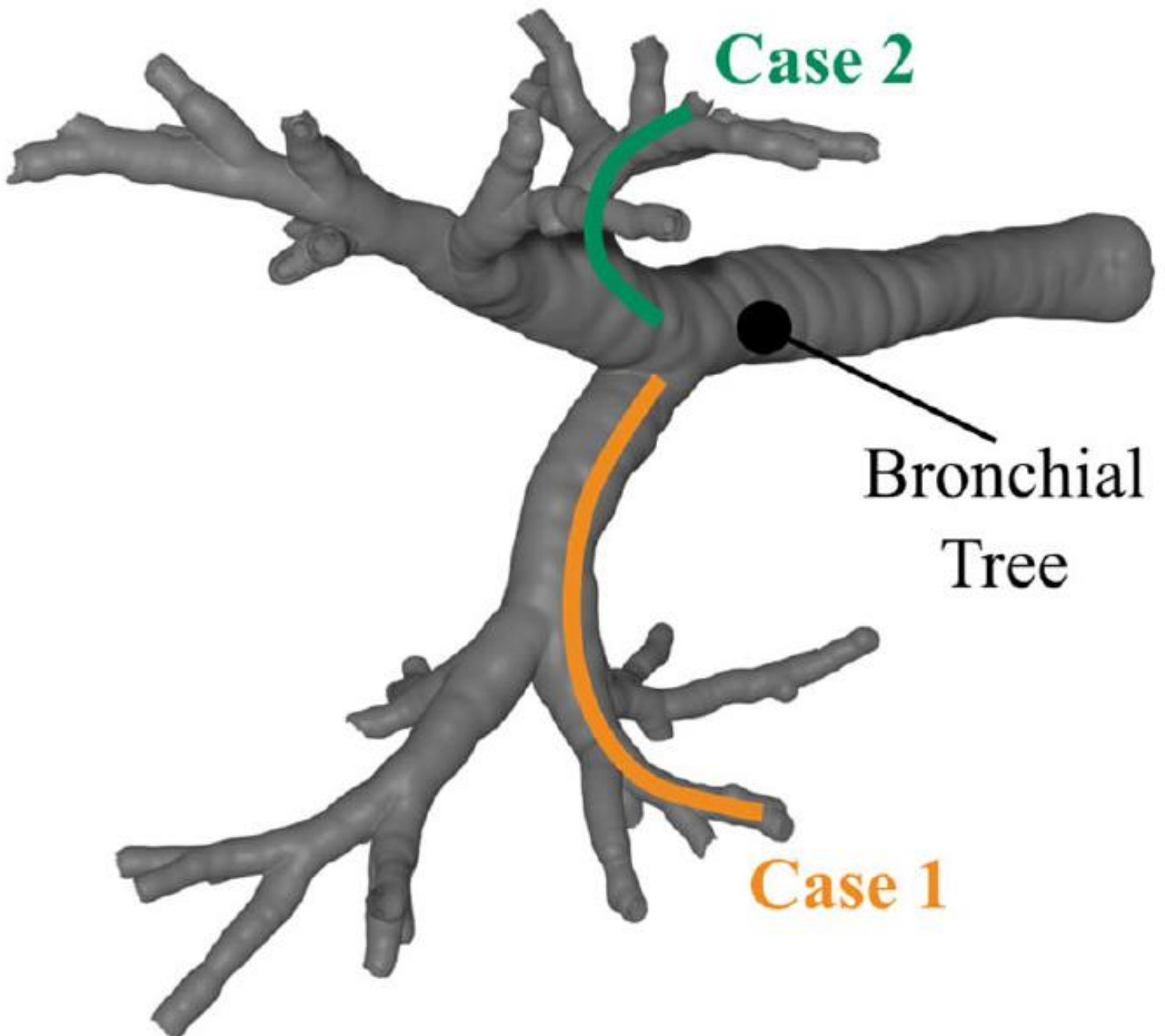
[We wrote this together at MICCAI 2022.](#)

His full interview is in [Computer Vision News of June 2022.](#)





*“Instead, we want to go through the mouth and trachea and navigate the branches of the bronchial tree!”*





Pietro Valdastri is Chair in Robotics and Autonomous Systems at the School of Electronic and Electrical Engineering, University of Leeds.

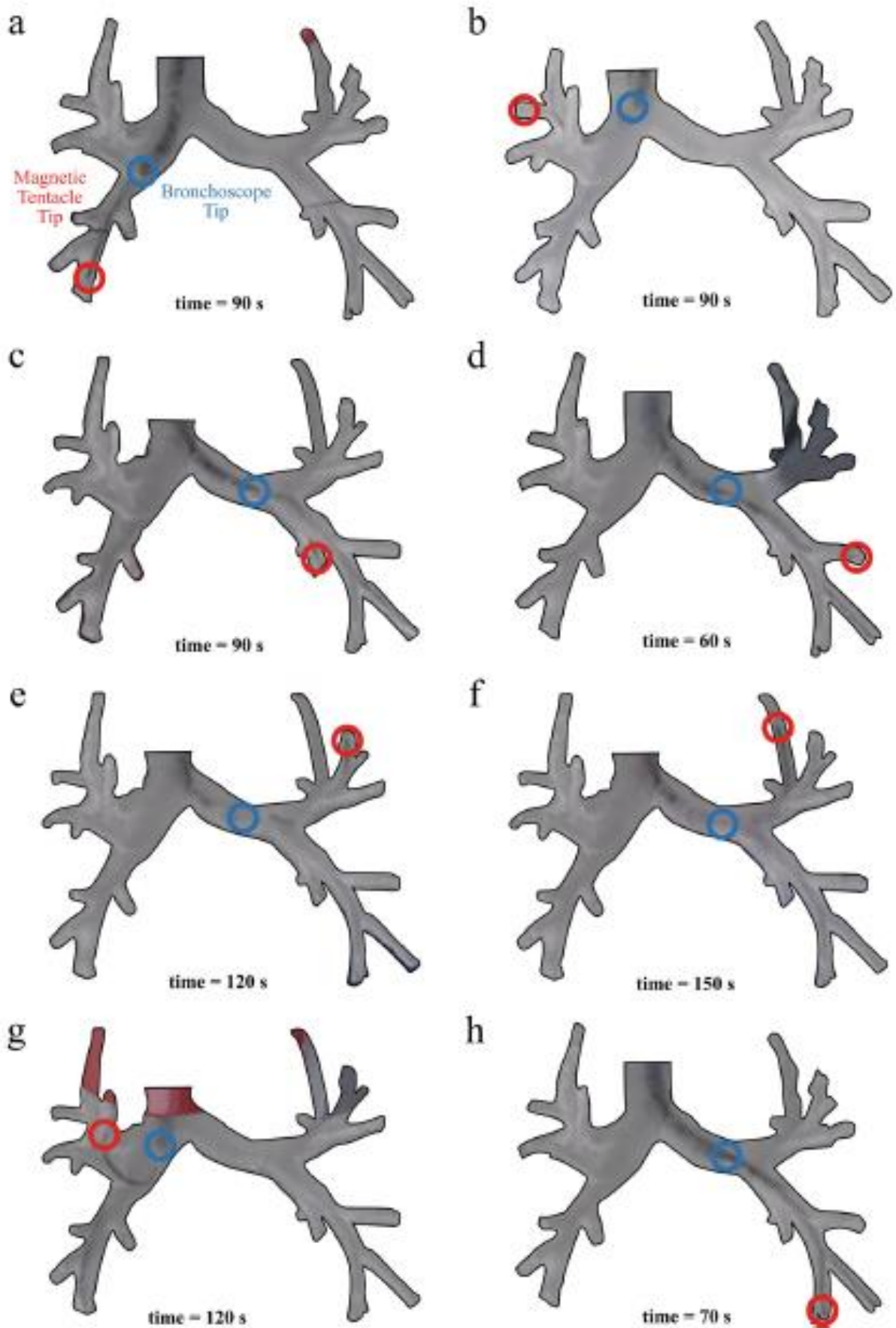
He is also the Director of **STORM** (Science and Technologies Of Robotics in Medicine) Lab, working on technologies to transform endoscopy and surgery. He speaks to us about its groundbreaking work.

**Magnetic manipulation** has emerged as an effective method for exploring deep inside the human anatomy in the most minimally invasive way possible. At **STORM Lab**, they work on enabling technologies that employ **external magnets in tandem with catheters containing embedded magnetic particles**, allowing unparalleled access to remote anatomical regions.

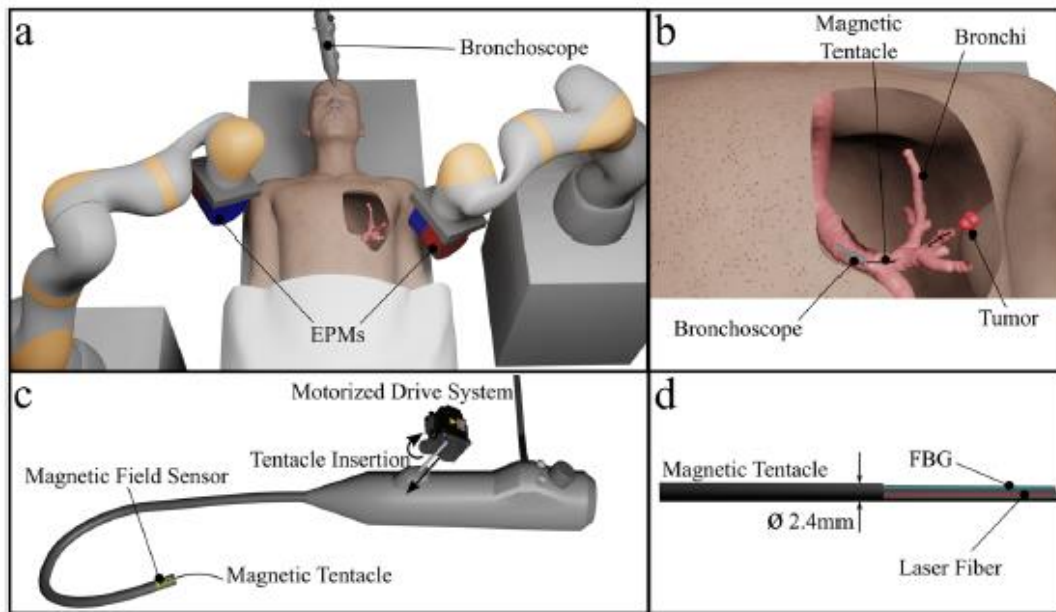
*“We’re using this technology to get as far as possible into the lungs,”* Pietro tells us. *“At the moment, it’s not possible to go very deep into the lungs from the throat. You can reach the peripheral areas by puncturing from outside, but this may puncture the lungs themselves, which can collapse. Instead, we want to go through the mouth and trachea and navigate the branches of the bronchial tree.”*

**STORM Lab’s magnetic tentacles**, measuring a mere 1.5mm in diameter, represent a significant advancement over existing tools, which are around 4mm wide and cannot get very deep. Its **magnetic catheters can facilitate navigation and the retrieval of tissue samples for diagnostic purposes**, as well as the delivery of targeted treatments, such as microwave therapy for tumors.

To enable their tiny size, **onboard sensors rather than traditional cameras enable intelligent navigation**, providing real-time feedback on the catheter’s position. Preoperative imaging helps plan a path, while an optical fiber senses the shape and localizes the catheter, which can be visualized in 3D.



**Final tentacle locations in phantom navigation experiments.** a–h Final navigation locations reached by the magnetic tentacles in eight primary targets of the sub-segmental bronchi. Locations of the bronchoscope tip (blue circle) and the magnetic tentacle tip (red circles) are indicated for each navigation, along with the associated completion time.

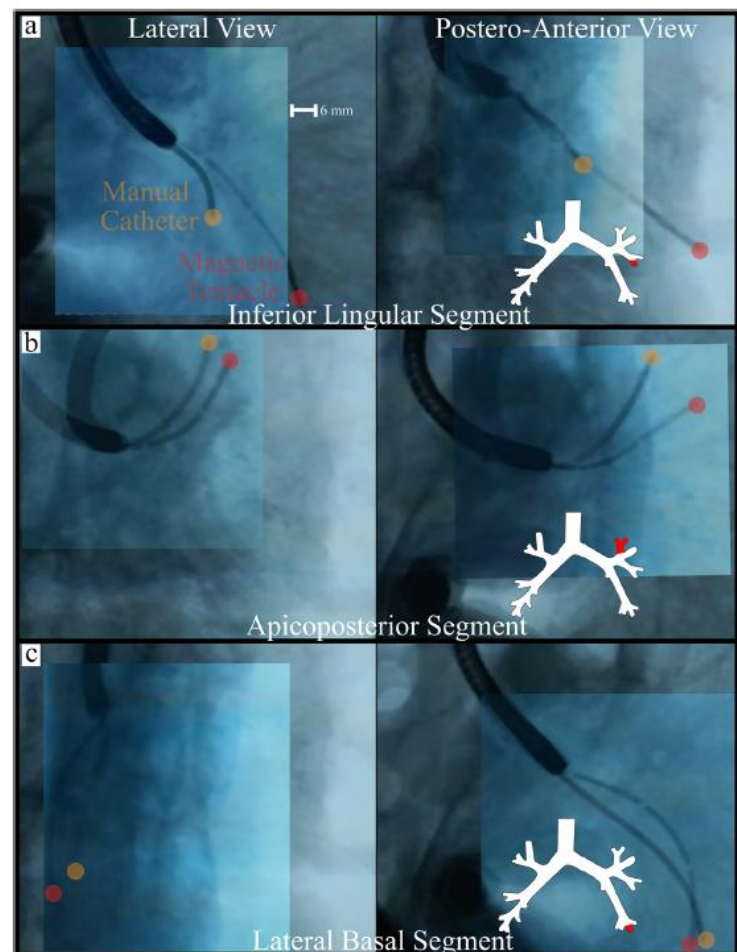


**Magnetic tentacles platform description.** **a** Overview of magnetic tentacle delivery bronchoscope and actuation system comprised of two robotic arms, each controlling the pose of an external permanent magnet (EPM). **b** Magnetic tentacle deployment and laser delivery to a targeted tumor. **c** Illustration of the tentacle delivery system and sensing. **d** Schematic of the magnetic tentacle showing the integrated shape sensing Fiber Bragg Grating (FBG) and laser fiber.

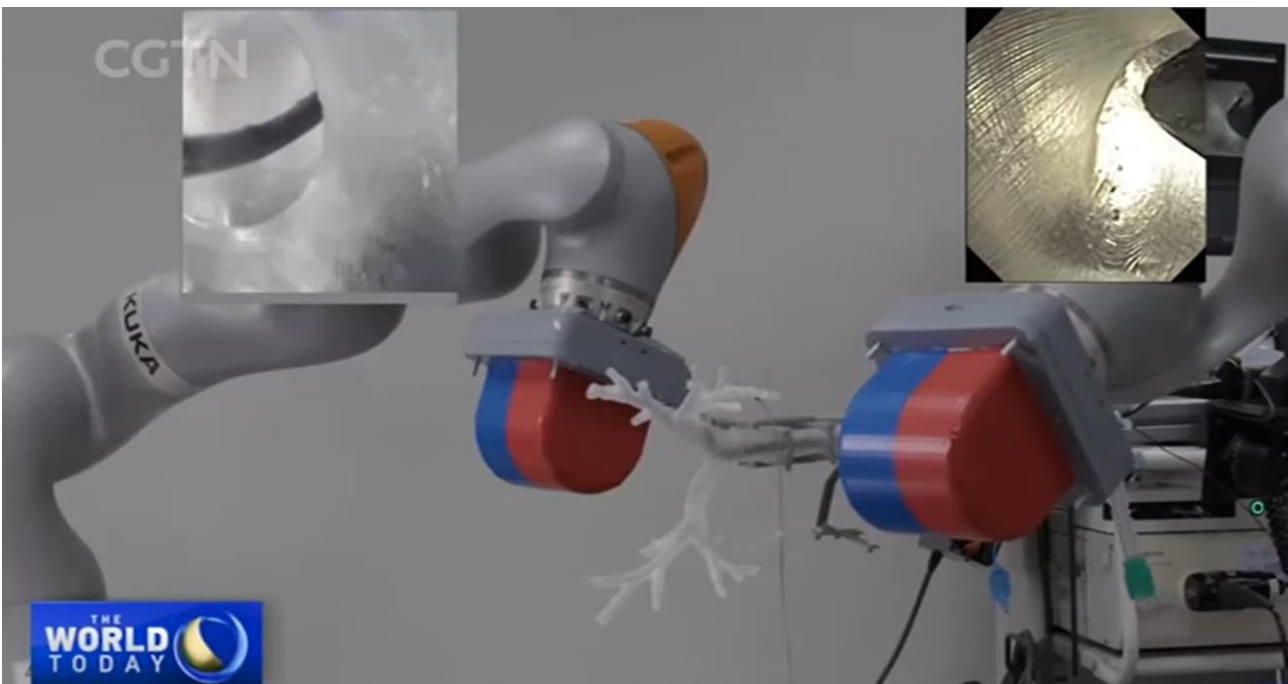
Standard practice uses fluoroscopy and a C-arm to visualize the catheters in real time. However, Pietro would like to see ultrasound used more often, as it poses no risk of ionizing radiation. *“That’s a challenge for the vision community,”* he identifies. *“Ultrasound is non-ionizing but difficult to understand from a cognitive perspective. I would love to see more progress in understanding ultrasound images.”*

#### Cadaveric experiment navigation results.

Fluoroscopic images showing the final navigation locations in three primary targets in the sub-segmental bronchi of the cadaveric specimen for the manual catheter (yellow filled circle) and magnetic tentacle (red filled circle) in the lateral view (left) and posterior-anterior view (right). Regions shown are **a** inferior lingular segment, **b** apicoposterior segment, and **c** lateral basal segment. Separate fluoroscopic images from independent navigation with the manual catheter and magnetic tentacle are presented as an overlay for comparison purposes.



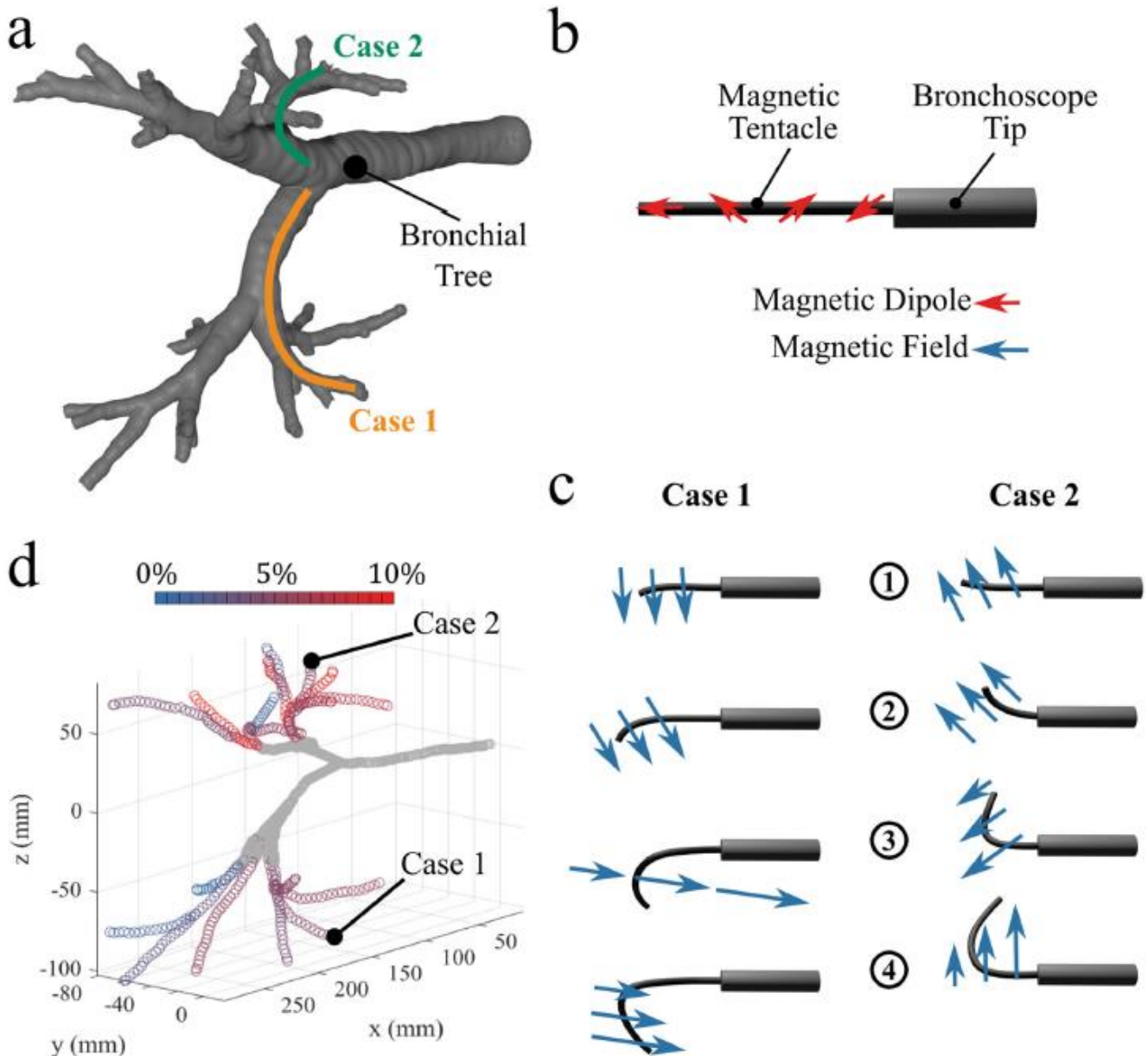
The team envisions their technology becoming a staple in surgical settings. Currently, there are at least three FDA-approved robotic bronchoscopes. The smallest is 3.5mm in diameter and has a camera, but a manual catheter is needed to extend further into the lungs. **STORM Lab hopes to integrate its robotic catheters with magnetic guidance into existing endoscopes**, enhancing their utility. It has already published a study demonstrating their use in human cadaveric lungs, and its work has received positive coverage, including a comprehensive feature by CGTN (watch video below).



*“We’re at a stage of preclinical evaluation in human anatomy,”* Pietro reveals. *“The next step is preclinical evaluation in living animals because you have breathing and a heartbeat. **When navigating the bronchial tree in the lungs, you need to compensate because the patient is breathing. You also need to compensate for motion due to the heartbeat.** If the animal trial goes well, the next step is human clinical trials.”*

Reflecting on previous projects involving **wireless robots**, he recalls potential hazards such as devices becoming lodged inside the body. However, with the current focus on **catheter-based systems**, the tool can be easily pulled out if something goes wrong, so it is an inherently safer design. Nevertheless, **working with magnetic fields can be challenging due to their unintuitive and nonlinear nature in space**. It helps to have a thorough understanding of how they function and **proficient robotic control skills to navigate and manipulate them effectively**.

*“It’s also about how we generate magnetic fields,”* he points out. *“There’s a line of research that uses **electromagnetic coils**. In coils, you can control the intensity of the field by controlling the current. We prefer to use **permanent magnets at the end of a robotic arm**. You can’t control the field with*

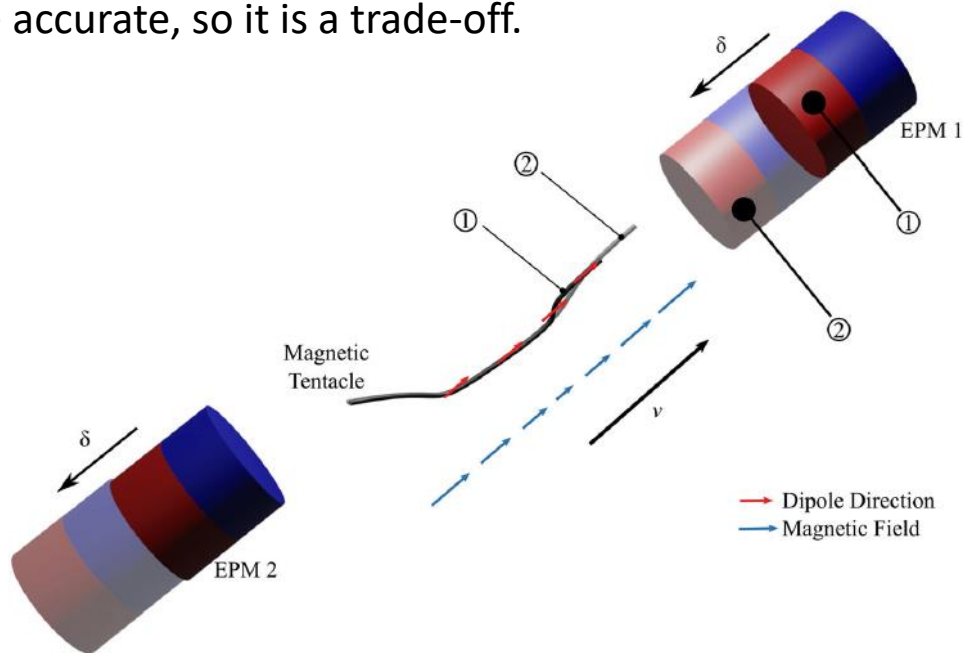


**Optimal design of the magnetic tentacles.** Analysis of patient-specific design over main branches of the bronchi. **a** Evaluation of general lumina in the left (Case 1) and right (Case 2) bronchi. **b** Example tentacle magnetization profile for optimal navigation in the principal sub-segmental branches. **c** Field and field gradient actuation (blue arrows) for the main cases (1 and 2) in the four insertion steps (1)–(4). **d** The lung geometry is colored according to the percentage mean spatial error at the optimization phase; color is scaled between 0 (blue) and 10% (red) error.

*current, but they're smaller. You can achieve the same force on a catheter with a small permanent magnet or a large coil. We like permanent magnets because they're compact and easy to integrate into an operating room, but they're also more challenging to control!"*

Pietro tells us he often monitors work on **magnetic field modeling** by **Bradley Nelson's group at ETH Zurich**, whose research focuses on **microrobotics and nanorobotics**. STORM Lab uses the dipole model to

predict the magnetic field and determine the position of the permanent magnet. However, it is an approximation with a lot of inaccuracies. The team would need a more complex and computationally less efficient model to be more accurate, so it is a trade-off.



**Magnetic tentacle actuation principles.** Description of magnetic steering based on field-magnetization alignment (1) combined with tip-dragging in a tangential direction to the anatomy's centerline ( $v$ ) realized via gradient pulling and relative translation ( $\delta$ ) of the external permanent magnets (EPMs) (2).

This thought leads us to a fundamental question: why use magnets in the first place? Pietro says the most viable alternative would be **small tendons integrated into the catheter**. However, these would have implications for its size. *"If you integrate tendons in the catheter, then you need a larger diameter and a stiffer device,"* he explains. *"With magnets, you just need to embed magnetic particles. Then, the magnetic force and torque for actuation and steering come from outside, so your catheter can be soft and thin."*

STORM Lab operates at the intersection of **robotics and medicine**, comprising a diverse 30-strong team of postdocs, PhD students, and lecturers. They are all robotics engineers with varied backgrounds, including electronic and mechanical engineering and computer science. *"It's also crucial for us to work with clinicians,"* he emphasizes. *"We have many collaborators here in Leeds and in the US, but it's key that we work with doctors on a daily basis."*

Its **magnetic colonoscopy robot**, licensed to a University of Leeds spin-out company, **Atlas Endoscopy**, is pioneering technology poised to transform this common medical procedure. Having already completed human trials, Pietro is optimistic: *"I would say keep an eye on Atlas Endoscopy,"* he teases.

**STORM Lab is looking for new talent. If you think you have what it takes, you may be joining them on the next leg of their journey of discovery!**

## Enhancing Breast Cancer Screening with AI

Breast cancer is the second most common cancer in women (skin cancer is the first) and a significant challenge to healthcare systems globally. The risk of developing breast cancer increases with age and is influenced by factors such as someone's genetic makeup and family history.

Like other serious health conditions and diseases, **early detection of breast cancer** is essential to improving survival rates, and routine breast screening is a critical part of this. Women over 40 are recommended to have a mammography examination once a year. In addition, MRI and ultrasound exams may be offered to high-risk people with a family history of breast cancer or those who are carriers of the BRCA gene.

However, traditional screening modalities can expose patients to unintended risks, like harmful radiation from mammography, and create unsustainable workloads for radiologists, raising the possibility that they will be tired and make mistakes. Incorporating **artificial intelligence (AI)** can help solve these challenges and optimize breast cancer screening efficacy, accuracy, and efficiency, ultimately driving progress in the fight against this disease.

Mammography involves the breast being pressed between two surfaces and a 2D X-ray image or a 3D tomography scan via tomosynthesis

being captured. A radiologist then examines these scans to identify suspicious lesions or areas requiring further investigation. In performing this scan, there is a **delicate balance between radiation dosage and image quality**. With the help of AI, radiation exposure can be reduced by training neural networks to reconstruct high-quality images from lower-quality ones. **Classic computer vision techniques and more advanced deep learning methods** can help detect and segment suspicious landmarks, relieving the pressure on radiologists and increasing precision.

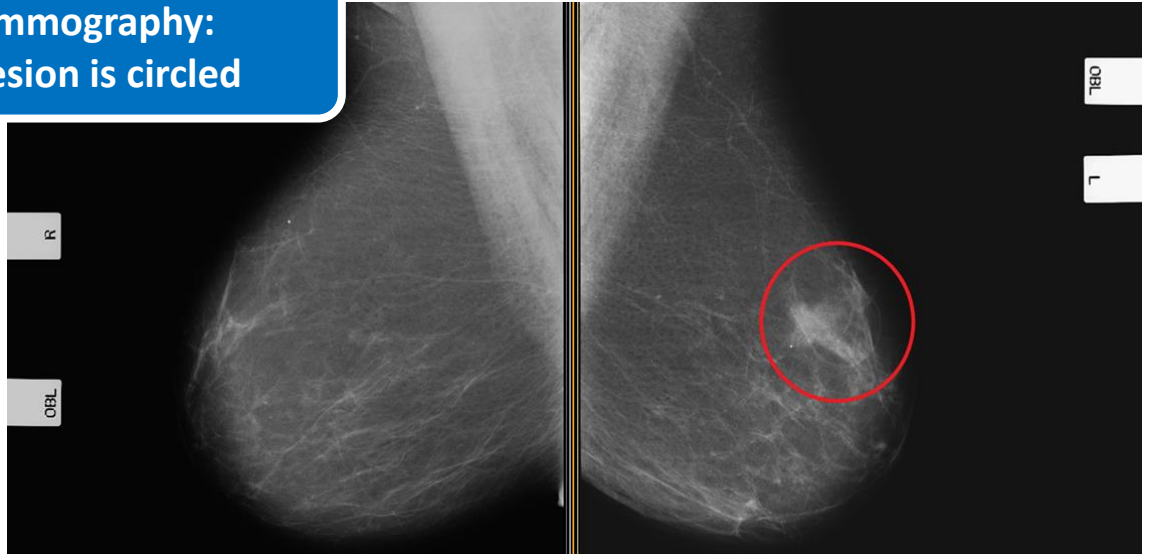
Breast MRI, though highly effective, has long examination times due to the acquisition of multiple sequences and the person's breathing, which causes movement that must be taken into account. Deep learning techniques can register each sequence to a predefined baseline,



**Breast MRI - MIP with subtraction showing enhanced procedure in left breast**



**Breast Mammography:  
suspicious lesion is circled**



such as **Diffusion Weighted Imaging (DWI) to T2**, eliminating relative motion between and during acquisition. Resolution can be further improved by **training deep neural networks to reconstruct MRI images from suboptimal acquisitions**. These methods streamline the imaging process by reducing the need for repeated acquisitions. Also, as in mammography, **computer vision systems can assist radiologists by analyzing breast MRI images, detecting and segmenting suspicious lesions, and providing quantitative data for interpretation and diagnosis**.

Ultrasound, particularly **automated breast ultrasound (ABUS)**, is a promising approach for examining women with dense breast tissue or abnormalities detected by other imaging modalities. In ABUS, a large transducer on the breast autonomously captures multiple images encompassing the entire breast. AI helps detect abnormalities automatically and, with robust data support, can even enable malignancy

classification, significantly reducing missed detection rates. Also, ultrasound frequently serves as a tool for biopsy guidance, **where AI techniques can improve the accuracy of needle guidance and tracking and reduce the duration of the procedure**.

Finally, standardized reporting guidelines, such as the **Breast Imaging Reporting and Data System (BI-RADS)**, ensure consistency and accuracy in interpreting screening images across mammography, MRI, and ultrasound. Parameters such as lesion size, volume, shape, homogeneity, restriction, and other characteristics can be **automatically extracted from the imaging data and presented to radiologists** for informed decision-making, saving time and bolstering confidence.

**RSIP Vision is committed to assisting in developing AI-driven solutions that will reduce screening duration and costs while raising early detection and survival rates for breast cancer for women worldwide.**

**Lei Li is a lecturer at the University of Southampton.**

**Lei, can you tell us what your work is about?**

My work is about AI for healthcare, specifically focusing on cardiac digital twins, like creating a virtual heart for the patient using the patient's data.

**What made you decide to work on this?**

It's a very exciting project for me! You can create a virtual heart for the patient, and then you can do lots of things on the virtual heart to help the patient with treatment, diagnosis – many things.

**Have you seen any results from your research so far?**

During my PhD, I was working on atrial fibrillation, mainly focused on developing some AI tools to detect the scarring area, which is quite a challenge. For a long time, the doctor needed to manually segment because it's a very tiny objective within the heart. What I have done is that I developed a very accurate tool that can automatically detect this area. The doctor can use this information to select patients and guide the operation, and also predict the recurrence rate of the operation, so it's very useful.

**Read 100 FASCINATING interviews with Women in Computer Vision**

***“It's a very exciting project for me!”***





**Does it have benefits for the patient, too?**

Yes.

**What are those benefits?**

For the atrial fibrillation patient, for example, they usually need to do a repeated operation procedure. It means they need to go to the hospital many times to get a fully successful operation. What I can do is predict the area of the scar, which can guide the doctor to find the optimal operation target, which can reduce the times the patient needs to go to the hospital.

**This sounds like a big project to improve healthcare.**

Yeah. [Lei laughs]

**Do you feel a lot of responsibility?**

Yes. Previously, I mainly focused on the image data. It's not enough because we know the heart is beating and we also need to consider the electrophysiological property of the heart. Currently, what I'm doing now is to create a virtual heart, which is not only anatomy but also some electrophysiological information to mimic the heartbeat of a real patient. It's also patient-specific. With this, we can do a virtual surgery on the virtual heart to find the optimal operation target for the patient. It's very promising for the future.



**You are not originally from Southampton, are you?**

No, I just moved here last month.

**You are Chinese-born and Chinese-educated. How did you end up in the UK?**

Good question. I'm educated in China. I grew up in China. As you may know, for a researcher, it's always good to go to a different place, and the UK is my favorite one, so I came here because I watched a lot of UK dramas like Harry Potter and things like that! *[she laughs]* I always wanted to come to the UK, and I got an opportunity during my PhD to be a visiting student at King's College London, supervised by Professor [Julia Schnabel](#). You've also interviewed her.

**You came for the drama, but have you met any actors since you have been in the UK?**

*[Lei laughs]* No, I didn't have that chance. I know that a lot of my friends met some actors in the subway, but I didn't get that chance. Most actors live in London, but during the time I stayed in London, Covid happened, so I stayed the whole year at home! *[she laughs]*

**You should have worked at King's, not at Southampton. In Southampton, there is no subway, right?**

Yeah! *[she laughs]*

**Maybe you can spend a weekend in London and find the actors! Who is the actor that you most wish to meet?**

Many choices. I can't remember all their names, but Laura Carmichael in Downton Abbey and the actor in Harry Potter and Keira Knightley in Pride and Prejudice - I really like her!

**What will you say if you meet them?**

I might say, I really like you, and I watched a lot of movies and drama from you!

**When you arrived in the UK, did you find a good environment for your research?**

Yeah, it's a very good question. I think the research atmosphere in the UK is quite different from China. In China, it's too many people, so it's quite competitive. People are always trying to publish a lot of papers, and some of them are maybe not useful,



but people are very hardworking and that's very different from the UK. After I arrived in the UK, I realized people don't need to work at night and neither on the weekends. That's quite amazing for me because in China, where I did my PhD, we always needed to work at night, and also during the weekends.

**Wow. That must be very challenging. How did you manage to spend your young years working on research instead of going out and having fun with your friends?**

I didn't feel it's not normal because many people surrounding me also did the same thing, so I thought that's the thing we should do. I didn't think much about that. After moving to the UK, I realized people don't need to work so much time, and a lot of their life belongs to themselves. Then I followed their style and realized I suddenly have a lot of time, which I haven't had before! [*Lei laughs*] That's quite new for me. At the beginning, I didn't even know how to spend this spare time because I didn't have that before. [*she laughs*]

**So now go and have fun!**

But one thing very interesting I realized is that even though I work less time compared to China, I don't think my research has been delayed, which means I'm more focused and more efficient, even though I don't work for such a long time.

**Is there anything that you miss from China in your European work? Food and family do not count!**

[*she laughs*] Yeah, I wanted to say food and family. Otherwise, I think I miss my previous research group, like my supervisor, Xiahai Zhuang. It's a bit shy to say, but I regard him almost like my father. He's super nice.

**What is the most important thing you have learned from him?**

The most important thing I learned from my supervisor is critical thinking. Previously, I never thought I will be a scientist, or I will be a PhD student. I was just a very typical Chinese student just learning the thing I learned from the teacher. The teacher tells me, I will believe it. I will never think whether this is right. After I did my PhD with my supervisor, he always taught me to think about the things I learned from others and read from papers; I cannot 100% trust it. Also, he always guided me, before I do



anything, I need to think about why I do this. I know many people have this kind of training, but I didn't have that before, so it's quite fresh for me. I started to become curious about the world. I realize there are so many questions around me, and I always keep asking and searching in Google, but I never did it before. I just trusted the things that happened in my life. I never thought about it. But after, with my supervisor, I started thinking about my life. It's quite different.

### **'Why?' is an important question.**

I just think once you start to think about the world and think about life, you'll realize the world is so interesting, and you can start to engage with the world. But previously, I didn't have the chance to think about that. Like my mind has been closed.

### **Where are you going, Lei?**

*[she laughs]* It's a big question! I just started a new position at Southampton. It's quite new. I still don't have students, but once I have students, I will try my best to mimic my PhD supervisor to guide the students. Also, research. My supervisor is a very good researcher. My current research is kind of similar but not exactly similar to my PhD supervisor's research. His research is more about algorithms to develop interpretable AI models, but I'm more interested in clinical applications. In the future, I really hope that I can develop tools,

develop the virtual heart, which can be applied in the hospital, which has still not happened in the UK. I know there are some cases in America or maybe in Germany, but I don't know any case of the virtual heart applied in the clinical application. I do hope that one day I can achieve that goal.



### **Is this the right moment to say that you are looking for PhD students?**

Yes, I'm looking for PhD students.

### **What kind of people do you want?**

Good question! After I sent out our PhD opportunity on LinkedIn, I received a lot of applicants' emails, and the one thing I realized is most of them are male. I do hope that I can receive some female researchers because I want to support that. Another thing is I'm looking for some students who like the research direction of the virtual heart model and are interested in AI.

**Read 100 FASCINATING interviews with Women in Computer Vision!**



Victor Campello (right in the picture) has recently defended his PhD at Universitat de Barcelona, under the supervision of Karim Lekadir and Santi Seguí.

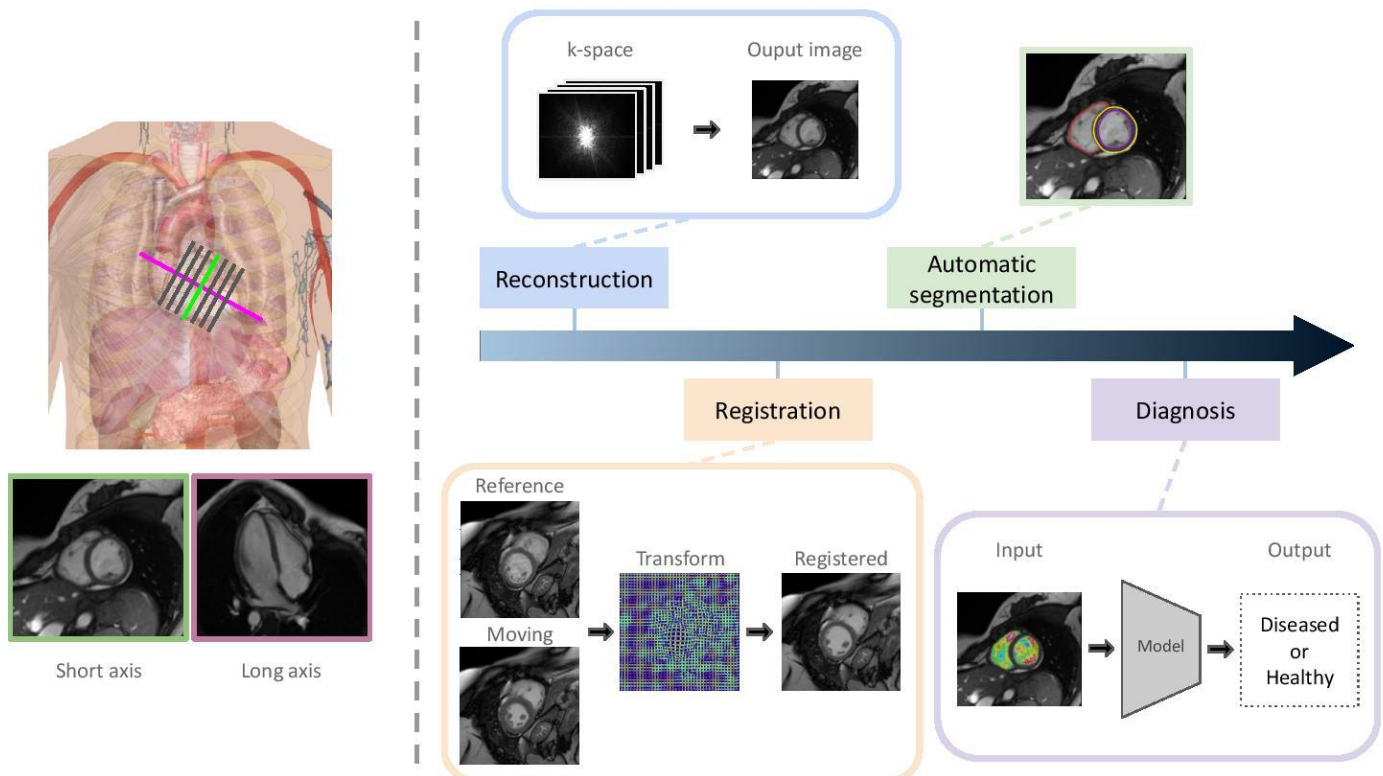
His thesis focused on model generalizability in a multi-centre setting for cardiac MRI.

He is now a Postdoctoral Researcher in an ERC-funded project that aims to develop and validate affordable and inclusive AI methods in low-resource settings in rural Africa.

**Congrats Doctor Victor!**

The field of medicine has experienced the appearance of AI models that can solve complex tasks such as automatic image segmentation or diagnosis (see examples in Figure 1). The application of such models in the

clinical practice, if adequately implemented, may result for example in a reduction of time-consuming tasks, an improvement in the diagnostic accuracy or a better characterization of a disease. However, important challenges need



Visualization of the heart orientation in the chest with the two main MRI views used in the literature and a depiction of four important steps in cardiovascular image analysis where AI has been used.



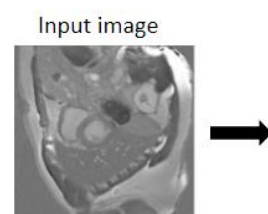
to be overcome before these methods are used in the daily practice. Victor focuses on one of these important challenges in his thesis: the **generalization** of models to unseen domains independently of other factors, such as the scanner manufacturer, the scanning protocol, the sample size or the image quality.

In his thesis, Victor established a collaboration with clinical researchers from six different centres from Spain, Germany and Canada to assemble a large multi-center dataset: [the M&Ms Dataset](#). These data were used in **the M&Ms Challenge**, organized at MICCAI 2020 in Perú (virtual edition), with the aim of comparing and analysing different techniques proposed by the participating teams to bridge the domain gap. The results highlighted the importance of well-established frameworks and extensive data augmentation.

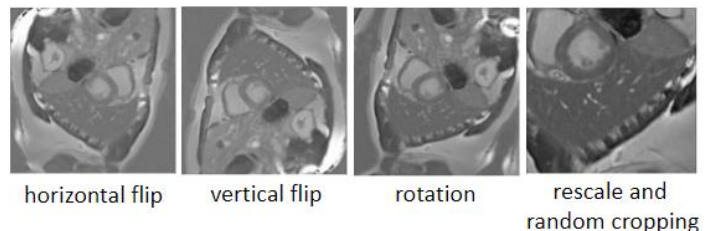
The second contribution focused on **model generalization on contrast-**

**enhanced imaging**, where the variability in image appearance across centers is larger due to the injection of a contrast agent and the disparities in the time elapsed between the contrast injection and the scan. Victor and his colleagues showed that extensive data augmentation (shown in Figure 2) is very important for generalization and that model fine-tuning can reach or even surpass the performance of multi-centre models.

In the final contribution of the thesis, Victor and his colleagues investigated how to **harmonize images and features** from multiple centres for an improved diagnostic accuracy on unseen domains. They showed that histogram matching-based harmonisation results in image features (radiomics) that are more generalizable across centres.



*Spatial-based data augmentation techniques*



*Intensity-based data augmentation techniques*

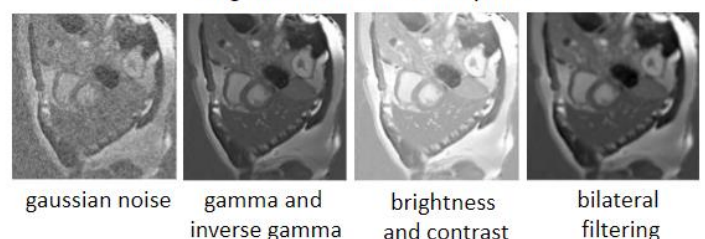
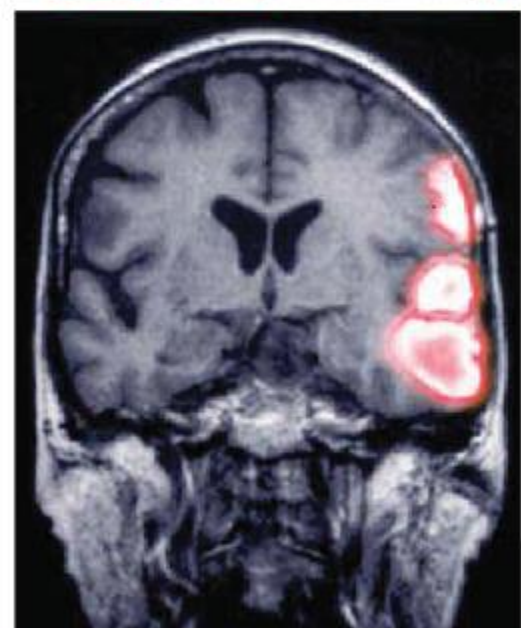
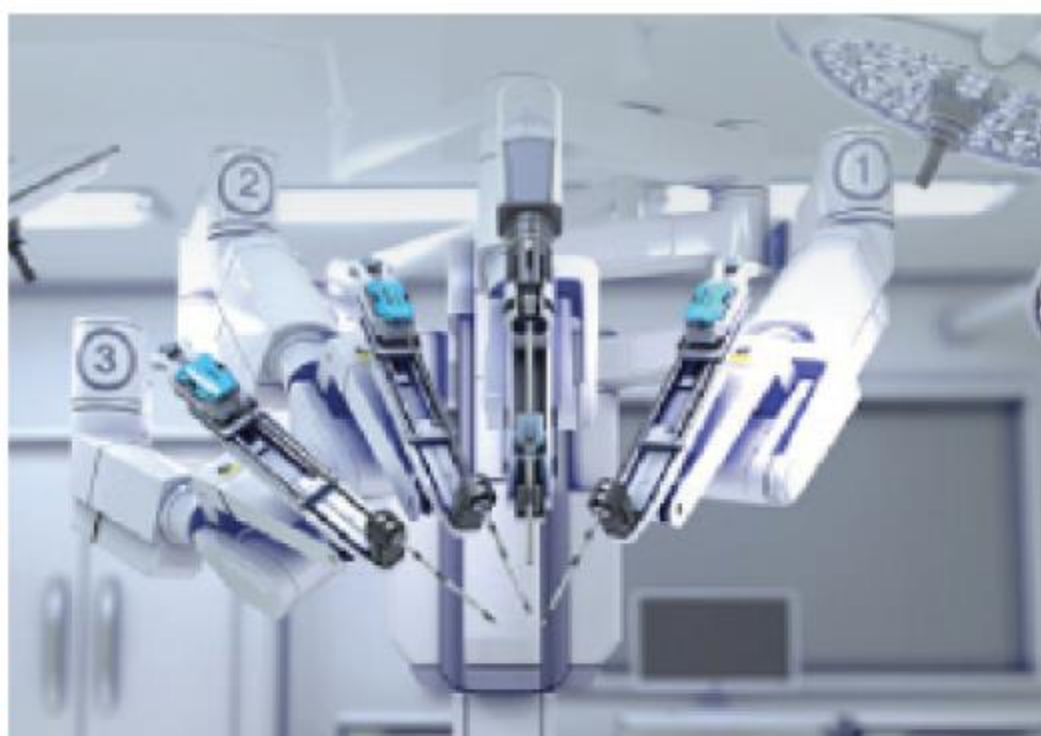


Illustration of the effect of spatial and intensity-based data augmentation applied to a contrast-enhanced cardiac MRI scan.



## IMPROVE YOUR VISION WITH Computer Vision News

**SUBSCRIBE**

to the magazine of the  
algorithm community  
and get also the  
new supplement  
Medical Imaging News!

