

October 2023

Computer Vision News & Medical Imaging News

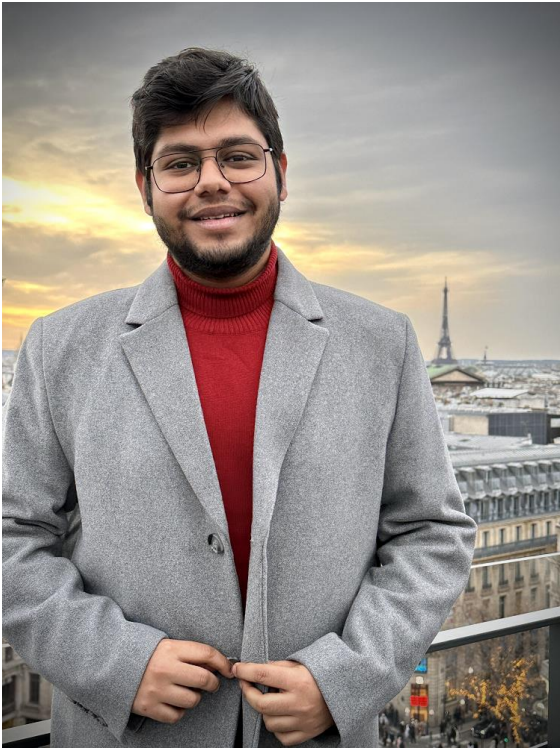
The Magazine of the Algorithm Community



A publication by



SGAligner: 3D Scene Alignment with Scene Graphs



Sayan Deb Sarkar is a Computer Science master's student at ETH Zurich majoring in Visual Interactive Computing.

Currently, he is interning with Qualcomm XR Research in Amsterdam.

In this paper, Sayan presents a novel method for aligning pairs of 3D scene graphs robust to in-the-wild scenarios. He speaks to us ahead of his poster this afternoon.

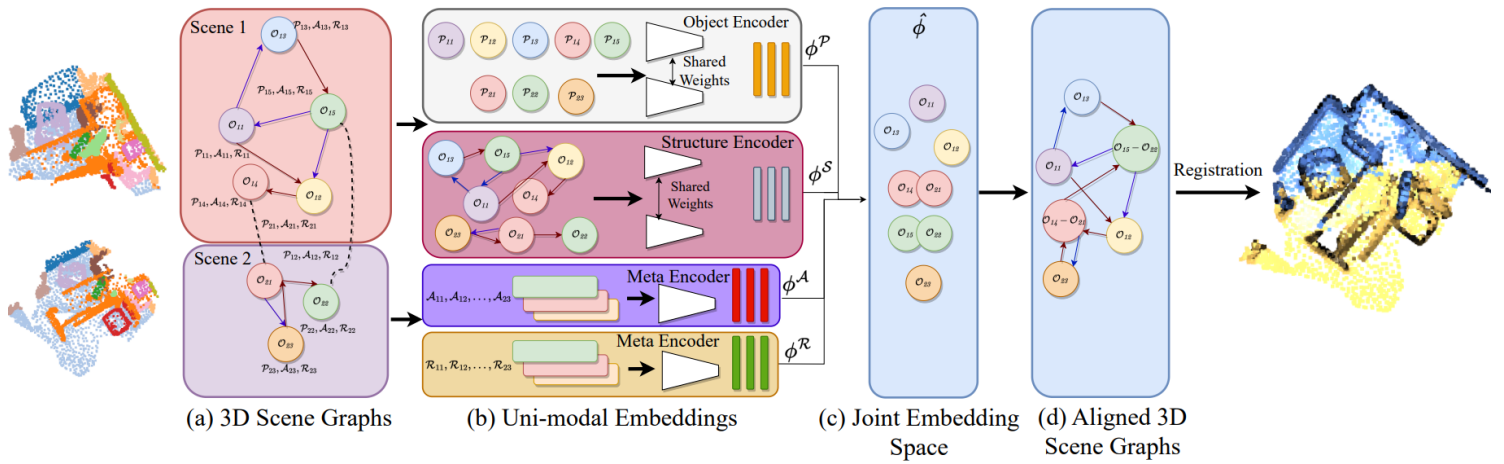
Generating 3D maps of environments is a fundamental task in computer vision. These maps must be actionable, containing crucial information about objects and instances and their positions and relationships to other elements.

Recently, the emergence of 3D scene graphs has sparked considerable interest in the field of **scene representation**. These graphs are easily scalable, updatable, and shareable while maintaining a lightweight, privacy-aware profile. With their increased use in solving downstream tasks, such as navigation, completion, and room rearrangement, this paper explores the potential of **leveraging and recycling 3D scene graphs for creating comprehensive 3D maps of environments**, a pivotal step in

robot-agent operation.

"Building 3D scene graphs has recently emerged as a topic in scene representation, which is used in several embodied AI applications to represent the world in a structured and rich manner," Sayan tells us. *"SGAligner focuses on the fundamental problem of **aligning pairs of 3D scene graphs whose overlap can range from zero to partial**. We address this problem using multimodal learning and leverage the output for multiple downstream tasks of 3D point cloud registration and 3D scene reconstruction by developing a holistic and intuitive understanding of the scene aided with semantic reasoning."*

Sayan demonstrates that **aligning 3D scenes directly on the scene graph level** enhances accuracy and

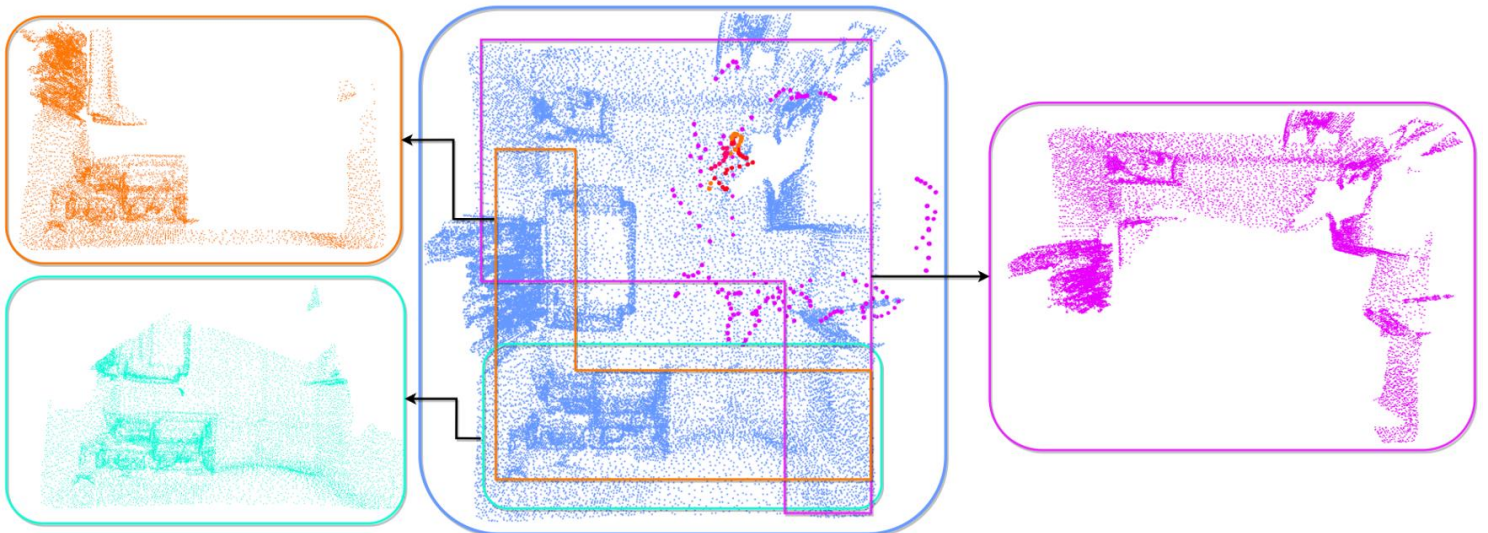


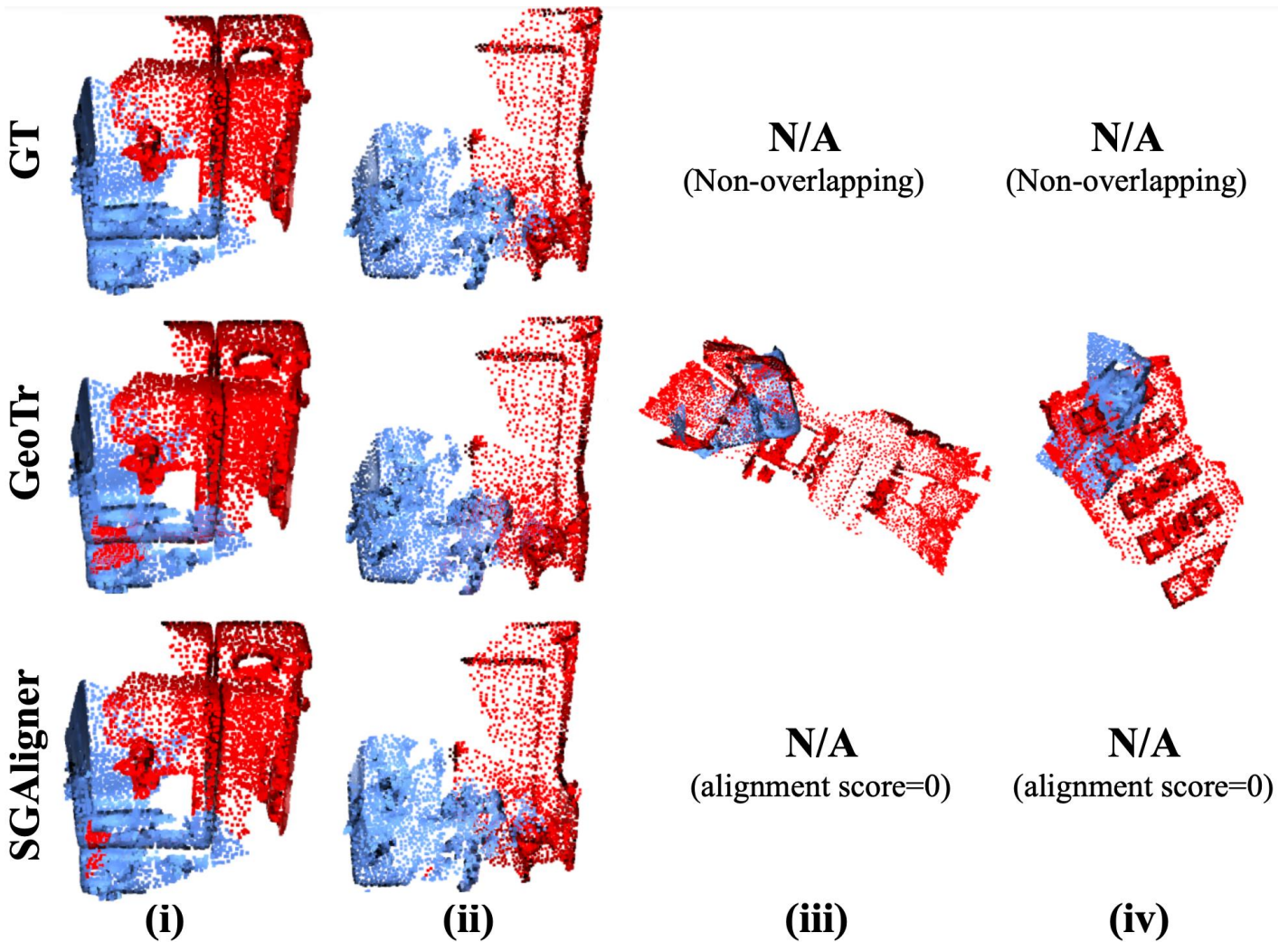
speed in these tasks and is robust to in-the-wild scenarios and scenes with unknown overlap. This finding opens up exciting possibilities to unlock potential in fields like **graphic design, architectural modeling, XR/VR experiences, and even the construction industry.** The ultimate goal is to move toward a gradient reality, where spaces are designed with user needs in mind, fostering immersive connectivity, communication, and interaction on a global scale. This work is the first step toward that goal **using semantic reasoning and aligning 3D scenes using a semantic meaning.**

The journey to develop SGAligner

has not been without its challenges. From a technical standpoint, Sayan tells us formulating the project and navigating the complexities of generating and aligning scene graphs was difficult.

“We wanted to be inspired by the language domain and then applied a similar setting in our computer vision problems,” he explains. *“That was one challenge figuring it out. The second part was understanding and visualizing which sorts of potential real-life applications we could plug into, and understanding the practicality and how to make this whole module very lightweight so that it is privacy-aware and can*





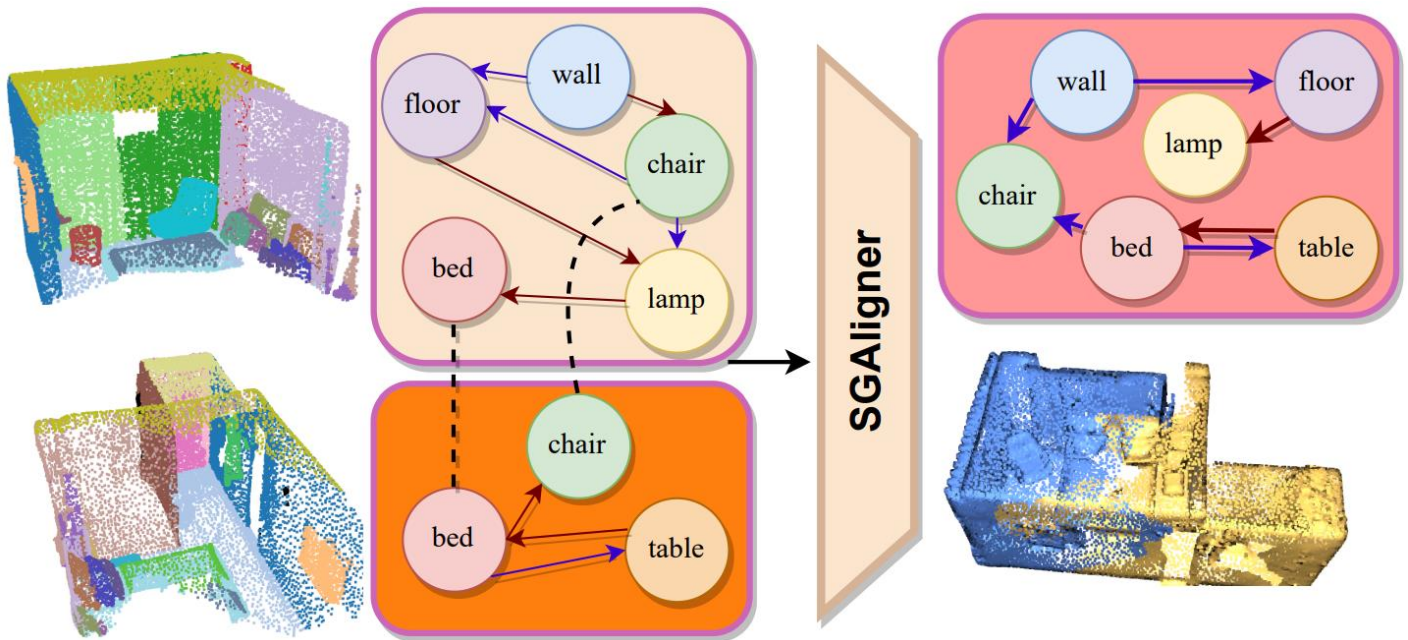
be easily shared among people.”

To his knowledge, **this is the first work to address the problem of aligning pairs of 3D scene graphs.** Its approach sets SGAligner apart – rather than relying on metric data, it operates exclusively on the graph level. This unique perspective confers robustness against various challenges, including noise, in-the-wild scenarios, and overlap. The implications of this approach are far-reaching, opening doors to applications in **mixed and augmented reality and even SLAM.**

Sayan started his master’s degree last September and took up this project in the first semester. One key personal hurdle was the balancing act of pursuing all this

while moving to Zurich and adapting to new surroundings. However, any obstacles were surmountable with the support of dedicated supervisors like **Ondrej Miksik**, [Marc Pollefeys](#), **Dániel Baráth**, and [Iro Armeni](#).

“I originally had reached out to Iro for the project when I was moving to ETH to start my master’s,” he recalls. “She has been very helpful in understanding where I need support and where I can be independent. Daniel is very experienced with point cloud registration and all the technical parts. He was the best person in the community to help me figure out the downstream applications. Marc is one of the grandfathers of 3D computer vision.



*He's been very helpful in driving the project. Ondrej always had a high-level understanding of the project, which helped me because, as a first author, it's easy to get lost in the technicalities when you work on a project. The last year of working with them helped me get my internship at **Qualcomm**, and I'm very grateful for all their help and advice."*

The impact of SGAligner is already being felt across the computer vision community. Making the code public and releasing the paper on **arXiv** has sparked interest among fellow researchers exploring various avenues for further development. Potential directions include **cross-modal alignments**, such as aligning point clouds with CAD models or other modalities, **applications in scene retrieval** from extensive databases, and **multiple downstream applications in AR scenarios**. The benefit of **SGAligner's lightweight and privacy-aware scene graphs** to the community cannot be overstated.

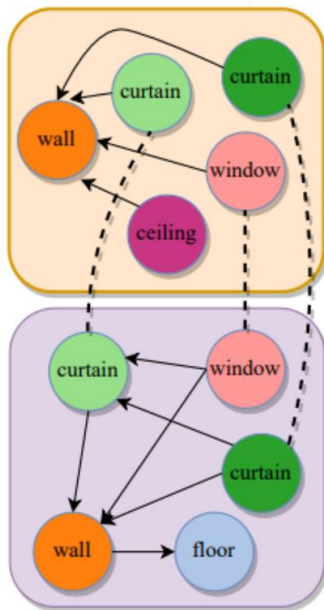
As lead author, this is the first time Sayan has had a paper accepted at a top conference – a fantastic achievement. Does he have any wisdom for those whose papers were not accepted this year?

"The computer vision community has grown manifold in the last few years," he responds. "Please do not be disheartened that your paper wasn't accepted because we have all been there at some point. Always have a high-level understanding of where your work could play into both industry and academia. It's not only about solving a new problem but also having real-life applications of the problem. SGAligner was a new problem, but we could also find multiple real-life applications where we could make a difference. Also, don't forget to do ablations and understand which other works might be relevant for comparison because it's always good from a reviewer's perspective to understand how your work performs differently from others."

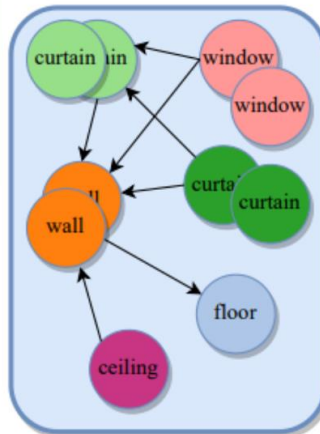
3D Point Clouds



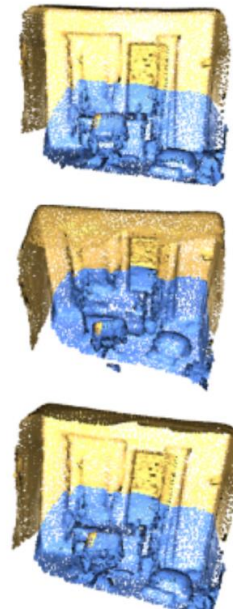
3D Scene Graphs



Aligned Scene Graphs



Point Cloud Registration



GT

GeoTr

SGAligner

Sayan has a research internship in **Marco Manfredi's** team at **Qualcomm XR**, working on improving the performance of lifelong SLAM systems and understanding how multimodality could plug into SLAM.

"I'm still a master's student, so I'm very happy getting exposed to this

sort of research," he smiles. *"It's a journey that started before ETH with Vincent Lepetit, and then with Iro, Marc, and everybody at ETH, it's continuing!"*

To learn more about Sayan's work, visit his poster this Friday at ICCV - Poster session of 14:30-16:30.

Computer Vision News

Editor:
Ralph Anzarouth

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



Publisher:
RSIP Vision

[Contact us](#)

[Give us feedback](#)

[Free subscription](#)

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

[Read previous magazines](#)

Copyright: **RSIP Vision**

All rights reserved

Unauthorized reproduction
is strictly forbidden.



ICCV23

PARIS

International Conference
on Computer Vision

October 2 - 6, 2023

**Feel at ICCV,
As if you were at ICCV!
Subscribe here for free**



SUBSCRIBE

1

Augmented Architecture for Vision and Language

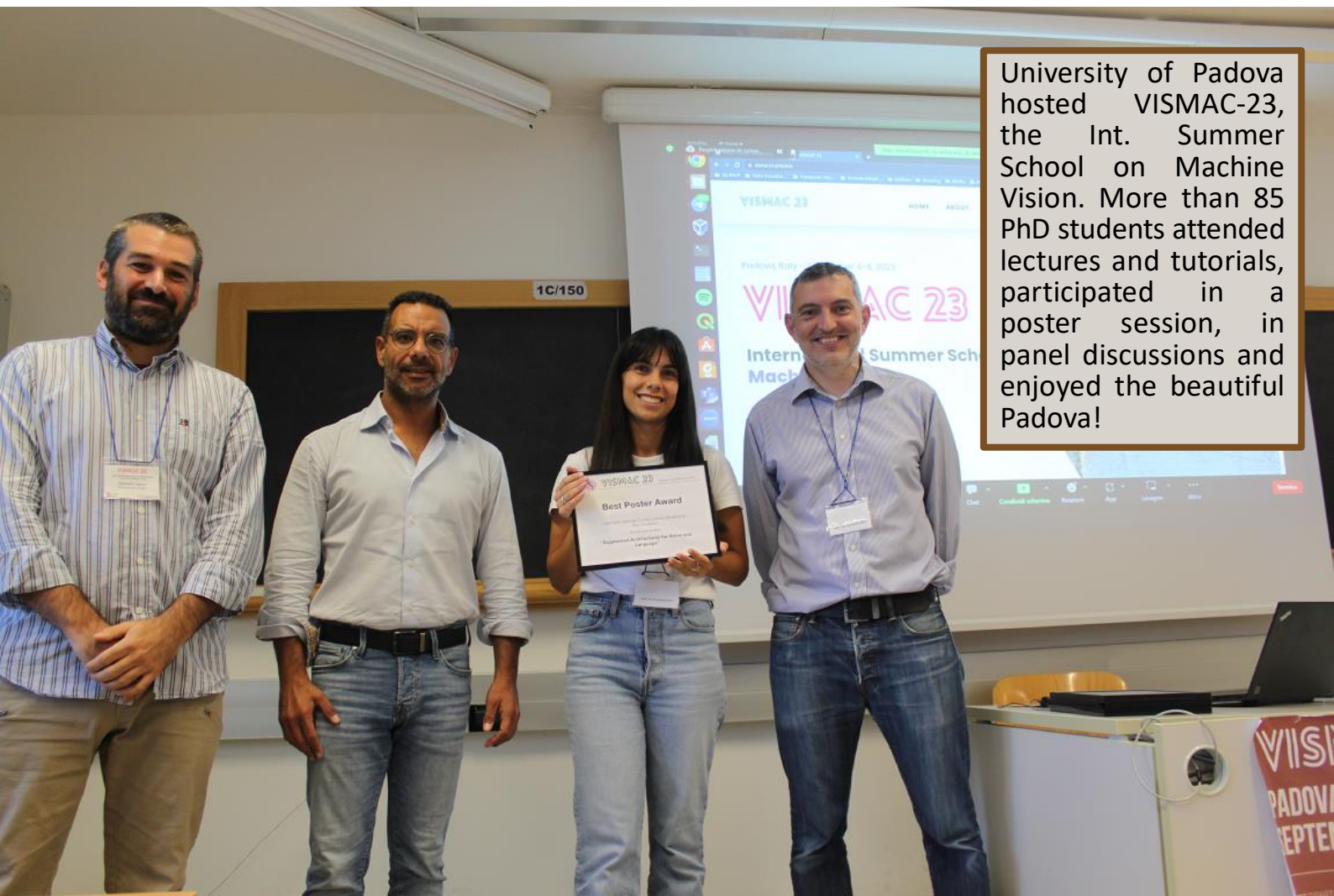
Hello, I'm Sara Sarto! I am a PhD Student at the University of Modena and Reggio Emilia.

Recently, I had the honor of winning the Best Poster Award at the International Summer School on Machine Learning (VISMAL) in Padova (Italy) thanks to my most recent research activities on augmented architectures for vision and language tasks.

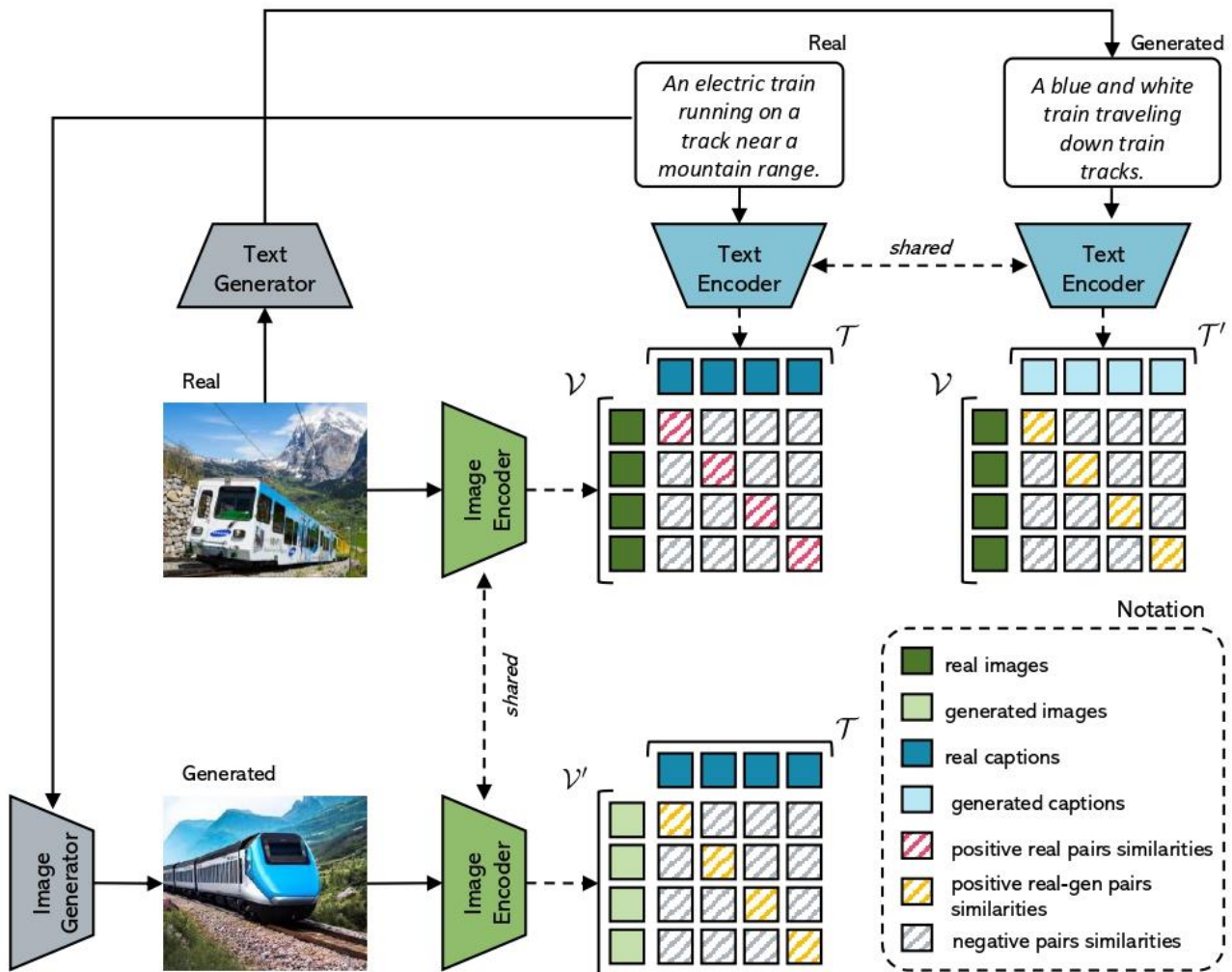
by Sara Sarto

During my first year of PhD I mainly focused on **image captioning architectures**. This interesting domain requires the development of an algorithm to describe visual contents with natural language sentences and, in the past few years, this field has garnered significant attention within the research community.

In this article, we'll explain how some kind of augmentation in this field can be a powerful source of information. During my presentation at **VISMAL2023**, I shared insights from my two latest research projects: one related to the evaluation of captioners (accepted at **CVPR2023**), the other to the generation of captions (accepted at [ICCV2023](#)).



University of Padova hosted VISMAL-23, the Int. Summer School on Machine Vision. More than 85 PhD students attended lectures and tutorials, participated in a poster session, in panel discussions and enjoyed the beautiful Padova!

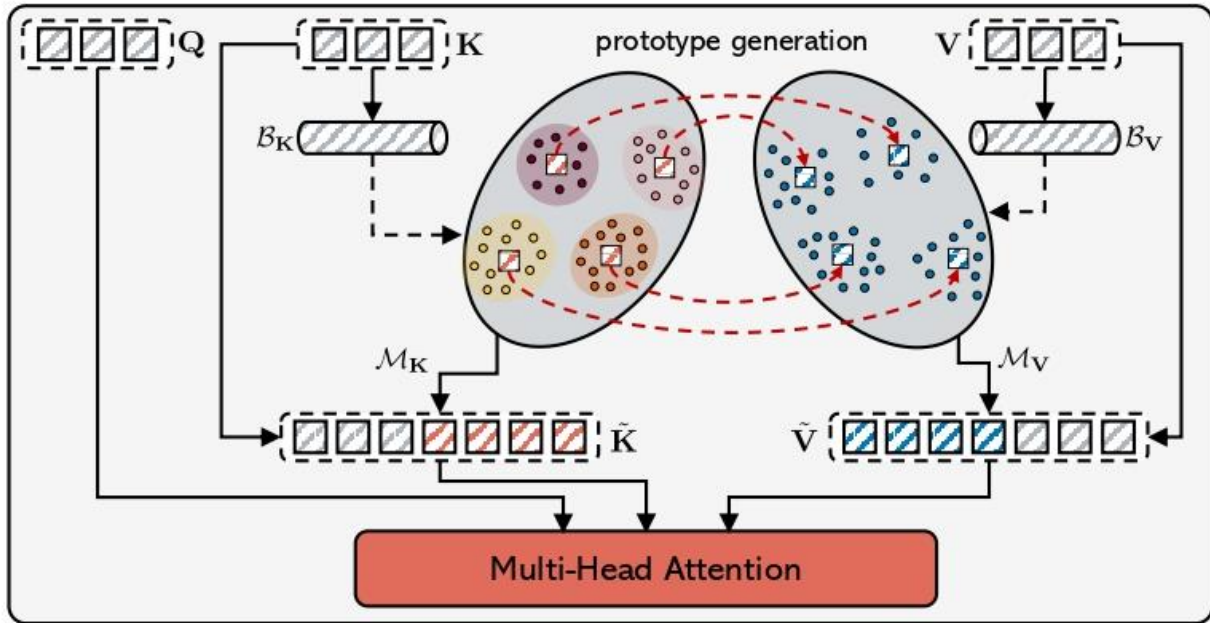


While the quality of generated captions has seen notable improvements, the automatic evaluation of captions has also witnessed a significant effort but until now has relied on metrics that utilize few human references or noisy web-collected data. Furthermore, these standard metrics often failed to align closely with human judgment. To bridge this gap, we propose a novel evaluation metric based on a *positive*-augmented contrastive learning that allows us to reach the greatest correlation with human judgment.

This new and efficient metric is called **PAC-S** and is just a consequence of a fine-tuning of the **CLIP architecture** using as augmentation some positive examples generated using two synthetic generators for text and images (**BLIP** and **Stable Diffusion**, respectively).

With our advancements in the evaluation aspect, we shifted our focus towards enhancing the caption generation task and we devised another augmented architecture called **PMA-Net**.

The idea started noticing that the attention operator, mainly used in captioner, is not able to attend past training examples, reducing its effectiveness. To address this limitation, we propose a prototypical memory network, which can recall and exploit past activations. In this case it is a memory augmentation in which the memory is fully integrated in the



Prototypical Memory Attention

architecture and represents past knowledge processed by the network itself. With this augmentation we surpass the state-of-the-art approaches. The results of both these works demonstrate the

effectiveness of augmented architectures in tackling visual and language tasks, shedding a light on potential future directions that could be done to improve both sides of the captioning task.





THE RSC PRESENTS THE CLASS OF 2023



PINEAU, Joëlle | School of Computer Science, McGill University

Joëlle Pineau has made deep, prolific and impactful contributions to machine learning and reasoning and planning under uncertainty, developing theoretical and algorithmic foundations, as well as innovative applications in a wide range of sectors including personalized and robotassisted healthcare and human-computer interaction. Professor Pineau is a leader of her international research community, whose work has significantly improved experimental machine learning practice, and a champion of responsible artificial intelligence.



*Academy of
Science*

On Friday, November 17, the 2023 Class of 101 new Royal Society of Canada (RSC) Fellows will be inducted for outstanding research and scholarly achievement. Among them, our community is proud to celebrate Joëlle Pineau, McGill University. The above is the full quote from the RSC's nomination. *“Honored to be elected to the Royal Society of Canada,”* Joëlle said in a humble statement. *“My biggest gratitude to my amazing students and wonderful colleagues at McGill, Mila and Meta who have made this possible - and have been by my side throughout this journey in research!”* Congrats, Joëlle!

Hilde Kuehne is an Associate Professor at the University of Bonn as well as an Affiliated Professor at MIT-IBM in Cambridge.

Read 100 FASCINATING interviews with Women in Computer Vision

*... every project is kind of unique!
Every project has ups and downs!
I could not say, okay, this was especially
a nightmare because of blah ...*



What is your work about, Hilde?

I'm working on everything in multimodal learning at the moment. Technically, it means trying to figure out how we can learn from different modalities and across different modalities. So we started with video, where it's obvious that you have more than one modality. Video is not only the vision part; it also has audio. In most cases nowadays, it comes with ASR. So there's also text! Actually, the interesting thing is that we figured out that one modality can actually be used to enhance learning for the other. I think that's a super cool thing because, similar to vision language models, it kind of frees us a bit from having to use annotation. It also opens up the space for anything free text. I think that's super cool!

Tell us why it is super cool.

Oh, that's a hard one. *[laughs]*

I am here for the hard ones!

Okay, so I especially come from video understanding and action recognition. One problem that we have with actions, probably more than with objects or anything else in the world, is that they are very hard to describe. People usually have a very good understanding of what an object is like. A mug is a mug, period. But actually, understanding actions highly depends on your world knowledge, on your expert knowledge for a specific task, and so on. Therefore, describing actions

by pure categories usually works for a certain subset of tasks. This is what we have in current data sets, but it's usually not enough to capture all actions that are going on in the world. Therefore, moving away from pure classification, especially in the context of action and video understanding, is very important. First, having foundation models that actually transfer much better than what we have at the moment, and second, actually to get closer or to do even more for real-world applications.

I understand now why it is cool. Is it cool enough to dedicate the best years of your career to research?

[laughs] Absolutely!

So what is best, teaching or researching?

[hesitates a moment...] Both have good sides and bad sides. If I had to choose at the moment, I would probably say research. However, teaching and research are not separate for me. I mean, obviously, there are lectures. But technically, teaching and research happen together. When we have good Master's students or even PhD students, and they do research, technically, we also teach them on the fly how to be good researchers. This is something that I really love. So, actually, it's both.

Isn't it funny that most of the research is done by people who are not yet proficient in research? They are just learning to research.



[hesitates a moment...] Perhaps yes, perhaps no. The interesting thing that we have is that a lot of the people who are doing the real research are researchers in training, if you want to see it like that. But let's say the interesting thing is when you look at what those people then later do, either they move on to industry and apply what they have [learned] to build crazy good products, or they actually move on to academia and start educating the next generation of researchers. In this sense, it makes sense that it's kind of like a self-reinforcing system.

You already have a few years of research behind you. Maybe you want to tell me what you consider

your best find till now. What are you most proud of?

Well, I have done some data sets, and I'm still surprised that they are still around by now. I would have guessed that each of them would last probably for two to three years, and then they would be replaced by something way cooler. They are both still around, and I don't know why.

Oh, you can mention them! We are not shy.

[laughs] Okay, I have done HMDB and Breakfast. It's actually very cool to see that people still find them useful. However, when you ask me what's the most important thing that I have done, honestly, it will always be one of my current projects. So, the current ones are always the most important to me, no matter what I have done in the past.

What are the current projects? Can you share something with us?

Yes, all the projects that I do at the moment are about multimodal learning. Technically, they all somehow deal with this question of how to bridge modalities. With this respect, many of them are actually not so much about building new architectures but understanding what current systems are doing and how to make this better. One of them is, for example, a paper that will be published at ICCV about learning by sorting. For example, we show that by changing the loss



function, we can learn embedding spaces that are better suited for K-nearest neighbor retrieval. And as retrieval is one of the cornerstones for multimodal learning, this is something that's actually pretty cool. We have a lot of interesting papers on the usage of language together with video or how we can actually make text better for video. We have some interesting ones, which will hopefully be available soon, but I cannot talk about them at the moment. *[laughs]*

Let's tell the ICCV people that they should come to the poster of Nina Shvetsova, Sirnam Swetha and Wei Lin. Come to the three posters of these young and fine people and ask questions. You might, by

chance, find Hilde there. Three posters are no mean feat!

Absolutely! Actually, there are four posters.

Which is the fourth?

Nina has two. Nina has Sorting and In-Style.

Okay, so you will have to tell me about Nina, who was able to get two first-author posters at the same conference... What is special about her way of working?

[laughs] Well, let's first say Nina is great! Nina is also working with me. Nina is my first PhD, so it's always something special. And first, just to not overstate, the sorting paper was a lot of hard work, and it got rejected twice or even three times. Whoever gets rejected always resubmits and makes it better. At some point, it will work. But the second thing is this In-Style paper, which is a bit more about this research on how to use language models to make video annotations better. So that's generally Nina's idea; it's all hers. I think it's super cool work, and hopefully, it helps the video community to solve a few of our problems.

What is the most difficult thing that you have done in this field until now?

Oh, my God. That's a good question!

Thank you.

[laughs] Um, I don't know, actually, because every project is kind of unique!

Every project has ups and downs! I could not say, okay, this was especially a nightmare because of blah... There is no crazy outline or crazy point where I would say, okay, this was a nightmare because of this.

Did you ever think, “This is too tough, I give up”?

[laughs] One thing that I always wanted to do that never worked, and I would still love to, is actually binary networks, like real binary neural networks.

Why don't you do it?

Because it's tough, it's just a super tough problem.

Perhaps some ambitious researcher in this community will say, I want to do that and will ask you for advice.

My advice is probably to do something else. [both laugh]

What is the best advice that you have given, and what is the best advice you have received?

[hesitates for a moment] So, the best advice I ever received in my life was probably when I was considering studying computer science I was



“If you want to study computer science, go home, start programming, and if you love it, just come back, then you're right for this!”



definitely not planning to do computer science in the first place. Actually, I was leaning more towards arts and design. But I went to the study counsel of the university, and he told me, *“If you want to study computer science, go home, start programming, and if you love it, just come back, then you're right for this!”* Obviously, I went home and started programming. I loved it, and so I went back, and that's the rest of the story. I guess I don't know what's the best advice I ever gave to people because I'm randomly blurring out stupid stuff all the time. [laughs] You only have to ask people what's the best advice they ever got. But I think if I had to give advice, it would be exactly that. If you want to do something, if you consider doing a PhD, try publishing. If you love it, come back.

Can you tell us about the MIT-IBM Watson AI lab?

First, the lab is a collaboration between IBM and MIT. Technically, it's a very interesting lab because it's

an industry lab, but it's run in a very academic way. So, it sometimes feels more like academia than industry. And the reason for that is our work is mainly project-based, like in academia. We have to hand in proposals for projects like in academia. Each project is then actually headed by one Pi from MIT and one Pi from IBM. So, it's always both sides involved. I think this makes it a bit special and super interesting as it's exactly at this intersection between academia and industry. This is actually where I feel most comfortable because I really like academia, I really like industry, and I always looked for a place where I can have a balance between both of them. I don't want to be 100% on one side. I also don't want to be 100% on the other side. I always want to be in between somehow.

Elementary, Mr. Watson! You have found the right balance between both.

Exactly!

We have spoken about the present, and we have spoken a little bit about the past. Let's speak about the future. What is your future?

As I just started in Bonn, I guess currently, it's mainly settling. Actually, I am starting to hire more people because I also got an ERC starting grant last month.

Oh, very nice! Do you need people?
Yeah.

PhD students or postdocs?

Actually, both.

Guys, if you read this and you are

interested, you have an incredible chance to work with awesome Hilde Kuehne. Don't miss it, or both Hilde and I will be disappointed. [both laugh] BTW, you are also a program chair at the upcoming WACV, and this is a baby that is very dear to your heart. I have one last question for you, Hilde. It's about ICCV. What do you expect from the upcoming conference?

I'm super looking forward to the workshops and to the poster session. I have to say, I love the poster session. I will try to stop by every video-related poster, I promise!

**Read 100 FASCINATING interviews
with Women in Science!!!**





Don't miss the BEST OF ICCV
and the BEST OF MICCAI
in Computer Vision News of November.
Subscribe for free and get it in your
mailbox!
[Click here](#)

A door to open the way for tomorrow's women to the professions of the future: this is the mission of AlxGirls – Summer Tech Camp.

AlxGirls is an Italian summer training camp that offers young female students the tools to be leading actors of the Fourth Industrial Revolution.

by *Darya Majidi, Roberta Russo, Monica Cerutti, Sara Moccia*

AlxGirls is open to talented Italian high-school female students aged between 17 and 18 years old. The participation in the **AlxGirls – Summer Tech Camp** gives the students the opportunity to take part in an interesting study program that will touch on the ethical and technical aspects of new technologies relevant to artificial intelligence (AI). The program focuses on AI and data science, with the goal to offer the participants all necessary skills to be an active part

of the change and overcome the limits of the gender gap.

The Summer Tech Camp was conceived in 2022 by **Darya Majidi**, CEO of *Daxo Group* and founder and president of *Donne 4.0 Association*. The idea came from the fact that women and girls are under-represented in the ICT sector. Only 1.4% of graduates in Italy study these subjects and only 14% of them are girls. The idea is therefore to bring the most talented girls closer to the knowledge of AI and aware of its impact on the jobs of the future.





Each year twenty female students are selected among those from all over Italian high schools through a competitive selection process. The selection takes place with a national call, usually in May, and the girls are evaluated on quantitative criteria, such as grades, and qualitative criteria, such as the motivation to participate in the Camp. This year, we made a call in May and contacted all high schools in Italy. Other than girls from scientific high schools, this year we have also selected girls who are studying at classical high school and other curricula.

The Camp had its first edition in July 2022 at the International School of Higher Education (SIAF) in Volterra (Italy). SIAF was jointly conceived in 1999 by Scuola Superiore Sant'Anna di Pisa, Cassa di Risparmio di Volterra and Fondazione della Cassa di Risparmio di Volterra, as a

training facility based on residency and the availability of a particularly functional and welcoming location, located in one of the most fascinating places in Tuscany. SIAF is a unique structure of its kind in Italy, with a Campus that can accommodate up to 200 guests offering classrooms, recreation areas, services, and teaching aids.

Each year, the program of AixGirls ranges from basic knowledge of AI (supervised and unsupervised learning, deep learning, natural language processing and image analysis) to ethical, legal, sustainability and social impact aspects relevant to the use of AI in public sectors, as healthcare, and in industry. The faculty is made of a unique talented team of women coming from universities, startups, corporates all belonging to the Donne 4.0 association.



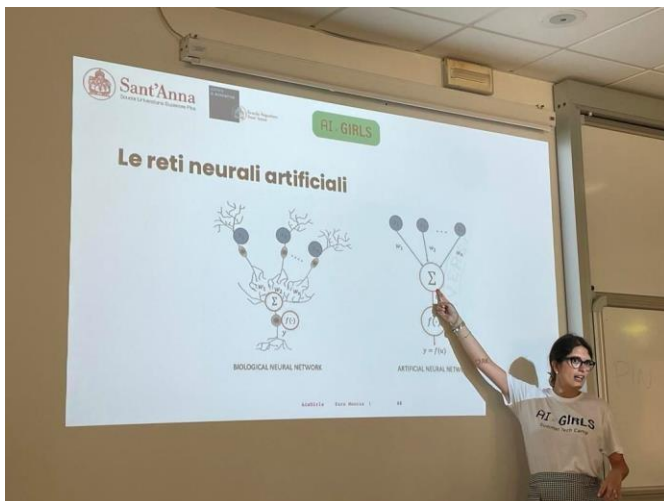
The mornings are usually dedicated to theoretical lessons, covering also aspects relevant to the impact of AI on **Sustainable Development Goals (SDG) of the UN2030 agenda**. *“The afternoons instead are “hands-on” labs during which the girls first learn how to develop an AI app and then, divided into groups, participate in a hackathon,”* Roberta, who led all the afternoon sessions and guided the girls with a lot of professionalism and passion, said.

This year, the results have been fantastic, above all expectations. There were some girls who had computer science and coding skills, but in groups they managed to develop very interesting applications. *“The girls had to think about a real problem they wanted*

to solve (also inspired by the SDG) and the to develop a simple, but fully functional, version of an app to solve that problem and prepare the slides to present their project to Fineco Asset Management CEO, who joined us from Dublin on Friday afternoon.

Events of this type are essential not only to show girls the great potential of technologies and digital, but also as an inclusive environment, with many female role models” - Roberta concluded.

During the camp, it was crucial to convey to the girls the “awareness” of how much gender stereotypes continue to influence their life paths, when instead it would seem to be issues related to the past: statistics show that in 2020 those



enrolled in Computer Science and ICT represent only 15% and the average salary 5 years after graduation is around 300 euros higher for men! If this data is used by AI systems without cleaning and evaluation, the biases will not only not be overcome, but amplified. "The starting point to counter this drift is to build a collective conscience. The AlxGirls community can become a point of reference for the new generations to achieve gender equality" - Monica said with conviction.

"AlxGirls reminds me of when I was 17 and I participated in a summer camp on robotics and biomedical engineering organised by Scuola Superiore Sant'Anna (Italy) at SIAF." - Sara said "At that time, [the arrow

shows her in the top right photo] I was still unsure of the university curriculum I would have undertaken in a year. I remember I was impressed by both the enthusiasm of the researchers and the research innovation I was completely unaware of. In those days, I decided to become an engineer working in the medical field. After 13 years, being on the other side, as an assistant professor in AI for medical image analysis and lecturer during AlxGirls, fills me with joy. It is like closing a circle."

The Camp is promoted by Fineco Asset Management, the Donne 4.0 Association and Daxo Group, which strongly aim to support female leadership in the professions of the future.



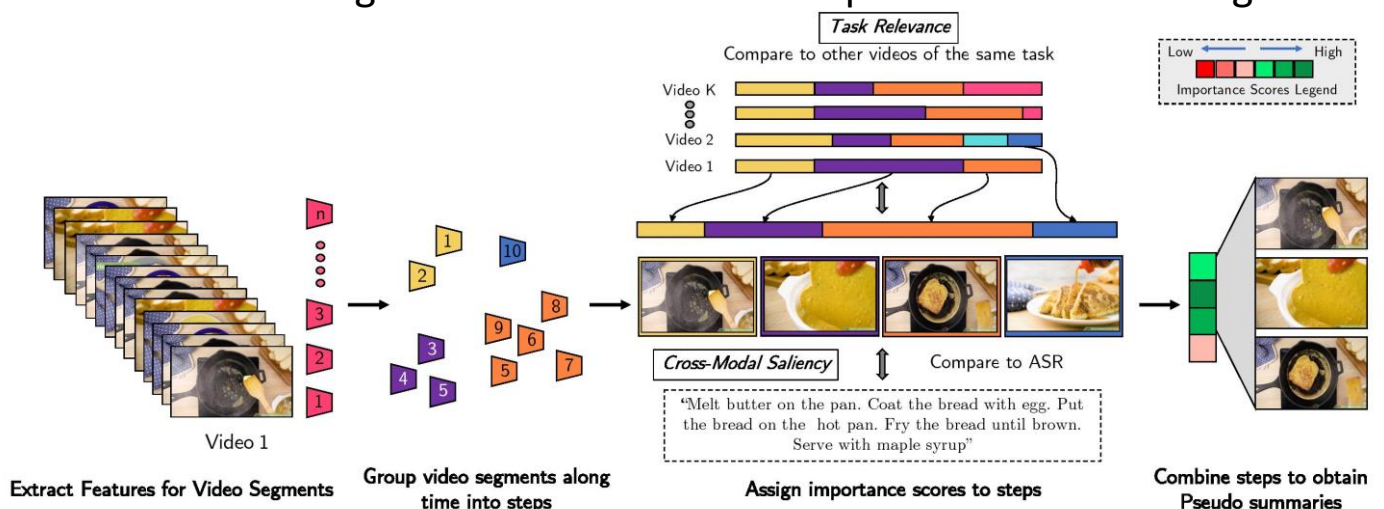


Medhini Narasimhan recently obtained her PhD in Computer Science from UC Berkeley under the supervision of [Trevor Darrell](#).

Medhini's research focuses on learning multimodal representations for long videos using little to no supervision, by modeling correlations across the different modalities. Specific applications of her work include creating short visual summaries of long YouTube videos, synthesizing longer videos from short clips, and parsing semantics of instructional videos.

She is currently a Research Scientist at Google Labs with Steve Seitz, continuing her research on video understanding, while also developing innovative products. **Congrats, Doctor Medhini!**

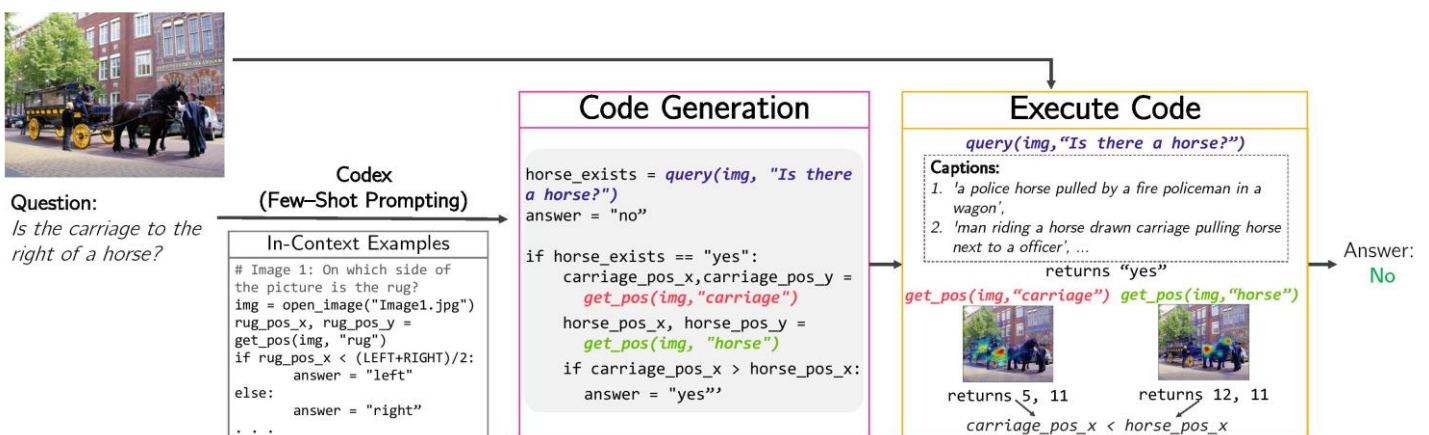
The internet hosts an immense reservoir of videos, witnessing a constant influx of thousands of uploads to platforms like YouTube every second. These videos represent a valuable repository of multimodal information, providing an invaluable resource for understanding audio-visual-text relationships. Moreover, understanding the content in *long* videos (~2 hours), is an open problem. In her PhD thesis, Medhini investigates the intricate interplay between diverse modalities—audio, visual, and textual—in videos and harnesses their potential for comprehending semantic nuances within long videos. Her research explores diverse strategies for




combining information from these modalities, leading to significant advancements in video summarization and instructional video analysis.

The first part of Medhini's thesis introduces a non-parametric approach to synthesize long videos from short clips by using representations learned via contrastive learning. This is achieved by repeatedly stitching together segments of the short video coherently to create dynamic yet consistent outputs. A learned distance metric is used for choosing segments, which allows for comparing clips in a manner that scales to more challenging dynamics, and to condition on other data, such as audio. In the next section, Medhini introduces her work CLIP-It which is a novel technique for generating concise visual summaries of lengthy videos guided by natural language cues. Specifically, a user-defined query or a generated video caption is used to create a visual summary of a video that best matches this natural language prompt. Next, she focuses specifically on summarizing instructional videos, capitalizing on audio-visual alignments between the narration and actions in the videos and similarity in the task structure across multiple videos to produce informative summaries. Fig 1 illustrates this method of creating a video summary using no external supervision.

To further enrich the comprehension of instructional videos, she then introduces a cutting-edge approach that facilitates the learning and verification of procedural steps within instructional content, empowering the model to grasp long and complex video sequences and ensure procedural accuracy. Lastly, her work explores the potential of large language models for answering questions about images by generating executable Python code. This involves first defining modules which are useful to answer questions and which use pre-trained image-language modules in the background. As seen in Fig 2, using a few sample prompts, an LLM can be instructed to orchestrate these modules into meaningful code snippets which can be executed to answer questions about the image in an explainable fashion. Currently, her research efforts are being directed towards exploring use of large vision and language models for video understanding.

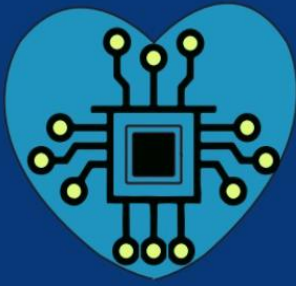


The image features a top-down view of a desk with various items. In the top left is a small potted plant. In the top center is a white mouse. In the top right is a box for a 'VIRTUAL MICCAI 2021' event. Two 'MICCAI 2021 DAILY' magazine covers are prominently displayed. The 'Wednesday' cover (September 27 to October 1) features a blue background with a city skyline and a robotic arm. The 'Thursday' cover (September 27 to October 1) features a blue background with a city skyline, a diagram of a brain scan, and two anatomical images of a heart. A dark blue banner with white text is overlaid on the center of the image. Below the banner is another 'MICCAI 2021 DAILY' magazine cover, this one for 'Computer Vision News', which includes a grid of images and graphs. In the bottom left is a white sticky note with the text 'click it!'. In the bottom right is a small container of colorful pens and pencils.

Are you going to miss MICCAI 2023 in Vancouver? You can be in touch anyway!

Follow MICCAI almost in real time:
[Click here](#) and subscribe for free
to MICCAI Daily (9-10-11 September)
with all the highlights from MICCAI

click it!



MEDICAL IMAGING NEWS

OCTOBER 2023





Wolfgang Wein is the Founder and CEO of ImFusion, a German technology company established over a decade ago that blends software development and licensing with consulting and research to address the unique needs of its customers.

ImFusion has crafted a software framework encompassing accelerated platform-independent libraries, front-end labeling tools, and domain-specific plugins **to help medical device companies transform cutting-edge research into innovative, minimally invasive surgical solutions that rely heavily on medical image computing.** Its commitment to not reinventing the wheel sets it apart from its competitors.

“We don’t have to advertise much,” Wolfgang begins. “There are more and more requests for what we do, which means the need is there. We’re very advanced in medical imaging and guided surgery.”

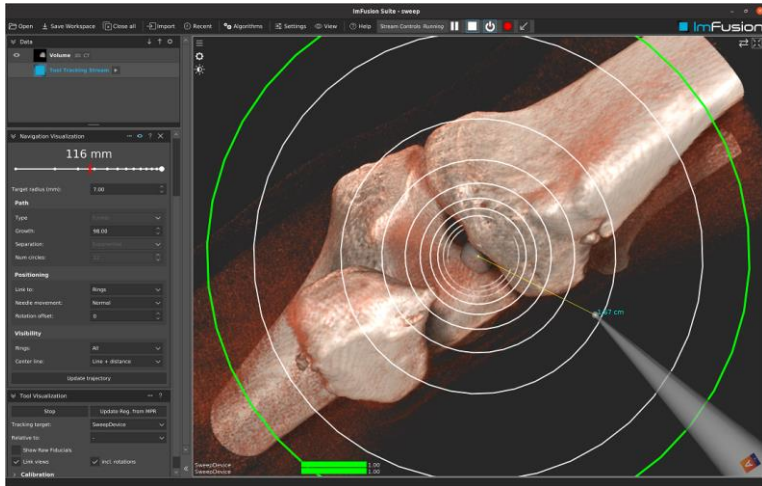
The team was initially composed primarily of engineers and PhD experts in the field; now, it has product managers and a back-office operations team. A unique blend of academic innovation, advanced software engineering, high-performance computing, and efficient numerics is at its core. Its software development kit (SDK) reflects this, offering customers a product-grade experience and streamlining the development process.

All this innovation is not without its challenges, but Wolfgang tells us the company’s adaptability and foresight have allowed it to navigate any hurdles successfully. While its core framework remains robust, ImFusion constantly monitors trends within different technology groups.

“There’s global competition, and the environment is very fast-paced,” he points out. “First and foremost, in machine learning, we must be very selective about what we implement ourselves and where we rely on large frameworks that the global community has adopted. Many years ago, we implemented our own random forest framework that performed better than the one available through OpenCV and others. Now, we’ve removed it from our build because of superior technology.”

Another shift he has observed is the development community’s **growing preference for Python over C++.**





ImFusion is acutely aware of this trend and is responding by enhancing its SDK with powerful **Python bindings**, which enables users unfamiliar with C++ to harness the capabilities of its C++ program libraries, ensuring it remains at the forefront of high-performance computing.

ImFusion's operations are distributed across seven departments, with **computer vision** being one of six technical divisions alongside machine learning, ultrasound, computed tomography, robotics, and SDK.

*"You could say that **almost everything relates to computer vision**," Wolfgang adds. "In the vision group that **Alexander Ladikos** leads, we focus on RGBD, real-time point cloud and geometry processing, endoscopic image processing, and industrial vision in some projects. Computer vision is broader than that, and you also have projects where this is combined with robotics. Some of the endoscopic image processing is associated to our computer vision*

group. We want to cover all interventional modalities. We also need to use ultrasound and X-ray and be at the absolute state of the art in processing all of those, which requires heavy computer vision."

ImFusion is actively involved in transforming patient care, with 90% of its business centered in the medical sector. A range of customers have been able to commercialize fully with its help, and devices running on its framework are being used on patients in surgical settings every day.

"That makes us very happy," Wolfgang smiles. "It also creates some of the most interesting and exciting research problems. There are certain interventional images in edge cases where we barely see anything in the images, so from all these regressions of the real-world use of our software, we can improve it even further."

This dynamic environment allows ImFusion to enhance its solutions continually, a unique experience that many in academia seldom encounter. The company also dedicates time each year to developing new technologies and methodologies, publishing its findings at leading conferences, including **MICCAI** and **MIDL**, and sharing discoveries with the global research community. As well as sponsoring the event, **it has two papers on the MICCAI program this year and its is staffing a booth.**

With over 40 full-time employees and organic growth spanning 11 years, ImFusion has created a strong foundation, but it remains vigilant about potential challenges. The loss of key customers and critical personnel is recognized as a possible risk. However, Wolfgang considers this unlikely, citing happy clients and an ambition to be one of the best places to work in Germany, aided by a new HR colleague focused on improving employee benefits and overall long-term satisfaction. Also, with a rising interest in what it offers, the company continues to focus on hiring new people.

“... it has two papers on the MICCAI program this year and its is staffing a booth ...”

“You could say there’s a risk in people licensing software for this because of the fast-moving pace of all the libraries and fundamentals changing,” he ponders. “Some large players, the big tech companies, NVIDIA, and others keep developing powerful libraries. We interface well with them, and we recommend them. It might be that, at some point, the focus will have to be less on our own software. We want to focus more on that, but if it doesn’t work, we’d retreat back to contract work. Therefore, overall, we’re

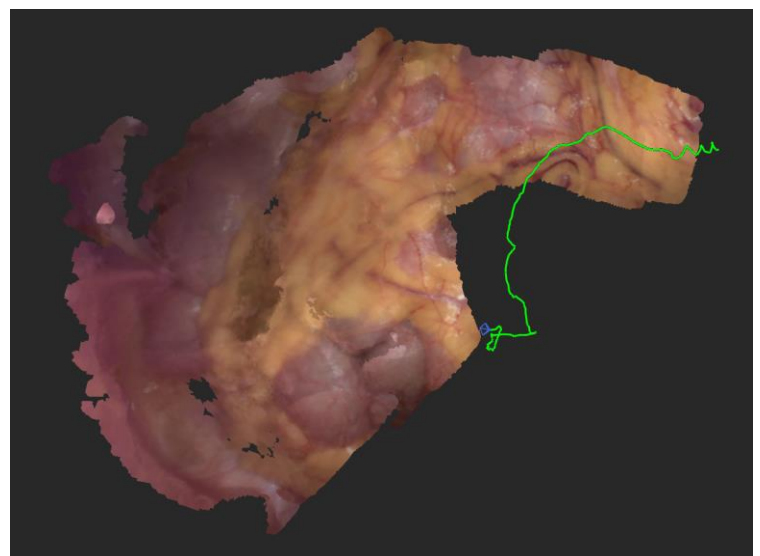
actually not a very risky business. I guess we’re not much different from [RSIP Vision](#) anyway!”

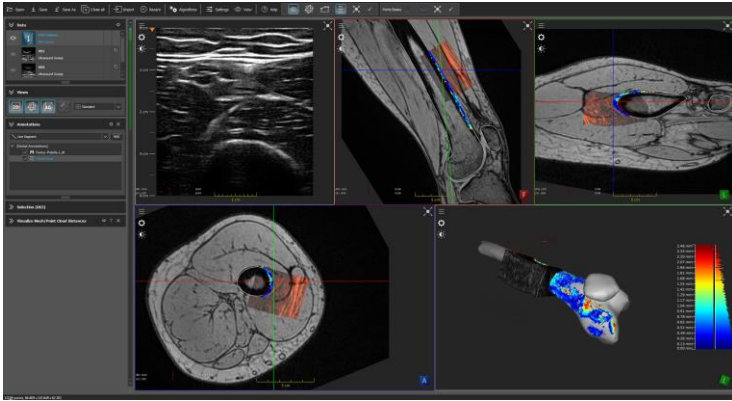
ImFusion is bootstrapped, meaning it operates without external investors, providing the freedom to move in the direction it feels would most benefit existing and future clients.

“We want to be sustainable and are not geared toward any quick exit,” Wolfgang continues. *“We’re serving multiple customers in a very responsible way with long-term relationships.”*

While the company’s role as a subcontractor for larger corporations can be volatile, he points out that its flexible approach, experience, and size allow it to navigate unexpected changes in project priorities and funding with relative ease. What might be a setback one year could be an opportunity the next.

“Nowadays, what’s important is that you have people who are the





absolute cutting edge in medical imaging, numerics, math, linear algebra, programming, C++, and object-oriented programming, and you combine engineers with data scientists,” he asserts. “You always need both. You cannot do a product in the medical space with one.”

“... you combine engineers with data scientists. You always need both. ...”

Understanding algorithms and a mix of smart engineering and numerical programming is essential in the highly regulated medical space, and this is baked into ImFusion’s framework. A dedication to collaboration with academia, medical device companies, and clinicians helps it to stay innovative.

“We cherish being exposed to academia and want to continue doing that,” he points out. “We’ve found some of our best talent through academic outreach. That’s why we’re very happy to be at MICCAI again.”

Wolfgang acknowledges the


profound impact of his own academic journey, telling us he owes a lot to his PhD advisor, [Nassir Navab](#), and the community of professors and research groups around the world who raised him in an international environment. He remembers how exciting it was to be a founding member of Navab’s group in Munich.

“A piece of advice he gave me very early on because I worked at Siemens after my PhD was to keep my academic profile,” he recalls. “Keep publishing. Keep giving invited talks somewhere so that people know you. We now live this in our company as well. We try to keep publishing. That means my team’s market value and visibility go up, so they could go elsewhere, but that’s fine. They have a public profile and are happy and highly respected. That gives us the possibility for growth. If you compare us to Google or Apple, we’re more like Google – less secretive and restrictive.”

If you would like to meet Wolfgang and the team and find out more about ImFusion, visit their booth at MICCAI 2023!



HAPPINESS IS... HOLDING TICKETS THAT SAY VANCOUVER!



October 8-12, 2023

See you at MICCAI 2023 in Vancouver.
Come over. In-person. It will be fun!



Maria Tirindelli has obtained her PhD last week at TUM, at the chair of Computer Aided Medical Procedure and Augmented Reality (CAMP) under the supervision of [Nassir Navab](#).

Since last year, Maria is working as a Research Scientist at ImFusion GmbH.

Congrats, Doctor Maria!

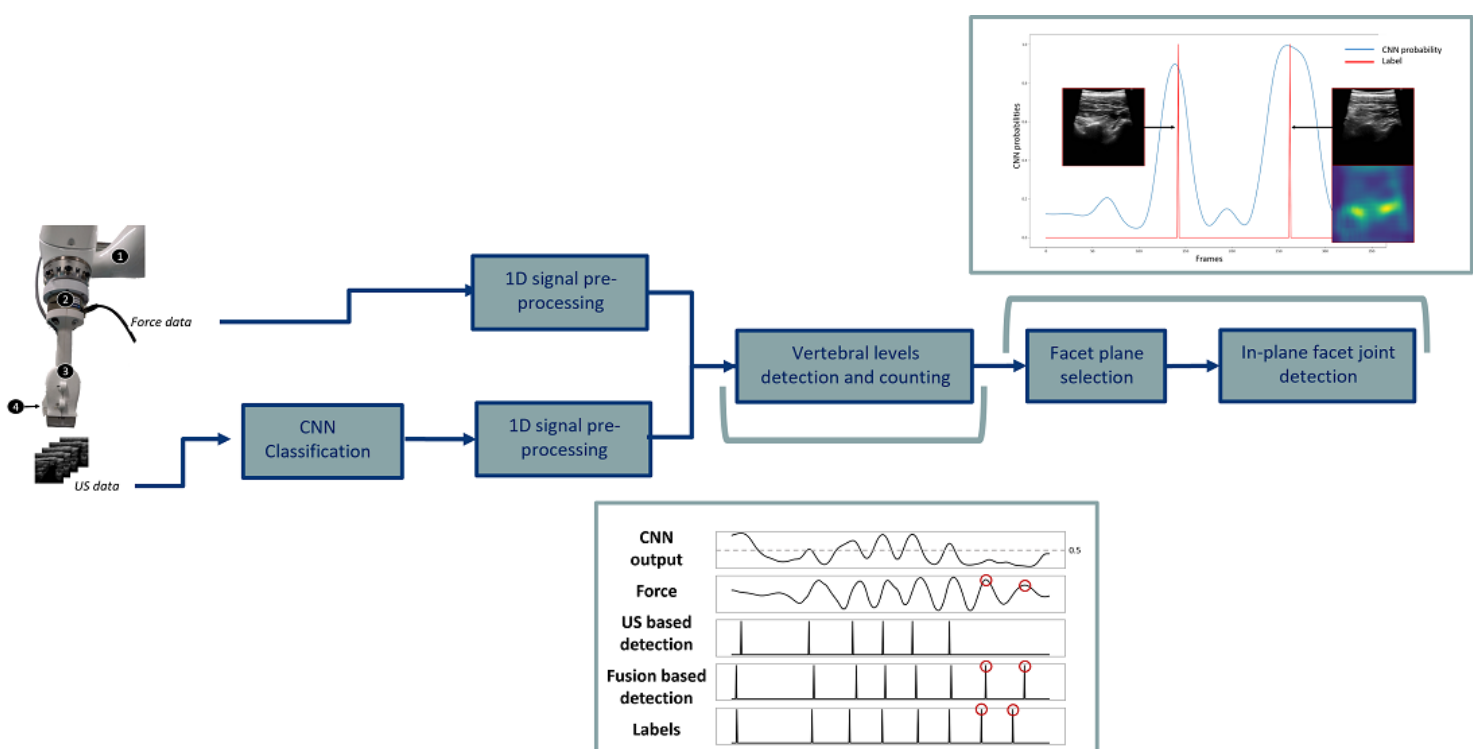
Ultrasound imaging is a known imaging technique in medical imaging, allowing for real-time data acquisition without the use of ionizing radiation. However, ultrasound imaging presents several challenges, among which the inter-patient and inter-device variability, noise, and artifacts. Additionally, the acquisition process strongly depends on the operator, making the procedure reproducibility low. To address these issues, robotic ultrasound has been proposed in the literature, to automate the acquisition of ultrasound data. Robotic ultrasound presents several advantages, as it reduces the reliance on the operator's expertise, it can guide novice radiologists in the acquisition process, and it relieves the operators from the task of manipulating the ultrasound probe. Furthermore, robotic ultrasound can be a valuable tool to enlarge the diagnostics and treatment reachability to remote areas, where the presence of medical staff can be limited. However, automatic data acquisition and interpretation remain a challenge.

This dissertation addresses the challenges of data interpretation and trajectory optimization for optimal ultrasound quality in robotic ultrasound.

In the first work of the dissertation, we propose a new method for automatic ultrasound acquisition and vertebral level identification for spinal injection. Specifically, we propose a setup consisting of a robotic arm, where a force sensor and an ultrasound probe are mounted. We then program the robot to move along the spine on the patient's back, while ultrasound and force data are acquired. Thanks to the utilized force control, we can then define a model for patient-robot arm interactions, to extract a force signal where vertebrae position can be identified along the spine. We then use a Convolutional Neural Network to extract vertebrae positions from the ultrasound data and we fuse the information, to guide the robot to the correct target.

The second work focuses on the identification of augmentation techniques that better reflect the physics of the acquisition of ultrasound data, compared to standard techniques. More specifically, we identify three possible augmentation techniques. The first technique introduces augmentations by introducing synthetic deformation, where rigid structures are left undeformed while soft tissues deform, consistent with what happens in real ultrasound acquisitions. The second technique introduces an augmentation that simulates variations in contrast and signal-to-noise ratio of the input signal. Finally, the third technique introduces an augmentation that simulates the occurrence of multiple reflection artifacts in ultrasound data. We further compare our results with the training without augmentation and classical augmentation for example segmentation and classification problems.

The final work focuses on the definition of optimal trajectories for robotically actuated ultrasound acquisition. To this end, we first define a method that extracts the optimal trajectory to ensure the maximal coverage of the volume to be acquired. Secondly, we utilize confidence maps to extract information on the scanned spatial points coverage. Based on this information, we define an optimization criterion that allows the extraction of an optimal trajectory in a way that the occluded points are reached from different orientations and the volume is properly scanned. We tested the proposed methods on three phantoms in a simulated environment, as well as on a phantom in a real-case scenario.



Foundation Model for Retinal Images



by *Christina Bornberg*
@datascEYence

Hello everyone, I am Christina and am interested in deep learning for ophthalmology which is the reason for doing the datascEYence column here in Computer Vision News! I myself work in the field and I through this format I aim to shine a spotlight on the remarkable work of fellow researchers who share the enthusiasm for image analysis focused on our visual organ.

featuring Yukun Zhou

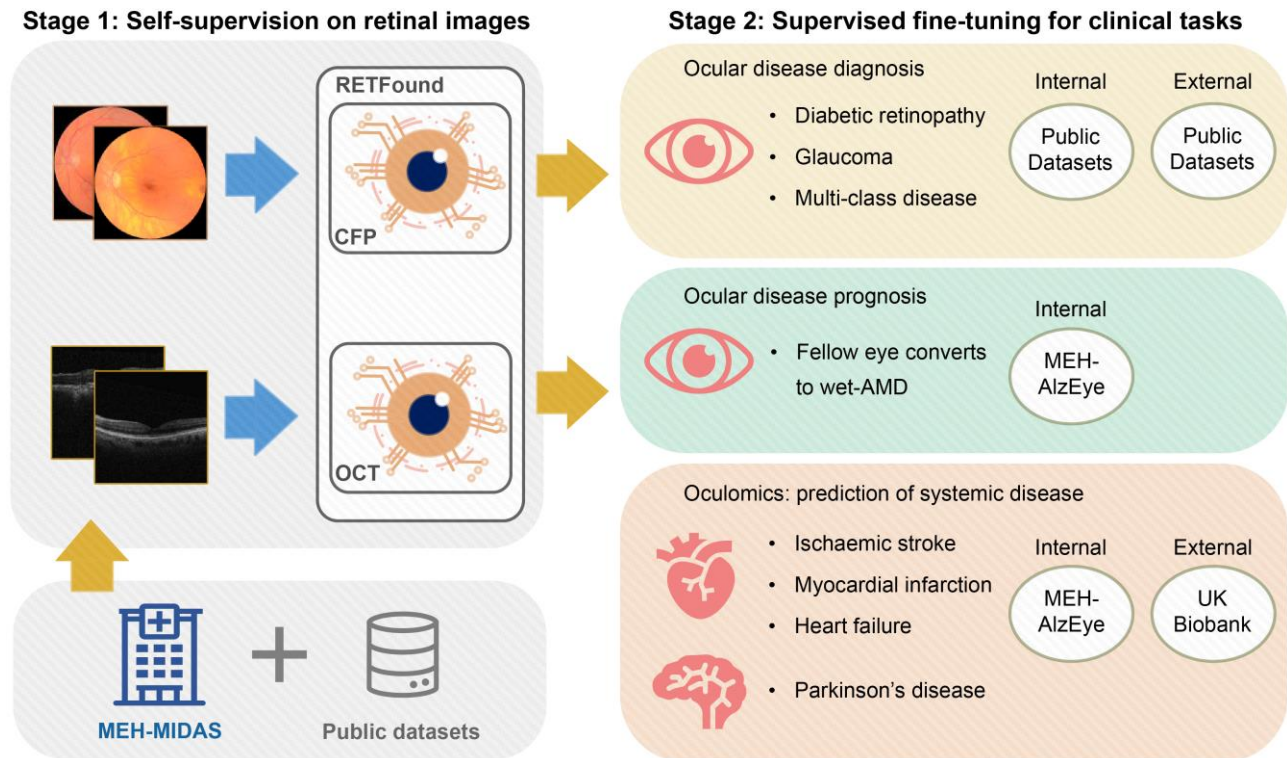
Yukun currently is a final year PhD computer science candidate at the UCL Centre for Medical Image Computing under the supervision of Daniel Alexander and at Moorfields Eye Hospital under the supervision of Pearse Keane. His journey into deep learning for ophthalmology started with a master's in mechanical engineering where he got introduced to signal processing and ended up reading papers on vessel segmentation in the eye. The similarity of image characteristics between especially color fundus images and natural images as well as his interest in learning-based methods made him want to pursue his PhD in this field.



Let's move on to Yukun's current research. If you are into deep learning in ophthalmology, you have probably already heard the news! **RETFound** was published last month in *Nature* under the title "[A foundation model for generalizable](#)

[disease detection from retinal images](#)".

I want to summarise the most important parts here and will give some special insights that I gained in the interview with Yukun!



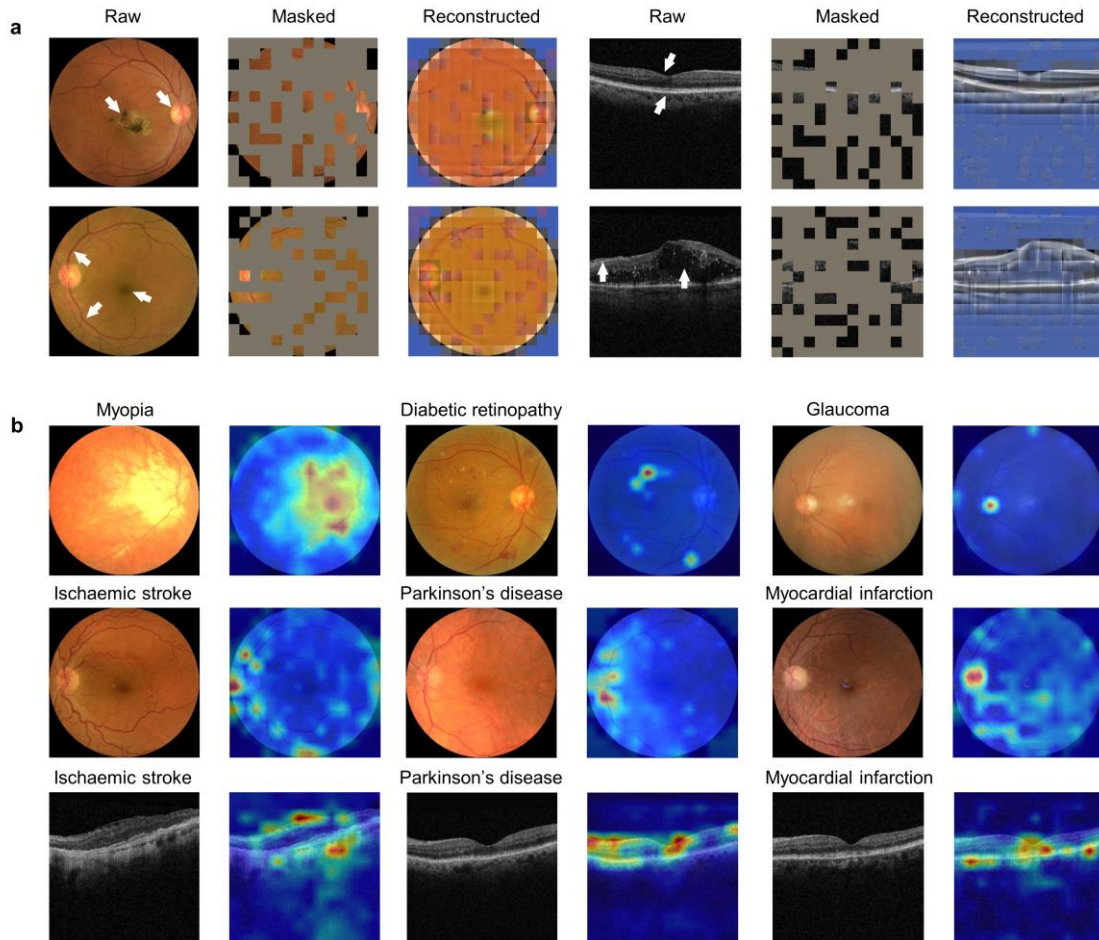
RETFound is, as the name already reveals, a foundation model. It's a model that learned a representation of the eye through self-supervised learning (stage 1 in the graphic) and can later on be fine-tuned on specific tasks without the need for a huge amount of data (stage 2 in the graphic).

The encoder Yukun applies is a large-scale vision transformer, while the decoder is a smaller vision transformer. Using an encoder-decoder architecture already gives a hint on it being an autoencoder. Specifically, they used a masked autoencoder. The addition here is the way the data is arranged. Instead of providing the network with the whole image, the image is divided into tiles and a fraction of the tiles are hidden. In their experiments, hiding 75% of the tiles was a good value for the color fundus experiment,

while for OCT images 85% of hidden tiles achieved the best results. You can see some examples in the figure!

Other non-generative techniques were actually only implemented after suggestions by reviewers. The contrastive methods (SimCLR, SwAV, DINO and MoCo-v3) ended up performing better than a supervised pre-training strategy, while slightly worse than the masked autoencoder. After all, this doesn't really matter to the researchers since they wanted to prove the general adaptability of a model trained in a self-supervised manner to a supervised downstream task - the specific self-supervised approach is just a tool.

After pretraining the autoencoder on firstly natural images and subsequently fundus images (or alternatively OCT images) the decoder is replaced by a multi-layer perceptron. For the fine-tuning on



labeled data, the whole network gets trained without freezing any layers. The downstream tasks include a variety of experiments on ocular disease classification (diabetic retinopathy, glaucoma), ocular disease prognosis (age-related macular degeneration) and oculomic prognosis (ischaemic stroke, myocardial infarction, heart failure, Parkinson's disease).

Yukun told me about the **importance of not only including ocular tasks but also oculomic tasks**. The first reason is quite technical: the goal was to verify the generalisability and adaptability of the foundation model. The second reason is more on the medical side - the eye is a window to the whole body's health condition and hence is an important organ in systematic disease understanding.

Finally, an explainability concept that computes the relevancy for transformers is applied in order to highlight regions that contributed to the classification. The method first assigns local relevance based on the Deep Taylor Decomposition principle and then propagates these scores through the layers. Some example heat maps can be seen in the figure!

If you want to reproduce/adapt/use RETFound, I have great news for you! The team released all codes (PyTorch and Keras) and the weight files which you can find in the **Code availability** section of their paper! They are also currently working on creating an application template together with software engineers from Google to minimize the operation required in use!

[More about AI for Ophthalmology](#)



Congratulations to [Maria Chiara Fiorentino](#) (third from left) for the GNB Award from Dipartimento di ingegneria dell'Informazione - Università di Padova for her doctoral thesis, "DL4US: Unlocking the potential of deep learning for ultrasound image analysis in gynecology and rheumatology"! Maria Chiara stated: *"This achievement wouldn't have been possible without the unwavering support of my colleagues, friends, and family. Your belief in me has been an incredible driving force. Heartfelt thanks to the GNB Award committee for this recognition. Research is a collective effort, and I'm eager to explore new horizons in my field. From the bottom of my heart, thank you all for your steadfast support. Stay tuned for more updates and discoveries!"*

DAICOW @ MICCAI2023



I am Camila González, a postdoctoral researcher working at the Computational Neuroscience Laboratory at Stanford University, School of Medicine. Since my undergrad days, I have been passionate about developing deep learning approaches that translate well to dynamic clinical settings.

I am excited to be organizing the first MICCAI tutorial on Dynamic AI in the Clinical Open World (DAICOW) together with a wonderful team of colleagues, to be held in conjunction with MICCAI 2023 on the morning of October 12th (starting at 8 am, but don't shy away if you can't make the early call 😊).

by Camila González

Have you ever worked with data from five, or even ten years ago? You probably noticed how different it is from more recent cases. If you did not, **your model definitely did.**

Many state-of-the-art methods for medical imaging rely on deep learning models that are susceptible to **distribution shifts**. Several factors cause changes in data acquisition, including ever-evolving scanning technologies and the presence of image artefacts. Likewise, naturally occurring shifts in disease expression and spread can cause the annotated

training base to become outdated. As a result, deep learning models **deteriorate over time** until they are no longer helpful to the clinician.

To maintain the expected performance, models **must adapt** to incorporate new data patterns while preserving their proficiency in the original evaluation set. **Continual learning** allows us to acquire new information without losing previous knowledge. This opens up attractive possibilities, such as extending the lifespan of medical software solutions and leveraging large amounts of multi-institutional data.

Yet **actually building, approving and deploying medical lifelong learning solutions faces several practical challenges.**

Our aim with this tutorial is to give participants hands-on insights into how various domain shifts affect the performance of deep learning models in dynamic environments and help them develop strategies to address these issues and correctly monitor performance. We hereby seek to breach the gap in the MICCAI community between technical research on continual learning and **the reality of deploying lifelong learning software in clinics.**

Join our tutorial to learn the technical, clinical and regulatory aspects of developing continual learning solutions. Let us take you through the process of building and deploying medical AI products that learn continuously over their lifetime in our interactive half-day event!

We will address the following topics:

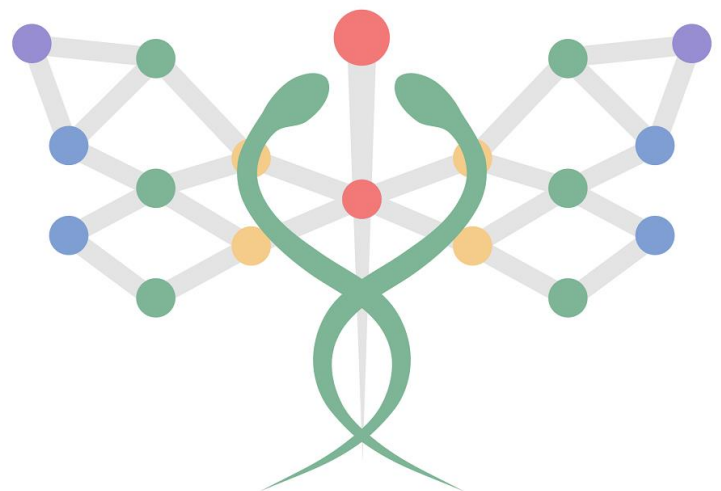
- ❖ **Data drift in medical imaging:** Common sources of domain shift and their effect on model performance, with a keynote from the fantastic **Prof. Jayashree Kalpathy-Cramer.**
- ❖ **Continual learning strategies and evaluation:** State-of-the-art methods and how to select the

appropriate strategy considering performance, flexibility and resource use, with a keynote from **Dr. Martin Mundt**, a *ContinualAI* board member.

- ❖ **Current regulations** for updating models in different global regions.

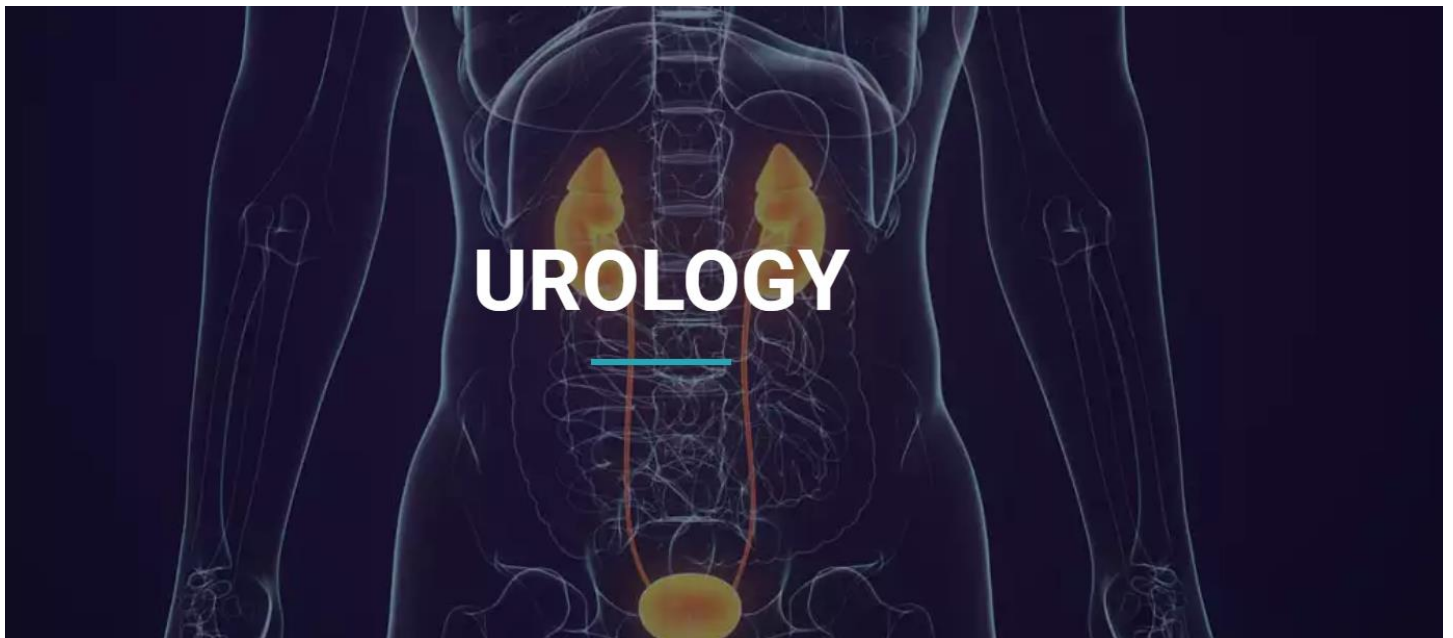
The event is aimed at a broad audience within our community. Registration is not required, but it **does** help us assess the number of participants, so **please let us know you're joining [here](#).**

Follow us on X at [@ContinualMedAI](#) for more updates 📺



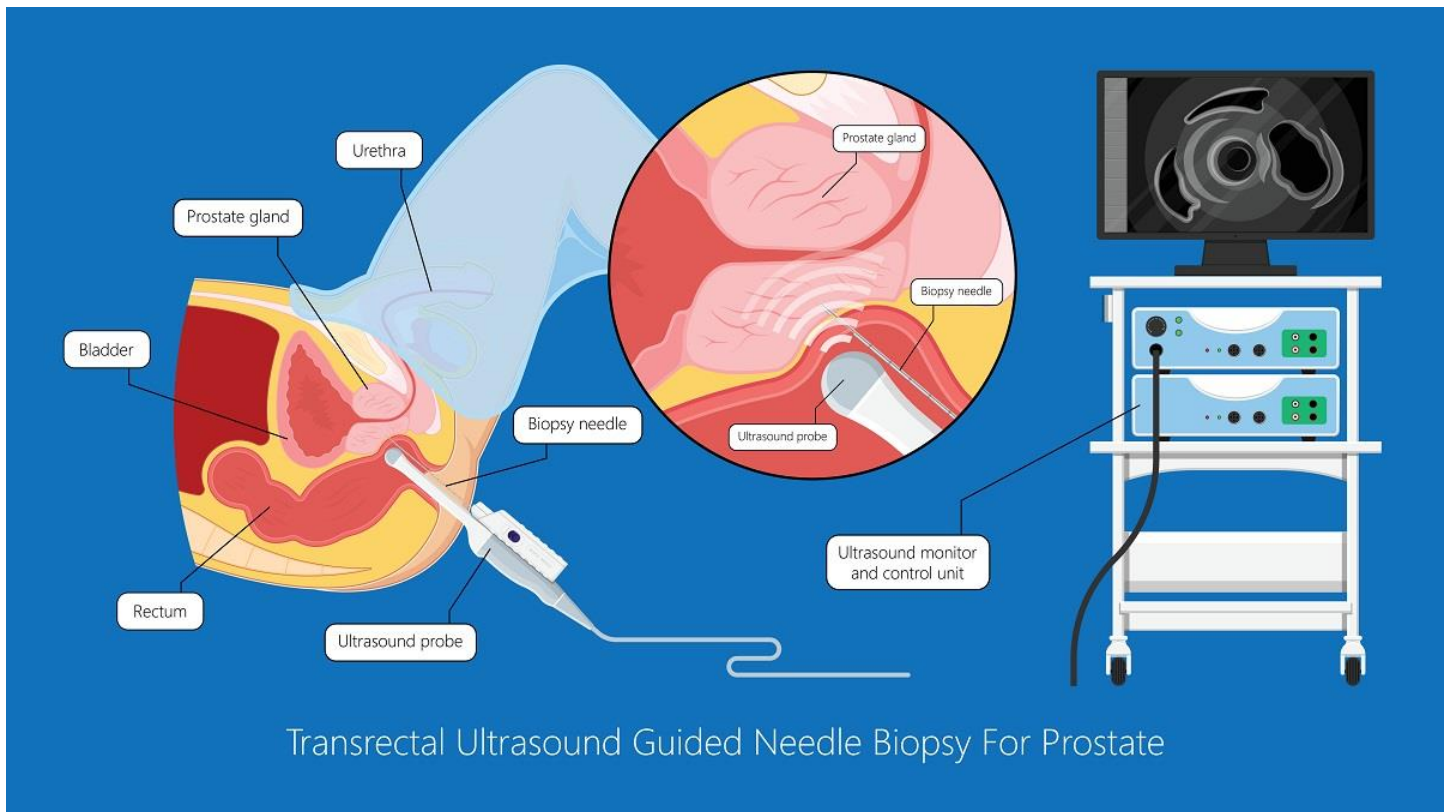
DAICOW
MICCAI 2023 

Prostate cancer is the second most prevalent cancer in men, affecting about 1 in 8 men during their lifetime. Often, patients will have an MRI scan as part of regular screening. MRI images identify regions and details such as the prostate boundary, zones and tumors, which radiologists use to calculate the Prostate Imaging Reporting and Data System (PI-RADS) score. The PI-RADS score ranges from 1 to 5 (low to high probability of clinically significant cancer) and provides the basis for diagnosis and treatment. Patients with a PI-RADS score of 3 or more usually undergo a prostate biopsy to detect suspected cancer.



Calculation of the PI-RADS score may be influenced by external factors such as experience, training or fatigue of the radiologist, which introduces variability into the process. **Artificial Intelligence** has the potential to serve as a valuable aid in **accurately and robustly calculating the PI-RADS score**, as AI algorithms output numerical values, which increase diagnosis objectivity and repeatability. For example, gradients and color distributions

can be quantified to analyze the tumor and determine the homogeneity of the tissue. The data obtained from AI algorithms assists physicians in calculating the PI-RADS score, leading to a more appropriate course of action. These advances in determining crucial details such as the size, location and state of the cancer allow physicians to choose the most suitable treatment for the patient.



Transrectal Ultrasound Guided Needle Biopsy For Prostate

MRI-generated images of the prostate and surrounding tissue may be used by physicians to improve biopsies. During a biopsy, the prostate is sampled using a needle that is guided by ultrasound to provide real-time images of the tissue. Detailed MRI scans can be registered with live ultrasound to improve the accuracy of the biopsy and target abnormal tissue. **Advanced AI algorithms** are being trained to compensate for differences between the ultrasound and MRI images to enhance registration, and thus improve guidance to tumors. In addition, probe tracking provides 3D anatomical information to enhance navigation.

At **RSIP Vision**, we work closely with physicians to improve the prognosis of [patients with prostate cancer](#). Physicians define the ground rules to detect and segment lesions, boundaries, and zones of the prostate. We incorporate these with **computer vision and deep learning techniques** to provide tools that help radiologists calculate the PI-RADS score in a more objective way and improve guidance during prostate biopsies.

[Read more articles about Medical Image Analysis and AI for Urology.](#)

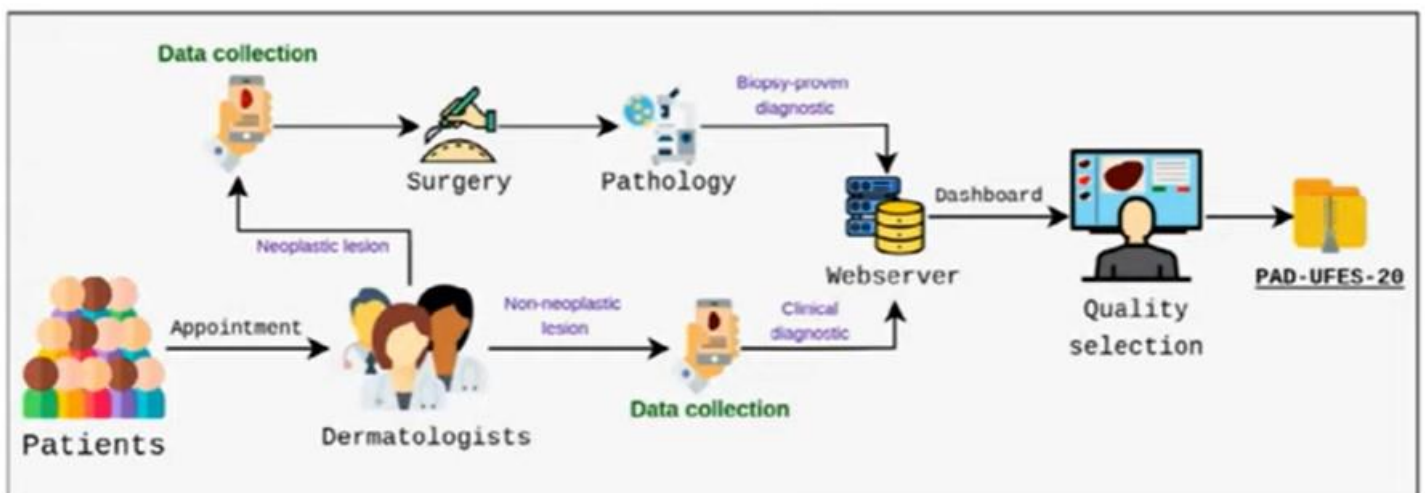


Datasets through the Looking-Glass is a webinar series dedicated to exploring the data-related aspects of machine learning methods. *“Our goal is to foster a community of researchers passionate about comprehending the profound impact of the data we employ on both algorithms and society, going beyond mere performance optimization. Our inspiration*

draws from a diverse range of subjects, encompassing data curation for dataset creation, metadata, shortcuts, fairness, ethics, and the philosophical dimensions of AI.”

The webinar is part of “Making MetaDataCount” project and is organized by Veronika Cheplygina (left in the photo) and Amelia Jiménez-Sánchez (right) at IT University of Copenhagen. The goals of the project involve the investigation of different types of shortcuts (based on demographics or image artifacts) that might occur and how these affect the performance and fairness of the algorithms, as well as investigate metadata-aware methods to avoid learning such biases or shortcuts.

Data quality selection



In our last webinar, we covered several topics about annotations regarding the integration of non-experts knowledge, the trade-off between detailed expert annotations and their cost and how to build a medical image dataset for skin lesion classification.

Andre Pacheco, an Assistant Professor at the **Federal University of Espírito Santo (UFES)**, presented PAD-UFES-20 dataset. The Dermatological and Surgical Assistance Program (in Portuguese: Programa de Assistência Dermatologica e Cirurgica - PAD) at UFES is a non-profit program that provides free skin lesion treatment, in particular, to low-income people who cannot afford private treatment.

Due to historical factors, the state of Espírito Santo witnessed an influx of thousands of European immigrants during the 19th century. Given Brazil's tropical climate, many of these immigrants and their descendants did not acclimatize well to this environment. Consequently, there is a notable prevalence of skin lesions and cancer in this region, and the PAD plays a pivotal role in providing support to these individuals.

Andre explained the challenges for building the dataset:

- ❖ Convince the doctors to collaborate.
- ❖ Design and develop applications to collect and store sensible data.
- ❖ Train doctors and students to use the app.
- ❖ Coordinate the data collection.
- ❖ Data quality selection.

Check out
the video!

Challenges

Coordinate the data collection



To learn more about the PAD-UFES project, the skin lesion dataset, and the challenges they've overcome; check out Andre's talk in the video above.

Veronika and Amelia had three successful editions so far (in February, June, and September 2023) with 10 speakers in total. The videos are available on their YouTube playlist.



Shan Lin is a postdoc at the University of California San Diego. Her research interests lie in the integration of perception, motion planning, control, and robotic manipulation, primarily aiming for autonomous robotic surgery and healthcare applications.

Shan has been selected as “Pioneer of Medical Robotics” to present her work at the Data vs Model in Medical Robotics Workshop at upcoming IROS 2023, where two stellar doctoral / post-doctoral candidates will present their bodies of work.

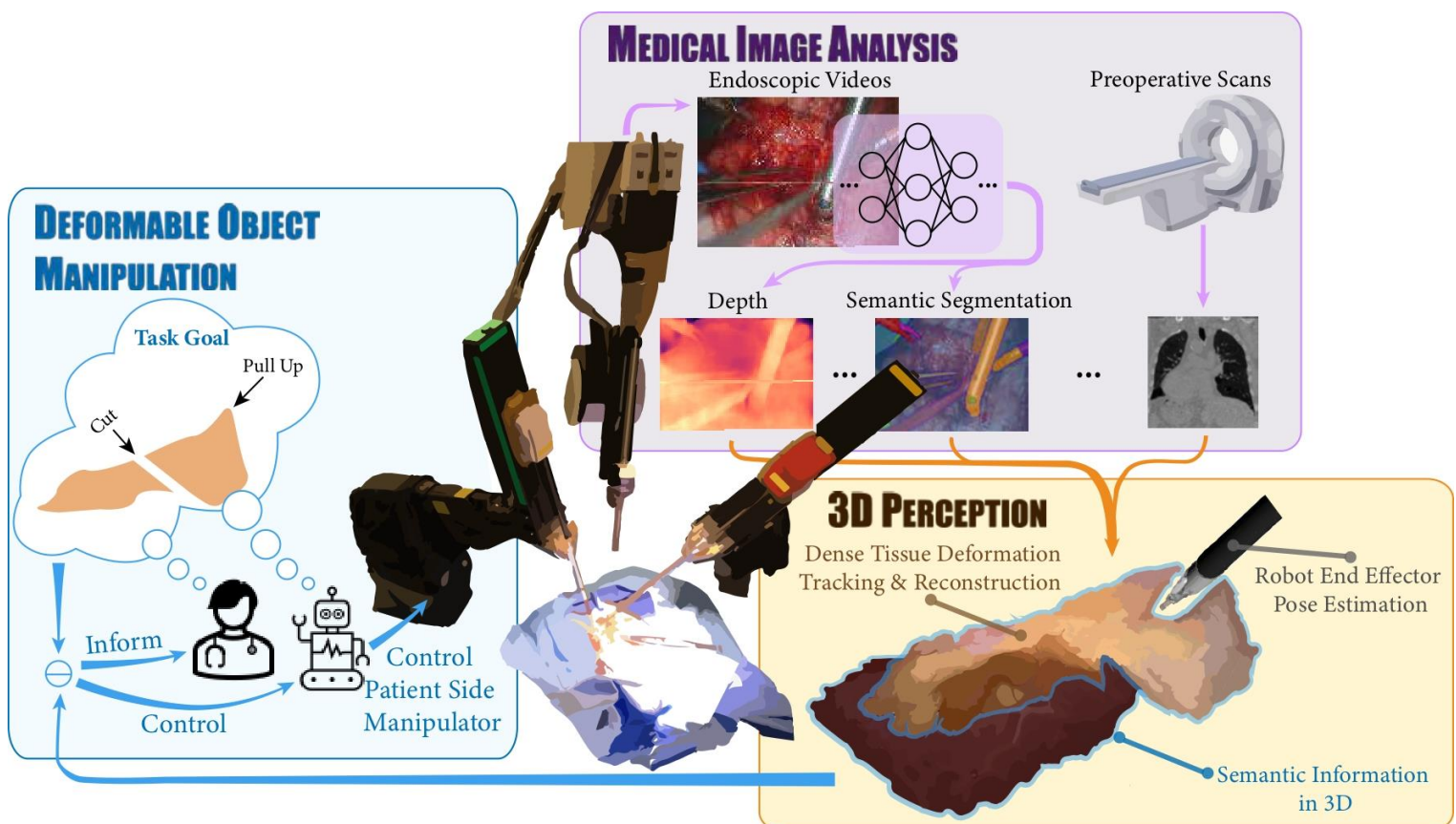
Her mentor Michael Yip will also speak at the workshop, which is organized by Giovanni Pittiglio, Yash Chitalia and others.

What follows is Shan’s Research Statement, which won her one of the two winning spots, the other being claimed by Alaa Eldin Abdelaal.

by Shan Lin

Robotic surgery has been revolutionizing the field of medicine, offering benefits such as faster recovery for patients and improved accessibility to timely treatment, especially in areas with limited medical resources like rural regions. However, achieving greater autonomy in robotic surgery requires a profound understanding and real-time tracking of surgical scenes, which remains an unresolved challenge due to the intricacies of the surgical environment. The surgical scenes present numerous obstacles, including blood, liquids, varying lighting conditions, and deformations. These complexities pose unique challenges for learning-based image or video analysis algorithms, as well as SLAM algorithms that were primarily designed for indoor dynamic environments and autonomous driving. My research, therefore, focuses on tackling these challenges in the domains of robotic perception and manipulation.

One primary focus of my current research for my postdoc at UCSD is non-rigid registration and 3D reconstruction of surgical scenes, with an emphasis on handling challenging conditions such as large deformations and achieving accurate performance. Endoscopic videos are an important type of sensory data that I am working with. They provide real-time information of the surgery and are widely captured in modern surgical procedures, thereby eliminating the need for substantial surgical workflow modifications. We have developed a comprehensive surgical perception framework, Semantic-SuPer, which integrates geometric and semantic information extracted from endoscopic videos to achieve more accurate data association and lead to robust tissue deformation tracking and reconstruction. Currently, we are working on further improving the capability of this framework to handle larger deformations, as well as enhance robustness to noisy input data and reduce error accumulation during longer manipulations. In addition to endoscopes, I am also investigating non-rigid registration techniques for other sensor sources. I mentored a project focused on developing a recursive registration network to track respiratory motions in lung 4DCTs, *i.e.*, sequences of 3D CT scans. Furthermore, I am extending the endoscopic video-based registration and reconstruction results to guide the manipulation of deformable objects. The proposed approach could potentially serve as compensation or an alternative for the methods that require accurate simulations of heterogeneous surgical scenes, which are hard to achieve.



Another aspect of my research involves extracting information from imaging data, including semantic segmentation and depth estimation. Such information plays a vital role in various downstream tasks, such as 3D reconstruction, navigation, and surgical workflow analysis, and the quality of this extracted information could directly impact the performance of these tasks. During my PhD, I focused on semantic segmentation, a procedure of partitioning an image into multiple distinct regions representing different objects or classes. I developed algorithms capable of accurately segmenting endoscopic images using sparsely annotated data, by leveraging temporal information, increasing the training set with synthetic images, and enhancing feature representations. Moreover, I demonstrated how segmentation can benefit one of its downstream tasks, objective surgical skill assessment. Presently, I am extending my exploration to depth estimation and working on leveraging cross-modality information to enhance depth estimation along with semantic segmentation.

While full automation of surgery is still far away, achieving smoother collaboration between surgeons and robots stands as the first milestone. My research will continue advancing perception algorithms and integrating multi-modal sensory data to create real-time, accurate "navigation maps" for surgery, empowering surgeons and robotic systems with a holistic understanding of the surgical scene, thereby facilitating informed and precise decision-making processes. Additionally, I will further integrate my perception results with tasks such as path planning, control, and manipulation, aiming to attain higher levels of autonomy and enhance treatment outcomes in surgery.



Shan Lin

University of California San
Diego

The other **Pioneer of Medical Robotics** selected by the workshop.



Alaa Eldin Abdelaal

Stanford University

See You - for real - in Paris!

49

Computer Vision News

Hey, you! Yes, you! Come to ICCV 2023 this week in Paris! Please! It will be glorious!

October 2 - 6, 2023



GUEST

Simone Crivellaro, MD
UI Health Chicago, Vice Chair
& Chief of Urology Robotic Section

Check out
the video!

FIRESIDE CHAT THE POWER OF AI IN ROBOTIC SURGERIES

Single Port, Ultrasound, and much more



Did you enjoy this October issue
of Computer Vision News?

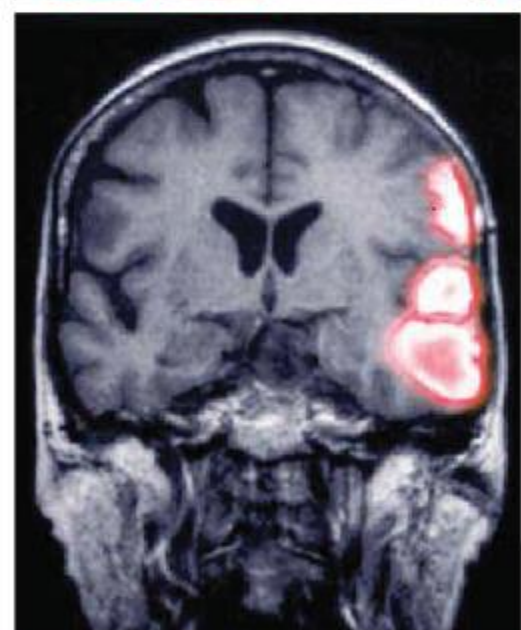
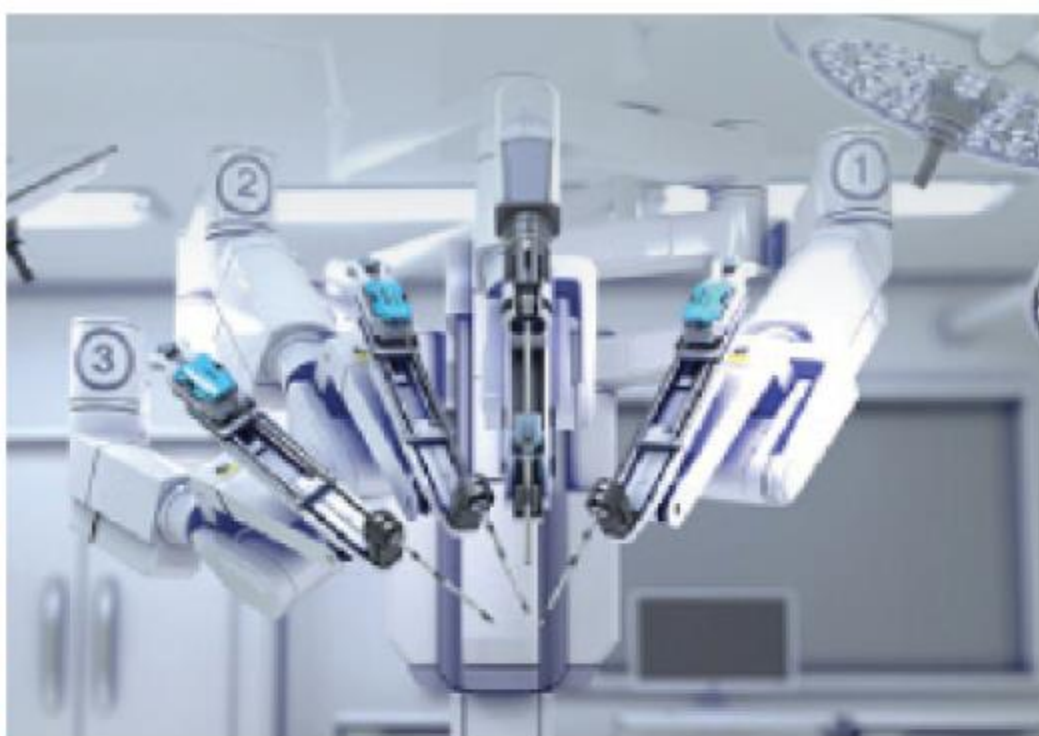
We are glad that you did!

We have one more important
community message to tell you.

It's an advert for something free 😊

Just turn the page,
and you'll know.

Keep in touch!



IMPROVE YOUR VISION WITH Computer Vision News

SUBSCRIBE

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

