FEBRUARY 2023

Computer Vision News

& Medical Imaging News

The Magazine of the Algorithm Community

Wolfram Scheible

This photo was taken in peaceful, lovely and brave Odessa, Ukraine.

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editors:
Marica Muffoletto
Ioannis Valasakis

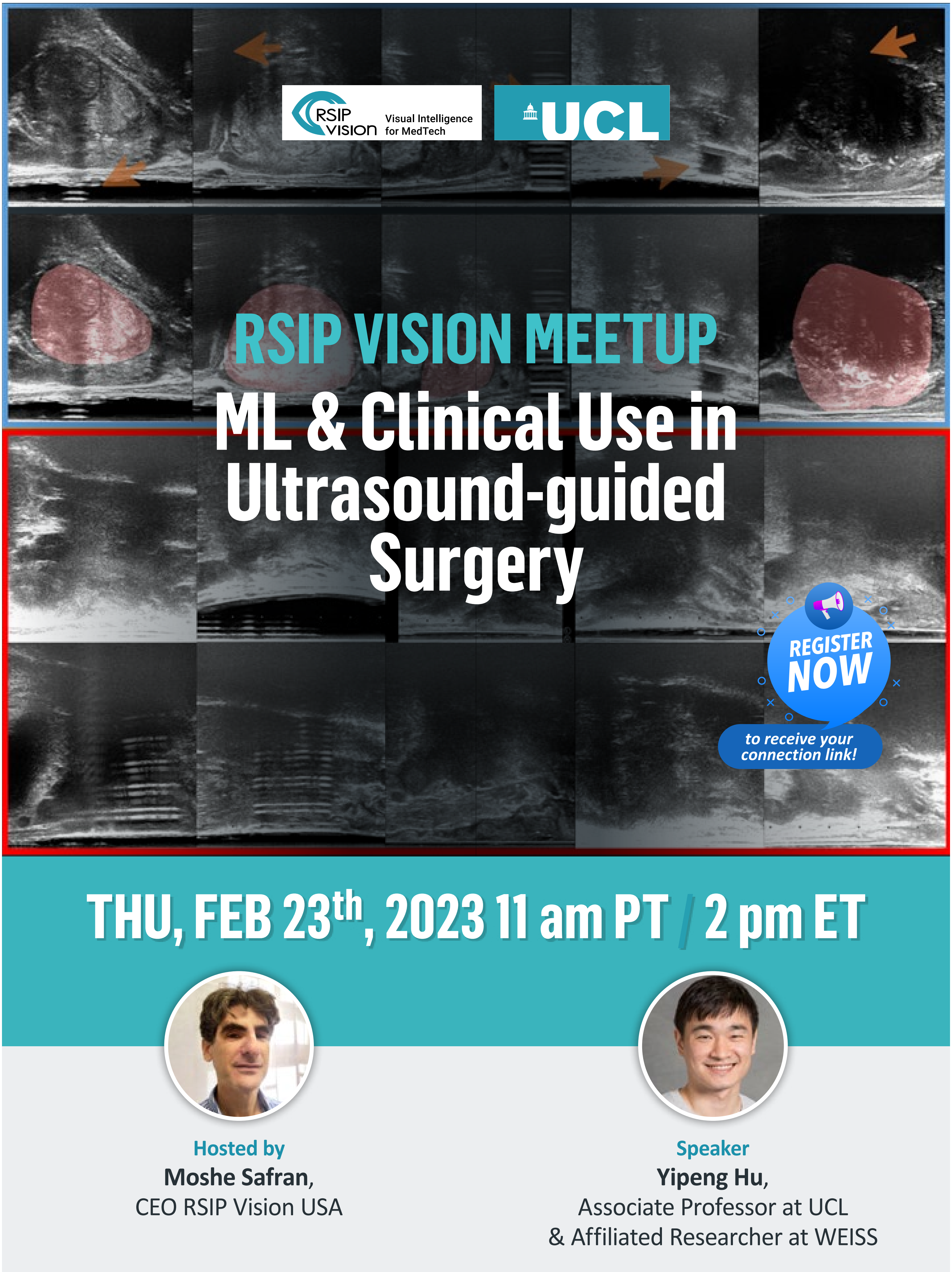
Publisher:
RSIP Vision

Contact us

Free subscription

Read Previous Magazines

Follow Us


RSIP VISION Visual Intelligence for MedTech
UCL
RSIP VISION MEETUP
ML & Clinical Use in Ultrasound-guided Surgery
REGISTER NOW
to receive your connection link!
THU, FEB 23th, 2023 11 am PT / 2 pm ET
Hosted by
Moshe Safran,
CEO RSIP Vision USA
Speaker
Yipeng Hu,
Associate Professor at UCL
& Affiliated Researcher at WEISS

TAME: ATTENTION MECHANISM BASED FEATURE FUSION FOR GENERATING EXPLANATION MAPS OF CONVOLUTIONAL NEURAL NETWORKS

Vasileios Mezaris is a Research Director and Head of the Intelligent Digital Transformation Laboratory (IDT Lab) at CERTH/ITI, where Nikolaos Gkalelis is a senior postdoctoral researcher specializing in deep learning for multimedia, video, and image analysis. Together with first author Mariano Ntrougkas, they recently won the Best Paper Award at the International Symposium on Multimedia (ISM 2022). Vasileios and Nikolaos are here to talk to us about their award-winning work.

TAME is a CAM-based method for **explaining the decisions of a convolutional neural network classifier**. It builds on previous explanation methods for classifiers in a relatively new and underexplored field, where there's been plenty of focus on improving the accuracy of classifiers but less on explaining why they reach their decisions.

"Explanation methods are extremely important for real-world applications," Nikolaos tells us. *"Until now, classifiers have been black boxes. The user sees the result of the classifier and can't understand why it has taken a decision. Usually, they don't know about deep learning and are not computer scientists. They just need an explanation for* ***why the classifier has correctly recognized or failed to***

Vasileios Mezaris

Nikolaos Gkalelis

Mariano Ntrougkas

recognize something."

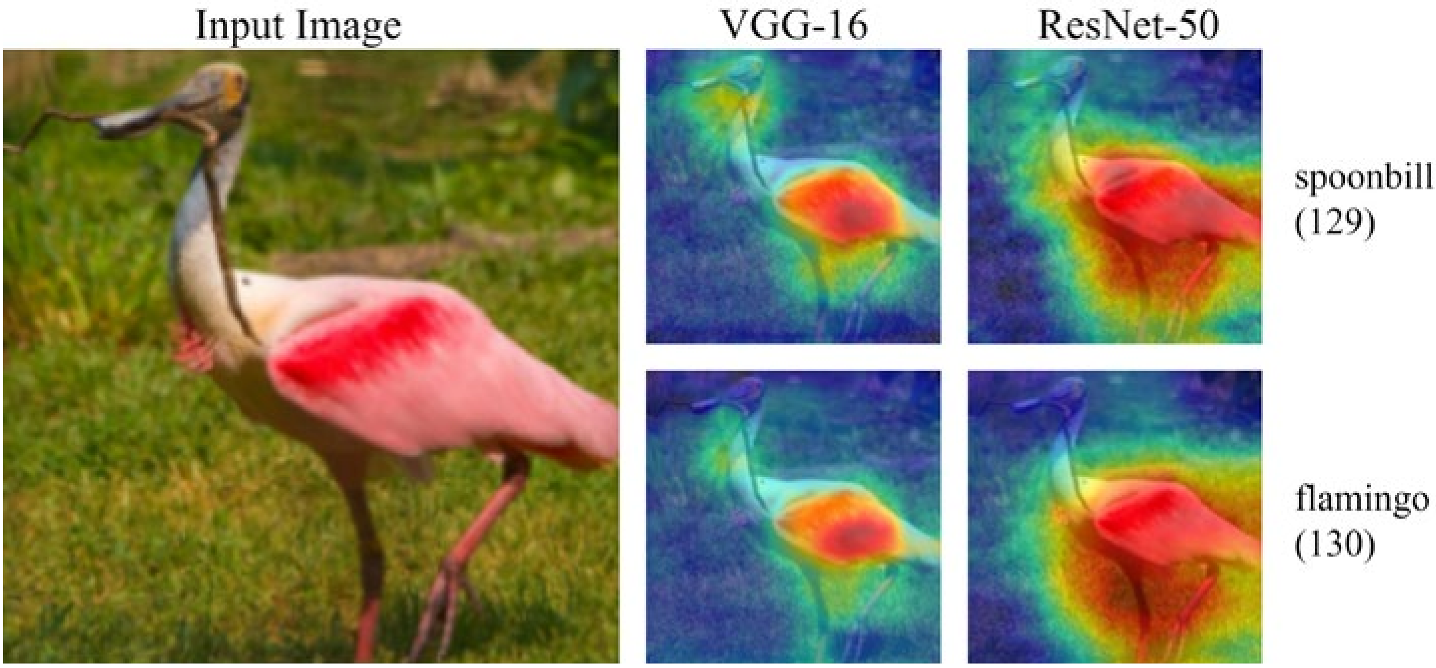
TAME uses computed weights of the feature maps from convolutional layers to derive an explanation mask and is unique in that it uses a training set to learn the optimal weight. Previously, methods used only one image to compute these weights at the inference stage, whereas TAME extracts knowledge from the abundance of training data that exists in the literature. After training, explanation maps can be computed with just a single pass.

TAME's two main novelties are how it formulates the loss function for performing the training and combines feature maps from different layers of the original DCNN classifier to generate an explanation map.

"We introduced an attention mechanism for learning from these feature maps," Vasileios explains. *"The way that's used both at the training and inference stages is unique. It's structured like an attention mechanism but isn't used the same way as a typical attention mechanism in the network, where it would usually go between network layers to be accessed at every forward pass. Instead, it is a branch to the original network that learns some form of weights to modify the original image at the training stage."*

A challenge TAME encountered was that the masks used a cross-entropy loss to learn the additional attention layer, but this loss was deriving wide masks that covered most of the area of the image to keep the confidence score high. The team needed to restrict the explanation mask produced to look at only the necessary pixels of the image.

To solve this, they extended the loss with two additional parts: **a variation loss** to penalize the fragmentation of the image and focus on the most important small parts and **an area loss** to activate smaller



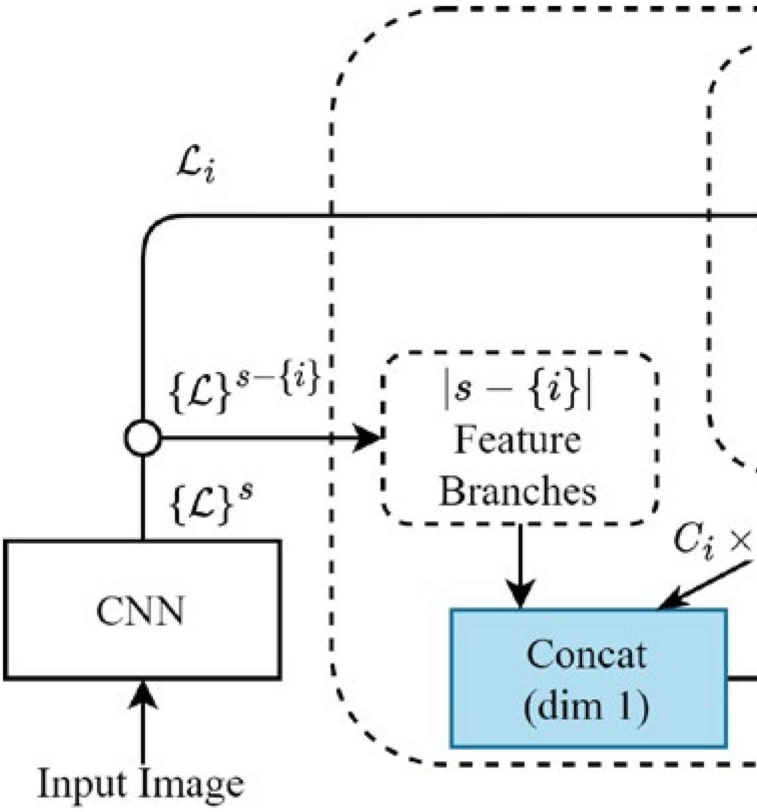
regions and penalize larger regions.

Another challenge the team faced was **taking information from multiple convolutional layers**.

"We added more attention branches to our initial method to process feature maps from not only the last layer but the three last layers to take information from more layers of the frozen network," Nikolaos explains. *"To do this, we had to investigate the different components of the feature branches, and we did ablation studies to see how to construct these feature branches in the attention module. We investigated using **ReLU or sigmoid activation**, which has exploding gradient problems meaning it can't process the feature maps so well. We also investigated different combinations of the feature branches to optimize them."*

CERTH is a multidisciplinary public research center in Greece. It has five institutes, including the Information Technologies Institute (ITI), where the IDT Lab is based. This research was supported by an EU Horizon 2020 project called CRiTERIA.

For Vasileios, Nikolaos, and Mariano, who also works in the IDT Lab and will graduate soon, taking home the **Best Paper Award**



is a fantastic achievement and no mean feat. What do they think made their paper stand out to the judges?

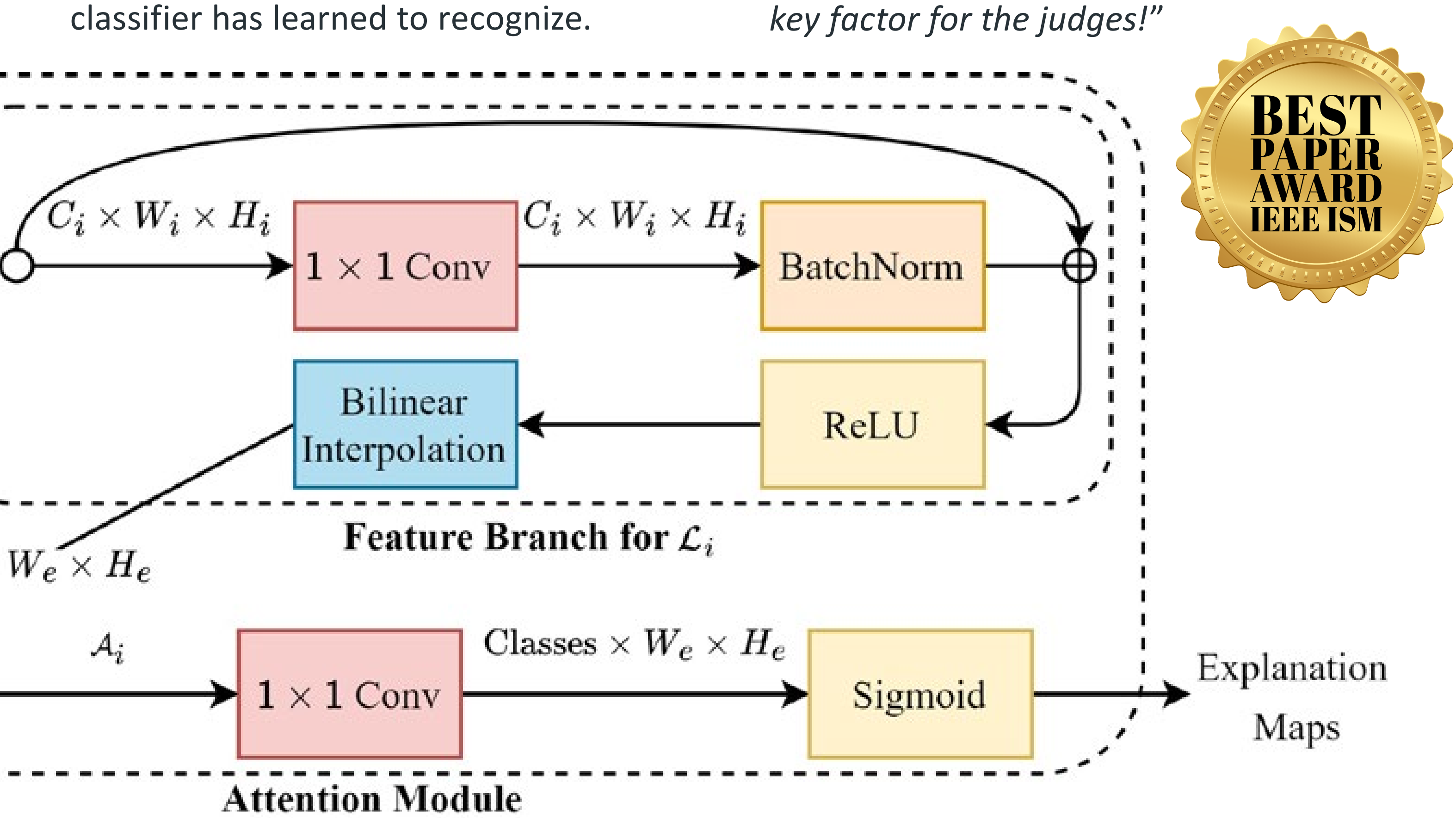
"This is something you can only guess, right?" Vasileios ponders. *"We presented* ***a well-performing method that generates good explanations****, and we do it efficiently with a single forward pass of the network. Our main competitors have the disadvantage of doing hundreds or thousands of forward passes through the network to generate the explanation for a single input image. That is how perturbation-based methods work. Our approach could do as well or even better."*

The team also showed the importance of extracting explanations in practice, with examples of how they can help data scientists to understand what the classifier has learned to recognize.

"For instance, we trained a classifier to recognize the concept of sunglasses, but by looking at the explanation, the classifier had actually learned to understand the concept of a person wearing sunglasses, not the sunglasses themselves," Vasileios reveals. *"These differences in what we expect the classifier to learn and what it has learned could be due to biases in the training data, demonstrating the practical importance of this approach."*

Nikolaos adds:

"Another thing is that we made an extensive ablation study to ablate all the different parts of the method, which convinces the reader and helps them to understand the value of each small part of the method. I imagine this was also a key factor for the judges!"



Katherine J. Kuchenbecker is a director at the Max Planck Institute for Intelligent Systems in Stuttgart, Germany. She is also an honorary professor at the University of Stuttgart as well as the spokesperson for the IMPRS-IS doctoral program.

Over 100 inspiring interviews with successful Women in Computer Vision in our archive!

Research is a team sport!

Wolfram Scheible

Katherine, you're not from Stuttgart, right?

No. Although my last name sounds German, I am American. I grew up in Los Angeles, California, and I had German ancestors. But I learned to speak German when I moved here to Stuttgart.

We would love to hear more about your work. What is Haptic Intelligence?

My main job is as a researcher, to lead a research team. I named my department Haptic Intelligence to try to capture an overarching theme. Most of the projects in our lab have to do with the sense of touch, with haptics, and also have to do with trying to accomplish a task or provide some useful technology. We work on two sides of the coin. On one side, which is more where I started in my research, our work is about providing haptic feedback to a human user, maybe in virtual reality or an operator controlling a robot. What kind of sensation should we measure or generate for them to feel and have a sense of touch to complement what they see and hear? And then, the other half of it, I started getting into when I was previously a professor at the University of Pennsylvania. It's here that I was in the GRASP Laboratory. I was the only one really working on human feedback, and everyone else was working on autonomous robots, computer vision, and artificial intelligence. It was in that environment that I saw the same kinds of signals I had been thinking I wanted to let a human feel them. Those are the exact same cues that a robot should be able to feel, to measure if it needs to manipulate objects in the world or physically interact with humans or the environment. So now, a lot of my research is also on tactile sensing, signal processing, sometimes for robots with machine learning to give them a

Axel Griesch

sense of touch, and then how to use those capabilities for manipulation and human-robot interaction.

Where do you stand now?

Good question! My mother is also a researcher, and she always said the conclusion to all research is: more research is needed. As researchers, we are expanding the frontier of knowledge and learning more things. When all is going well, that always opens up more questions. Compared to other fields, such as vision, I think we understand the sense of touch less than we understand visual feedback, the fields of graphics and computer vision, or even audio capture and processing, or audio synthesis. Touch is still at the earlier stages in terms of how well we understand it scientifically for capture and synthesis and even just fundamentally. I believe that's because touch is distributed throughout our body, and so it's actually a myriad. It's several different sensations, not just one. It's not localized, like in our eyes and our ears. It's very, very interactive. I know vision is also active. You move your head around to get around obstacles. But touch, as soon as you reach out and touch something, you change it. You're exerting forces, so you cannot perceive it without changing it. The motions we make in the world as a human or a robot are the questions we're asking physically. The physics of our own body and of the object we touch dictate the response. It's just very complicated. But that's good for me as a researcher! There are still many things we can investigate and study.

A researcher said some time ago that we had passed the point of haptics being a niche area. What do you think of that?

I agree. When I started my PhD in 2002, haptics was growing, but still very small. It was captivating to me. I was an athlete, and I played volleyball at Stanford at a very high level. I also was a dancer. I am still a dancer! I've always been drawn to technology, but

Alejandro Posada

I got my PhD because I wanted to teach. I did not realize at the time how much I would love research!

also people and these interaction sensors and actuators creating mechatronic systems. I could see that there was good momentum; it was feasible. There were several labs across the country and in the world working in haptics. It was part of the overall robotics domain. But you needed such specialized knowledge. There were not very many commercial devices that you could buy. There was nothing like a dataset or any standards. Now we're moving past that, and I see more and more people in robotics, in human-computer interaction, starting to think about haptics and getting excited about it. One other challenge I want to mention for the diffusion of haptics research is that I think compared to computer vision or audio stimuli, it's harder to share the experience of a haptic interface. At our conferences, researchers not only present papers and posters, but we also present demonstrations, hands-on demonstrations where you bring your system. We brought our hugging robot HuggyBot to Hamburg, where it hugged more than 100 people over three days, so that colleagues from around the world could feel what it's like to hug a robot. Because if you just read in a paper what it's like to touch this thing in virtual reality or hug a robot, it's really hard to understand. I think it's harder to transfer that digitally. It's almost like the very problem we're trying to solve. It makes it not as easy for other people to understand what we're working on or what works well. It's harder to evaluate and quantify the better techniques. But, yes, I agree, haptics is not a niche area anymore. It's really growing.

Actually, it's normal that you agree because that researcher was you! Five or six years ago, you said that in an awesome Futurism video. What progress has been made since then?

Thank you! Actually, I had forgotten about that video. I recorded it only shortly after I moved to the Max Planck Society here in Stuttgart. I enjoyed doing that interview, and it was a good outlook. Since then, I've established my department. I've been here now for six years. I have about 40 people working with me, and I've graduated three PhD students. I think, like, six or seven of my postdocs are now faculty members, and some of the others are on the faculty job market. That's also an indication of universities getting excited about hiring people with this kind of expertise, seeing the promise of interactive touch technology and touch capabilities for robots. In those intervening years, we have seen surface haptic... touch feedback on flat screens is

Wolfram Scheible

The conclusion to all research is: more research is needed!

really growing. It is still not very present in commercial products. But Microsoft announced maybe a year or so ago that they want to add haptics and stylus-based interactions. Also, in virtual reality, I had a start-up for a while. It's now closed. We were too early on giving haptic feedback in virtual reality, but at the time, it wasn't even stabilized on how head-mounted displays would work, what would be the operating standards, and how can this technology work. Now we're starting to see many companies, also research labs, coming up with more feasible approaches for haptic touch feedback that adds to the interaction in virtual reality. Then the third thing I'll highlight is that touch sensing is also a huge gap. Most robots that you can buy have almost no sense of touch. They have proprioception, like their own joint angle sensors. Like a robot knows where it is in space if it's calibrated. It often has good cameras and can perceive things at a distance. But there's almost nothing like our skin. We know that if you impair the human sense of touch, if you make your skin numb, humans become much worse at manipulating things. It's much more difficult to brush your hair or do any activities of daily living. Your sense of touch is constantly giving you feedback about your body and your manipulation of the world. We're now starting to see tactile sensing go towards where it could be integrated into a robot. But again, it's complicated how to get it integrated and not and be very lightweight and robust and cheap. There are a lot of technical challenges, and then what can a robot do with that information? We're still in the very early days. It's like when the first cameras were being made. We have some of the first tactile sensors, but it's growing. At every robotics conference that I go to, there are workshops on tactile sensing, and so there's a growing robust community there as well.

Another application area I work on a lot is robotic surgery. My PhD advisor helped

Shari Kuchenbecker

invent the Da Vinci robot, and it gives the surgeon a very good 3D view. They move the instruments around, but they don't feel anything. And so, again, they're not directly touching the patient. They're sterile, sanitary, not being contaminated, but they cannot feel that reaction. That makes manipulation more difficult because we are used to reaching out. In your car, for example, you want to adjust the radio volume. If it's a physical knob, you can find it without taking your eyes off the road. You can keep watching and approximately reach out, recognize the physical features and manipulate it. This is actually a big problem in the auto industry. If it's a graphical display with digital controls, I need to avert my eyes and look at it to understand exactly where and how to manipulate it. Car companies are very interested in adding haptics, even to screen-based interaction.

My first PhD student here in Germany, she and I were making a hugging robot at the time when the pandemic hit. And it ended up being much more relevant. We were able to recruit many people to join our studies about this HuggyBot robot who hadn't hugged anyone in months because they had been isolated in quarantine. People really liked hugging our robot! It's soft, warm, and responsive. We also sterilized it. We completely washed its outfit and sterilized it between people, and they wore a mask. We had the robot wear a mask also, so they would feel safer.

In Italy, we have a game called the Game of the Tower. I don't know if it exists in other countries. You have two people, and you have to throw one of them from the top of the tower. If I asked you, you have Katherine, the researcher, and Katherine, the teacher. Which one do you throw?

Well, I already did this because I was a professor at the University of Pennsylvania, where my job description was both teaching and research. I taught Introduction to Robotics to about 170 students, senior undergraduates, master students, and a few PhD students. Every fall and every spring, I taught about 100 undergraduates

dynamics with lots of programming assignments. I also was running my research lab. And then, I was lucky enough to get the opportunity to become a director here in my institute to move to Germany to have a new adventure. This job is almost solely about research and also about leadership, mentoring, running a doctoral program, and running my institute. But I don't teach classes here. I might get to a little bit in the future at the university, local university. Of course, I teach my group. I teach my team. But it's not as much a headline as it was. That is very surprising to me because I got my PhD because I wanted to teach. I did not realize at the time how much I would love research.

Since you mentioned mentoring, I have to appeal to a third Katherine. Many young scholars are asking themselves now which road is better for them. Maybe we can use your experience to tell them.

Actually, I just gave a talk about this on Tuesday. Yesterday, today, and tomorrow, we have 140 amazing applicants virtually interviewing for positions in this doctoral program that I run. The advice that I gave them and that I would share with you is; first, the most important choice is *who* you work with. You are signing up to apprentice with a scientist to learn how to think like they think and work on what they work on. And to be open and meet several people as it's a matching process. It's not like getting admitted to a Bachelor's program or a Master's program where it's just on academic excellence. There has to be a synergy. You're finding your academic parent, and there will be hard times along the way. In PhD and in research, not everything works. We tell wonderful stories of everything that works. But behind every paper, every demonstration, there are many, many things that didn't work, that were frustrating, ideas we had that didn't function as we thought, problems, bugs in our code, wrong assumptions, confusion, and miscommunications. Research is a team sport, and there are things we can do to make it go smoother, but there will be difficult times. Always remember, what is your motivation? I find the researchers who succeed most are really curious and excited and love to learn, and are willing to read.

You really need to understand every piece of the system that you're working on. The pieces in the system interact. This big dataset that you downloaded or are using or this code that you're borrowing, you put them together. The way you're plotting the data, the way you're telling the story, all of the things have to work together. You really need to understand each piece if you want to make progress. I think it was Max Planck who said *"insight must precede application."*

Axel Griesch

You have to really understand stuff and the tools before you can apply them to make them do what you want. Otherwise, they will misbehave.

When was the last time that you, as successful and experienced as you are, thought, "*I can't do this*"?

Honestly, I've never really felt like I should give up. I'm extremely lucky that I love what I do, and I have wonderful conditions and great people and other people supporting me. But just this morning, I was feeling quite discouraged when I got up and was sitting at breakfast with my husband. I just was feeling kind of grouchy because I had stayed up really late last night editing a journal article that the revision is due on Sunday. My student had worked very hard, and it's been such a busy week that until last night, I didn't have time. But I was very proud that I finished editing this whole paper. It's 15 and a half pages long. It's years of work. Then I had to watch the talks for the candidates that I was interviewing today. I was so tired, I was, like, falling asleep during their talks. It was pre-recorded talks. I just felt so frustrated that I didn't have enough time to do all the things that I needed to do. I had to go to bed at some point and then wake up. I still had really good interviews with people. My husband cheered me up a little bit. Also, bicycling to work and eating a good breakfast. Yeah, there are ups and downs. Sometimes you feel like it's too much. I just say, step by step, do the next thing in front of you. We can figure out what is the best that we can do, whatever is right in front of us. I just try to do the best I can on that. Then say, okay, what is the next thing? And through this, you can accomplish a lot.

Actually, what you are doing now is just like your husband: you are cheering us on


Axel Griesch

The most important choice is who you work with!

by saying that things are going to work out. Sometimes it doesn't work. How can a student recognize that this is not working and it's time to change the path?

You have to talk with the other people you're working with. It's very rare to do research completely alone. At least you should have a mentor. Maybe also have some team members who are on the project. Stay in frequent contact with them and check in with them. Even if you can also talk to people who aren't on your project team, your partner, your friend, your mom, or your cat. Describe the problems that you're having. Just putting your problems or your challenges in a frame to communicate to them, to someone else something else often gives you a new perspective to solve your own problem. Research is a team sport. Talk to the other people in your lab, in your group, in your program, or old friends. Don't stay alone and isolate. Listen to what other people tell you. I tell my students, I'll give you a lot of advice, and hopefully, four out of five of the things I tell you are right. But sometimes, I will be wrong. You need to get other opinions. You need to check, and you need to challenge me. Read some papers, come back to me and say, you know, I know you think it should work, but I don't think it's going to work because of this and this and this. And then we can really engage about that and make a decision together.

It's probably good to have more than one project going at a time. That way, one is always going better than the other. Sometimes pausing, taking some time away from one particular endeavor, lets you make some progress and feel good about something else. Even if that's something like updating your CV or making the video for your paper or redoing a plot, or, I don't know, doing your laundry. Then you can come back to the hard thing, and your brain keeps working on these things in the interim, and you might have new ideas. I also personally like getting outdoors and doing some exercise brings positive energy and gives me a fresh perspective. Also, talking to people, my friends, and my family.

Your robots... [*laughs*]

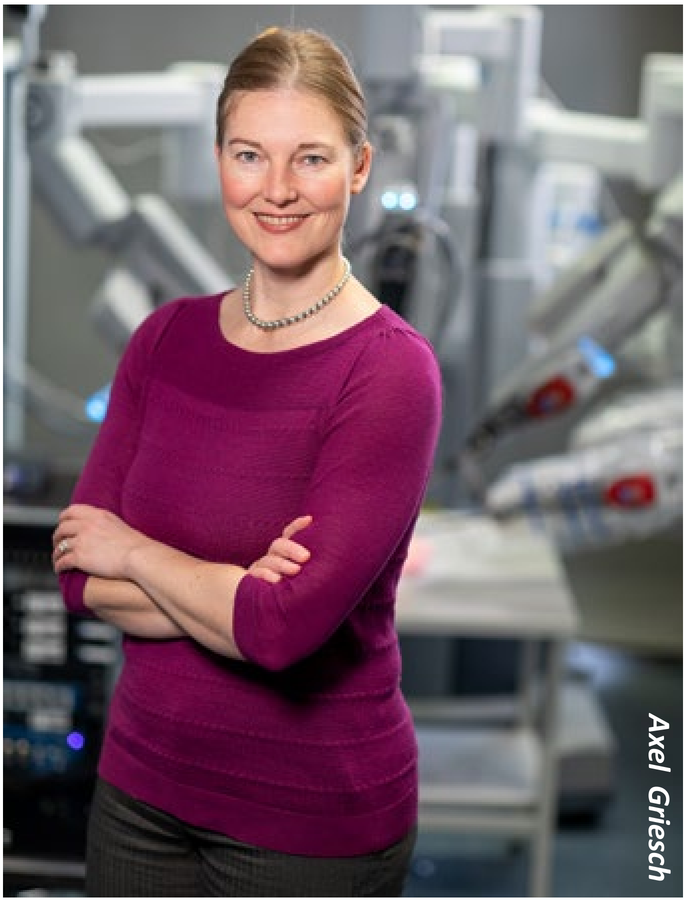
Yeah, maybe I'll go down the hall and get a hug from HuggyBot. [*laughs*] No, I find the people in my life are much more meaningful than the robots. Though I do like the robots, too!

One message for the community?

My first impulse on how to respond was to point to the sign that's behind me. It says, *"It always seems impossible until it's done!"* I bought this when I was on a vacation trip. It just resonated with me how things often feel in research, that we work on something for so long, for years. We get it to work. We submit our paper, and it gets rejected the first time. This is normal. Reviewers see things we didn't see. We are actually a global scientific community helping each other solve problems and explain our work and figure things out. If you keep persisting, eventually, it often does pay off. Not always. And then all of a sudden, it's done, and you have your paper accepted, or you have the job you wanted, or you finished your PhD. You figured it out.

Or you get your grant!

Or you get your grant! It felt like a miracle when I got my first grant. I remember it felt like the skies opened up and were shown upon me. Until that moment, I wasn't sure



Axel Griesch

"Schritt für Schritt", step by step...

I would be able to succeed as an academic. You need funding for your work, and that means you need to have good enough ideas and explain them to other really smart people who will give you money. That is indeed a big responsibility to take taxpayer money or a company's money, or a foundation's money and try to create new knowledge. But I believe it's worth it. You have to have patience and persistence and a positive attitude, and work hard. For that, I think, again, teamwork can really help you get to the end. So even though it might feel impossible, it's achievable. The Germans say *"Schritt für Schritt"*, step by step, and then you get there.

Over 100 inspiring interviews with successful Women in Science!

Congrats, Doctor Larissa!

Larissa Triess has recently finished her PhD at the Karlsruhe Institute of Technology in cooperation with the Mercedes-Benz AG. Her research focused on domain adaptation and 3D LiDAR data generation to improve perception algorithms for autonomous driving. Larissa is now a Machine Learning Engineer in the autonomous driving department at Mercedes where she continues her work on 3D scene understanding. Congrats, Doctor Larissa!

Autonomous vehicles require a detailed understanding of their environment. Therefore, vehicles are equipped with a variety of sensors, such as cameras, LiDARs, and RADARs. They capture the surrounding and real-time neural networks then predict the location and type of objects in the scene, such as vehicles, persons, traffic lights [Triess2020IV] [Triess2021NeurIPS].

The training of the neural networks is usually done with large amounts of annotated data. Ideally, it covers the application domain as good as possible. However, there are various effects that can cause domain shifts. Figure 1 shows a selection of such domain changes that lead to impaired performance in perception models [Triess2021IV]. It is not feasible to acquire data and annotations for each domain, especially if it involves safety-

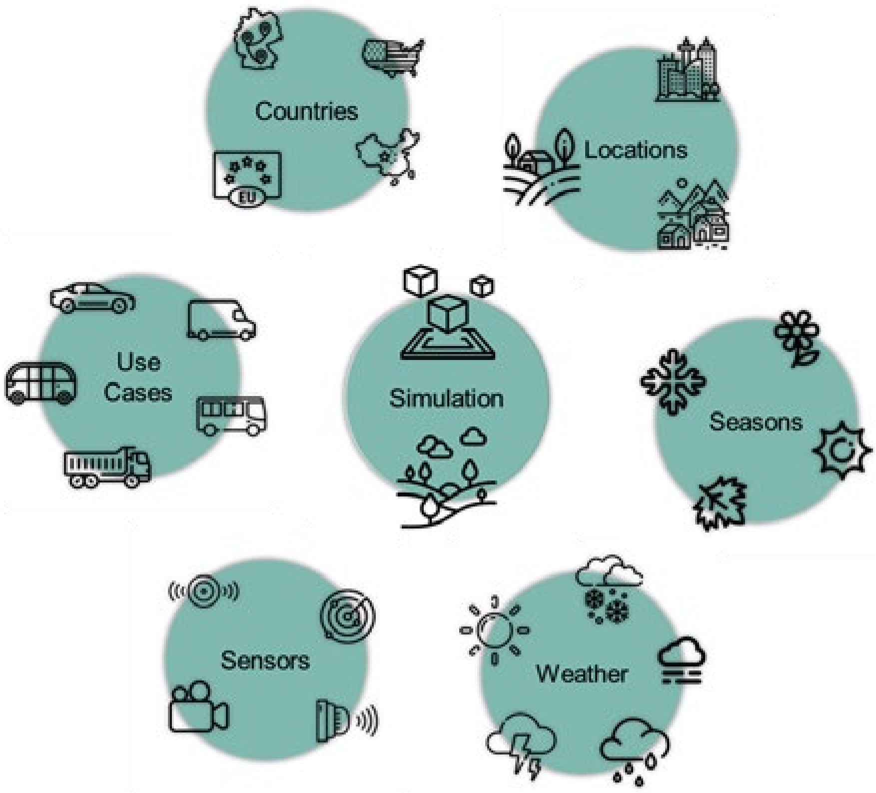


Figure 1: Examples of domain shifts for autonomous vehicles.

critical scenarios. Therefore, Larissa investigated three different approaches on how to achieve more robust performance in the application domain without any additional annotation effort:

1) It is common to insert additional objects into the scenes at training time. This is especially helpful for detecting under-represented dynamic objects. Larissa's work focused on the generation of such 3D objects [**Triess2022AISTATS**].

2) Domain mapping is a type of domain adaptation where data from a source domain is transformed into a target domain. This is achieved with generative models that synthesize realistic high-resolution LiDAR data [**Triess2019IV**]. A perception model can then be trained on the transformed data and be applied in the application domain without being exposed to a domain shift. In her work, Larissa proposes a quantitative metric to judge the quality of the transformed LiDAR data [**TriessGCPR2022**]. Figure 2 shows the concept and example output of the metric that provides local realism scores for the point clouds. In a detailed analysis, Larissa shows how the data quality and the perception performance are correlated [**Triess2022IJCV**].

3) For learning domain-generalized features, Larissa uses a method that simultaneously learns the semantics and the geometry of the scene. Since the geometry can be learned in a self-supervised fashion, it is advisable to learn them jointly for the source and target domains. The domain-generalized features then help to capture the semantics of the target domain.

To summarize, Larissa's dissertation presents a comprehensive framework for domain adaptation and semantic segmentation of LiDAR point clouds in the context of autonomous driving. For updates on her future work, follow Larissa on **LinkedIn**.

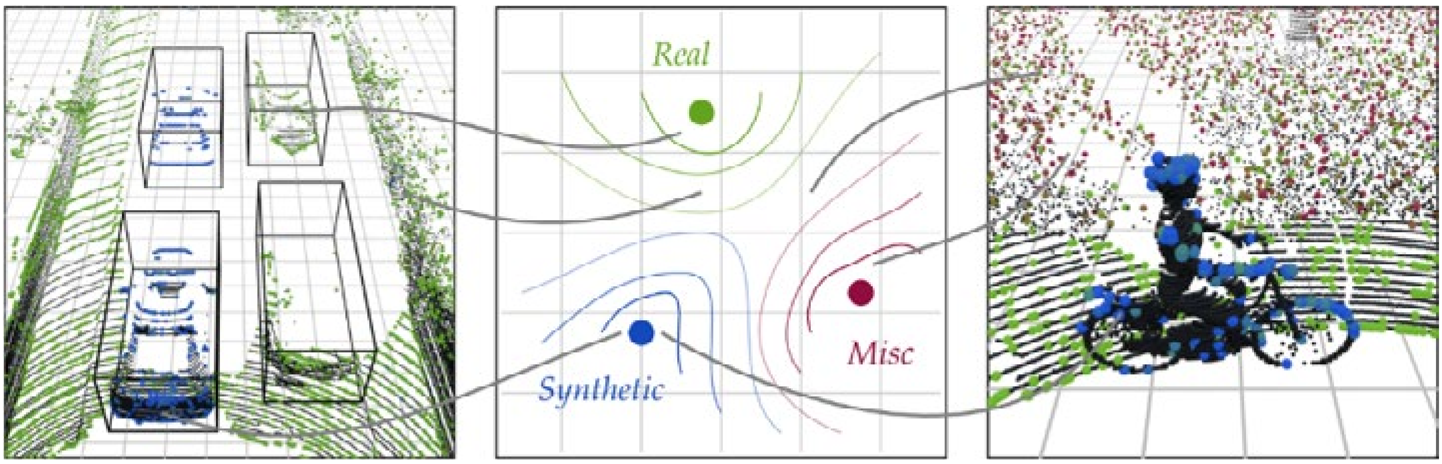


Figure 2: The realism measure has a tripartite understanding of the 3D-world (middle). It estimates the realism of local regions within the point cloud. Both example images are real-world scenes (green) with augmentations from a simulation framework (blue) and point distortions (red).

NLDL - NORTHERN LIGHTS DI

Robert Jenssen is a Professor in the Machine Learning Group at UiT The Arctic University of Norway, and the Director of Visual Intelligence, a research center with 11 Norwegian partners led by UiT. Michael Kampffmeyer is an Associate Professor, Head of the UiT Machine Learning Group, and a PI at Visual Intelligence. Stine Hansen, a researcher, and Srishti Gautam, a PhD student, are part of the UiT group and VI. All four are organizers of the Northern Lights Deep Learning Conference, which just took place in Norway, and are here to tell us more about it.

The **Northern Lights Deep Learning Conference (NLDL)** is an annual event organized by UiT and the Visual Intelligence research center dedicated to any topic related to deep learning. It began as a workshop in 2018, when there were some small local image conferences in Scandinavia but there was no forum dedicated to deep learning methodology.

"When we first organized the workshop, we already had some relatively big projects running here in our group, including industry and public sector collaboration," Robert, NLDL's General Co-Chair alongside Michael, tells us. *"If anyone was going to organize a workshop in Norway on neural networks and deep learning, we felt it should be us."*

The workshop started small, with only 30-40 participants, but over the years, the team discovered that more and more wanted to attend, so they announced it more broadly.

"We were quite deliberate in inviting

EEP LEARNING CONFERENCE

people outside our group and giving them responsibilities, such as session chairs," Robert continues. *"Once it became a conference in 2020, we had full paper submissions, so we needed lots of reviewers and program chairs too."*

This year's sixth edition had attendees from 18 countries and featured a various topics ranging from **fish segmentation** to **representation learning**. Most authors, almost a third, were from Norway, followed by Denmark, Germany, the UK, Sweden, and the rest of the world.

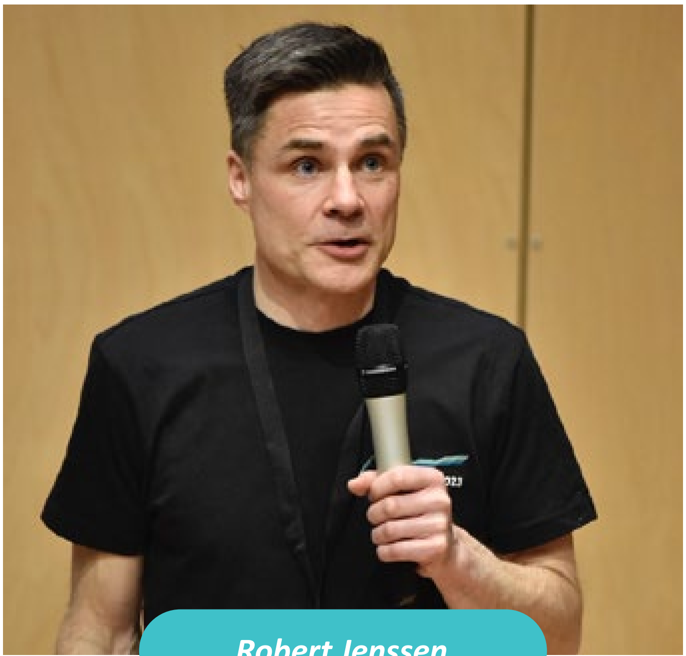
A Winter School ran across the week, with tutorials from prominent researchers on the first and last days. Around 150 early career researchers signed up, the majority from Europe but including some who traveled from as far afield as the United States and Japan.

"We had lectures about using the cluster and putting the code on your server, and it was very hands-on," Srishti tells us. *"That's something missing in many conferences and schools we attend. We don't have many courses on that subject, so I thought it was important. The people I spoke to were really interested in all of the topics. They went home super happy!"*

There was a stellar line-up of keynote speakers, including **Polina Golland** from MIT, **Mihaela van der Schaar** from the University of Cambridge, and **Christian Igel** from the Pioneer Centre for AI at the University of Copenhagen.

"Polina Golland talked about keeping a sharp focus on deep learning for the social good," Robert tells us. *"Like technology applications coming together into something new through deep learning that could have a really high impact in the world. Several of the keynote speakers stressed that."*

Other highlights included a dedicated industry event and Diversity in AI, a



Robert Jenssen

bottom-up initiative from young students and researchers, which Srishti and Stine were involved in organizing.

“We had two amazing male speakers working towards increasing diversity in their institutions and companies,” Srishti recalls. *“What was common in both of their talks was that we should never see diversity as a limitation, but instead, it’s an added competence.”*

Stine adds:

*“We had the **Women in AI** event last year but wanted to be broader this time. We’re planning to hold this event again next year.”*

The conference did not host traditional Best Paper awards. Instead, organizers recognized reviewers’ work by awarding Best Review awards to those who provided high-quality feedback on submissions.

“Much work goes on behind the scenes reviewing all the papers, so we wanted to show some gratitude for that,” Michael explains.

Holding the conference in January is a deliberate decision to coincide with the wintertime in Northern Norway, which can be rough but also beautiful, with its glorious mountains, fjords, and plenty of snow. Social events in the evenings offer the opportunity to witness one of nature’s most spectacular phenomena, the Northern Lights.

“Of course, people want to come here and talk science for a week, but we have to be honest, they also really want the Northern Lights chase!” Robert laughs. *“This year, we went to a small mountain near the city center, Tromsø. You need to take a cable car to get there. In the evening, 250-300 people went up there. They didn’t get the full jackpot, but they got some Northern Lights and even caught it on camera. People come here because it’s maybe a bit out of the ordinary and has this added Northern Lights dimension. The New York Times recently ranked Tromsø in the top 10 list of destinations that you should visit.”*

Michael smiles: *“The main reason being the conference, of course!”*

The organizers hope to have a bigger venue next year, as registration had to be closed

one week before the deadline due to the high demand. As for dream speakers, Robert is happy to report that they have been very fortunate so far, and he is sure that will continue.

"Mihaela van der Schaar and Polina Golland had already said they wanted to attend our conference, and we were very happy to have them this year," he says. *"In the past, we've had people like Bernhard Schölkopf,* **Serge Belongie***, Arthur Gretton, Lars Kai Hansen, and* **Julia Schnabel***. I don't want to offend anybody by suggesting a specific name for 2024, but I have high hopes that we'll continue to attract speakers at this very high level."*

Is there anything else they would like to see next year?

"Very strong Northern Lights!" Michael laughs.

As we wrap up, Robert highlights the importance of the conference for UiT as part of its big push on deep learning and AI, including, through Mihaela and Polina, an increased focus on machine learning and deep learning in the healthcare direction. UiT is part of a network of AI universities and institutions in Norway called NORA - Norwegian Artificial Intelligence Research Consortium, which is also a collaborator for NLDL.

"In Norway, it's not only us here at our university, but it's an effort by a bigger Norwegian community," he reveals. *"That's something we're really grateful for!"*

If you are curious to read about Stine's thesis, see her own feature in the Medical Imaging section of this magazine.

Computer Vision News has found great new stories, written somewhere else by somebody else. We share them with you, adding a short comment. **Enjoy!**

Shutterstock Rolls out a Generative AI Toolkit to Create Images Based on Text Prompts

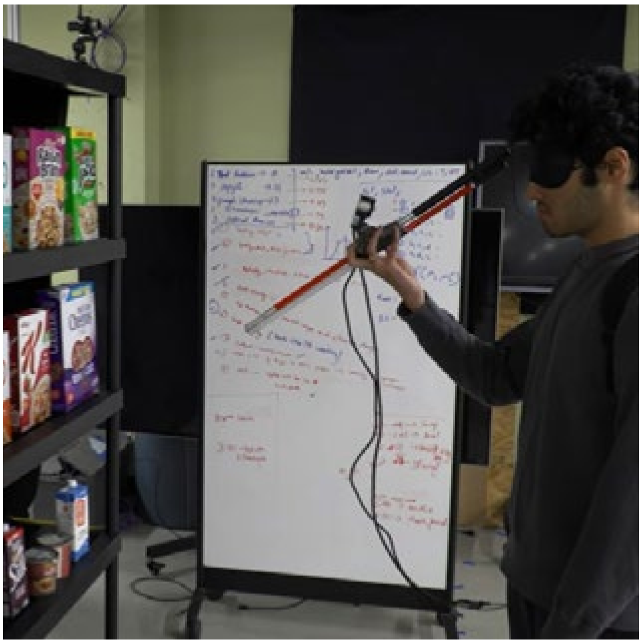
It's going to be interesting in the world of stock images! **Shutterstock** did not lose any time to partner with **OpenAI** to help develop OpenAI's **Dall-E 2** artificial intelligence image-generating platform. And now they bring their own generative AI tools to their users: customers of Shutterstock's Creative Flow online design platform are now able to create images based on text prompts, powered by OpenAI and Dall-E 2. On the other hand, Getty Images seem to stake their bets in the opposite direction: they play it safe, banning all AI-generated content over fears of future copyright claims and legal risks. Read More

How Artificial Intelligence Is Helping us Decode Animal Languages

Curious? You should be. AI and machine learning have been used over years to **analyze and translate human languages** to very impressive state of the art levels, enabling new communications avenues between humans and machines. Ask **Katie Zacarian**, the CEO and co-founder of **Earth Species Project (ESP)**, a non-profit organization that uses AI **to decode animal communication**. The idea is to apply past breakthroughs of AI for human language to animal communication. This is not less impressive! It involves large data sets that contain **visual, oral and physical communications between animals**. Read More

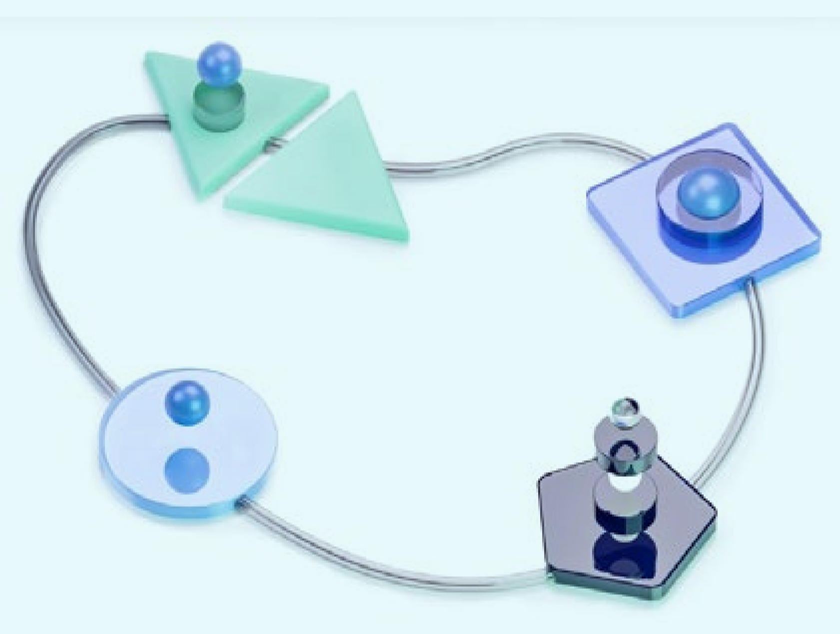
Show Me the Way, AI Stick!

It is not the first time that Spotlight News tells you about AI providing helpful solution to **visually impaired persons:** daily life presents them a variety of challenges that those with normal vision may not be aware of and we are happy to witness great advances in supporting the population affected by visual impairments. Mobility is not the least of these challenges and AI is always one step closer to solving it. This time, a team of researchers from Colorado developed a new kind of **walking stick, that maps and catalogs the world around it** thanks to a camera and computer vision technology. **Watch the Video**



A Robot Was Scheduled to Argue in Court, Then Came the Jail Threats

AI does no always win and not all the stories have a happy end. A British gentleman planned to have a "**robot lawyer**" help a defendant fight a traffic ticket in court. The person challenging a speeding ticket would **wear smart glasses that both record court proceedings and dictate responses** into the defendant's ear from a small speaker. The system relied on the same AI text generators everyone's talking about. To cut a long story short, his idea met a strong opposition from the side of corporations feeling threatened by it. Some of them answered with threats of prosecution and prison time. That's how the story ends this time... Read More

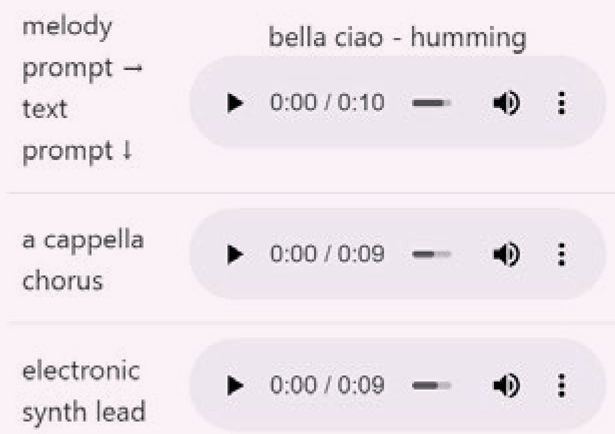


Open-Source Active Learning Toolkit for Computer Vision

What machine learning engineer and data scientist doesn't want to **understand and improve their training data quality and help boost model performance**? That's the promise of an AI-assisted platform provider called **Encord**. They have released a free open-source industry agnostic all-in-one toolkit designed for that purpose. One of the motivations disclosed by the company is that self-driving cars and diagnostic medical models using AI suffer from a "**production gap**": proof-of-concept models perform well in research environments but struggle to make predictions accurately and consistently in real-world scenarios. Read More

MusicLM: Generating Music From Text

Let's end this month's AI Spotlight News with a musical note: why couldn't we use AI to generate the kind of music that best fits our mood and desires? That's what they did at **Google Research**, where a team of researchers is introducing **MusicLM**, a model generating high-fidelity music from text descriptions such as "*a calming violin melody backed by a distorted guitar riff*", why not? Follow the link for examples. They also publicly release MusicCaps, a dataset composed of 5.5k music-text pairs, with rich text descriptions provided by human experts. Read More



COMPUTER VISION EVENTS

World Artificial Intelligence Cannes Festival
9-11 February
Cannes, France & online

RE•WORK
AI Summit West

San Francisco, CA
15-16 February

SPIE Medical Imaging

San Diego,a
California, USA
19-23 February

ICCTech Int. Conf. on Computer Technologies
Kuantan, Malaysia
23-25 February

CMVIT Machine Vision and Information Technology
Xiamen, China
24-26 February

ACM/IEEE Int. Conf. on Human-Roabot Interaction
Stockholm, Sweden
13-16 March

Digital Health Rewired

London, UK
14-15 March

Emerging Medtech Summit

Dana Point, CA
20-23 March

Computer Vision Summit

Tel Aviv, Israel
23 March

ISBI

Cartagena de Indias, Colombia
18-21 April

FREE SUBSCRIPTION
(click here, its free)

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

Fill the Subscription Form
ittakeslessthan1minute!

SUBSCRIBE!

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

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

Due to the pandemic situation, most shows are considering going virtual or to be held at another date. Please check the latest information on their website before making any plans!

MEDICAL IMAGING NEWS

FEBRUARY 2023

AI FOR PARTIAL NEPHRECTOMY

RSIP Neph Announced a Revolutionary Intra-op Solution for Partial Nephrectomy Surgeries. This next generation robotic surgery innovation utilizes AI-based visual intelligence to provide vital information that is overlaid and presented on the video feed in real-time during surgery.

RSIP Neph announced its new innovative intra-op solution that utilizes advanced Computer Vision and Artificial Intelligence algorithms, to enhance **the next generation of robotic surgery**. This solution has been under development with the largest urology department in Israel and it will assist surgeons during the surgery itself, by providing additional important information on lesions and body structures which are hidden to the human eye.

This new solution includes **advanced artificial intelligence algorithms** and a deeper understanding of the existing capabilities of the robot to perform these kinds of procedures, allowing it to radically enhance the way this procedure is currently performed. Segmentation, registration, and tracking capabilities provide the surgeon with a new view of crucial information about obstructed unseen tissues and structures. This view will be overlaid on top of the currently available robotic display, allowing the surgeon to perform this task with minimal risks. CT scans are currently taken prior to the operation, to inform the surgeon about the size, shape and location of tumors and other relevant structures. This pre-op information is merged with the intra-op ultrasound, and then fused to the intra-op stereo cameras. With our new technology, this information is overlaid on top of the surgeon's display in 3D, bringing the look and feel to a completely new dimension (**watch the video in this page**).

Some renal tumors do not require full removal of the organ due to their size and position within the kidney. In such cases, partial removal of the kidney limited to the lesion itself and surrounding areas may be more suitable than total removal of the organ. This partial removal, also known

as **partial nephrectomy**, is preferable to full removal of the organ, as it preserves healthy tissue that can continue to function. This is particularly critical in case of future damage to the second kidney.

Until now, partial kidney removal has been performed with some extent of guessing by the surgeon about the optimal cutting location. The gross estimation is due to the limited view that the surgeon has of the lesion which needs to be removed. This may lead to excess removal of healthy tissue or worse, insufficient removal of the lesion. It also limits this procedure, since it can only be performed by very experienced surgeons.

"This new technology will alleviate the uncertainties that surgeons experience during partial kidney ablations," said **Dr. Arnon Lavi, MD, Senior Urologist at Haemek Hospital in Afula, Israel**.

"Removing the need for guessing will allow the operating team to avoid critical mistakes and lead to better surgical outcomes, with very significant benefits to the patient."

RSIP Neph is part of the **RSIP Vision group** and until now, was active in a stealth mode. It develops innovative visual intelligence solutions for the Urological field, through advanced AI and computer vision applications. The solutions address diverse clinical use cases, including Kidney related procedures, PCNL, Lesion ablation, Calix segmentation (4-phased CT) soft tissue tracking and Non-rigid registration. Its technology is based on its founders' extensive track record and years in the field, trusted by the largest, industry leading medical device companies.

RSIP Neph

Full announcement

A PEAK INTO THE WORLD OF SPECTRAL IMAGING WITH SPECTRAI

By Marica Muffoletto (twitter)

Welcome to new and old readers of Computer Vision News! This month, we review something quite different than usual. In fact, we are going to talk about a domain-specific deep learning tool, named SpectrAI, solely dedicated to the analysis of spectral data through neural networks. SpectrAI includes both command line and graphical user interface (GUI) options. It's fully written in the well-known Pytorch library (Python), but thanks to the Matlab-based GUI, it is also well-suited to deep learning newbies who can select models and hyperparameters much easier than through code.

But what kind of data does SpectrAI analyse?

Not really the images we commonly discuss in this section of the magazine! This technique produces 1D (a spectrum, λ) or 3D data (a spectral hypercube, x × y × λ) obtained by acquisition at different wavelengths. A spectral dataset can be thought as either a collection of images at a specific wavelength stacked together or as a collection of wavelengths at different pixel locations. Moreover, spectral imaging is not only used in the biomedical imaging field, but it has also been applied to environmental monitoring, materials characterisation, and much more.

However, given its atypical nature, spectral images might not be suitable for standard techniques we apply to the analysis of other medical images. That's why SpectrAI provides numerous built-in spectral data pre-processing and augmentation methods, neural networks for spectral data including spectral (image) denoising, spectral (image) classification, spectral image segmentation, and spectral image super-resolution.

Installing SpectrAI is super easy. It requires Python 3.8 so make sure you have it installed on your machine. We used this line on terminal (for Mac) to set it up as default python3:

```
ln -s -f /usr/local/bin/python3.8 /usr/local/bin/python3
```

We can install it through pip: pip3 install spectrai

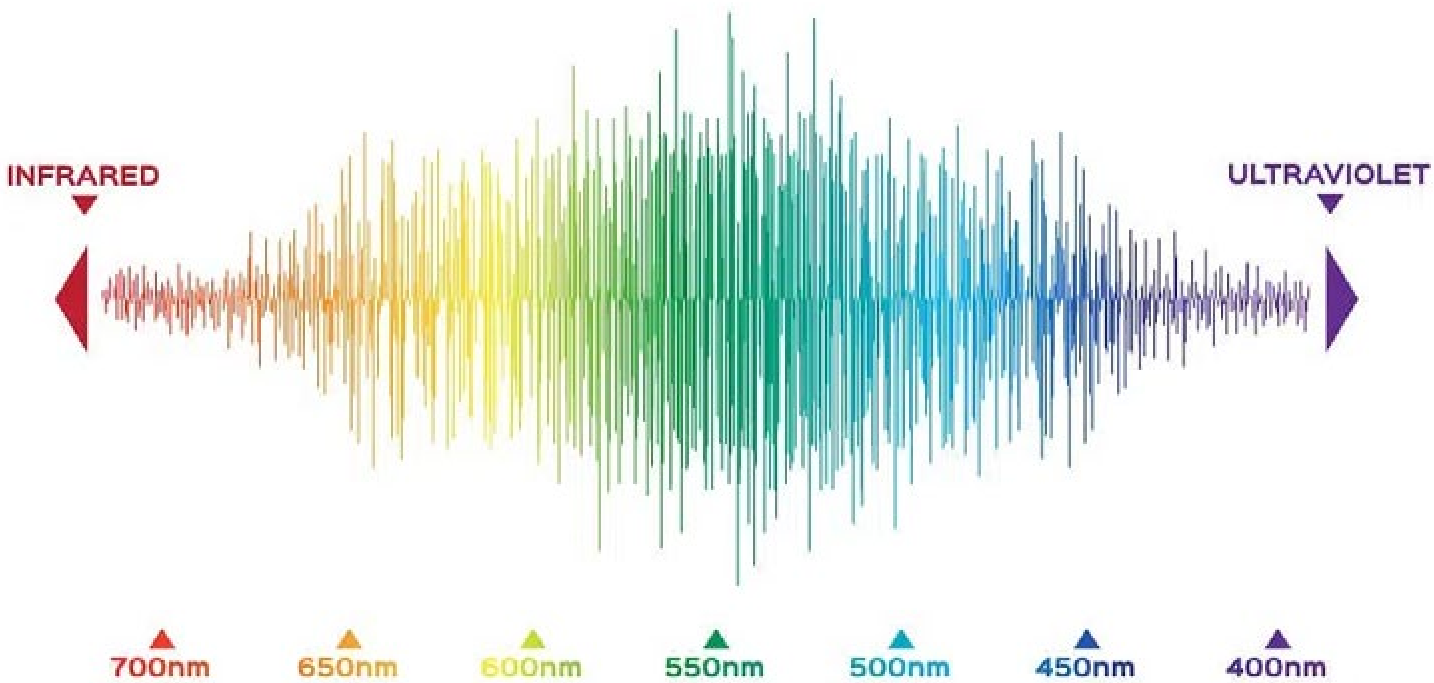
Or cloning the github repository. I suggest this option as more complete.

```
git clone git@github.com:conor-horgan/spectrai.git

cd spectrai

pip3 install -e .
```

Once the toolkit is successfully installed, to train a network, one can run the command: *spectrai_train*. This calls the python script train.py which automatically sets pre-processing and network parameters, by reading a config file. By default, the config file called is one for multi-resolution task.

For a new task, you can choose between several available networks (DenseNet, EfficientNet, RCAN, ResNet, SegNet, U-Net) and standard augmentation types (e.g. flip, crop) or spectral specific ones (e.g. spectral shift, spectral flip). Depending on your application, you can choose the desired combination and create your own config file, which can be executed by typing the following command:

```
spectrai _ train --config custom _ config.yml –verbose
```



See a nice image of spectral image lambda stack here.

The default config file looks like this:

```
# Default image super-resolution configuration
Task_Options:
  task: 'Super-Resolution'                # deep learning task [Calibration/Classification/Denoising/Segmentation/Super-Resolution]
  classes: 0                              # no. classes for Classification or Segmentation, zero otherwise

Training_Options:
  training_option: 'Train From Scratch'   # training option ['Apply Pre-Trained Network'/'Train From Scratch'/'Transfer Learning']
  pretrained_network: 'None'              # path to pretrained network, 'None' if training_option = 'Train From Scratch'
  pretrained_classes: 0                   # no. classes for pretrained Classification or Segmentation model, zero otherwise

Network_Hyperparameters:
  network: RCAN                           # network architecture [UNet/ResUNet/ResNet/RCAN]
  dimension: '2D'                         # convolution dimension for images ['2D'/'3D']
  activation: ReLU                        # activation function [ReLU/LeakyReLU/PReLU]
  normalization: 'None'                   # normalization layer [None/BatchNorm/LayerNorm/InstanceNorm/GroupNorm]
Training_Hyperparameters:
  epochs: 1                               # no. training epochs
  batch_size: 2                           # batch size
  learning_rate: 0.0001                   # learning rate
  input_image_size: 16                    # input image size (only applies to Super-Resolution)
  target_image_size: 64                   # target image size (only applies to images)
  spectrum_length: 500                    # spectrum length
  optimizer: Adam                         # optimizer [Adam/Adagrad/SGD/RMSprop]
  scheduler: Constant                     # learning rate scheduler [Constant/Step/Multiplicative/Cyclic/OneCycle/ReduceOnPlateau]
```

```
  criterion: L1                          # criterion [L1/'L2 / MSE'/'Cross Entropy'/'Binary Cross Entropy']
Preprocessing:
  spectral_crop_start: 0                 # start index for spectral crop window
  spectral_crop_end: 500                 # end index for spectral crop window
  background_subtraction: 'None'         # spectral background subtraction [None/'Automatic Least Squares'/'3rd Order Polynomial'/'5th Order Polynomial'/'Minimum Value Offset']
  data_normalization: 'Max Value'        # spectral normalization [None/'Max Value'/'Area Under The Curve']

Data_Augmentation:
  horizontal_flip: 1                     # horizontal flip (images)
  vertical_flip: 1                       # vertical flip (images)
  rotation: 1                            # rotation (images)
  random_crop: 1                         # random crop (images)
  spectral_shift: 0.1                    # spectral shift (spectra, images)
  spectral_flip: 1                       # spectral flip (spectra, images)
  spectral_background: 0                 # add spectral background signal (spectra, images)
  mixup: 1                               # mixup (spectra, images)

DataManager_Options:
  data_format: 'Image: H, W, C'          # data format ['Image: H, W, C'/'Image: C, H, W'/Spectra]
  data_directory: 'True'                 # is data input a directory [True] or a file [False]
  train_input_data: 'None'               # path to training input data ('None' if N/A)
  val_input_data: 'None'                 # path to validation input data ('None' if N/A)
  test_input_data: 'None'                # path to testing input data ('None' if N/A)
  train_target_data: "test_images/cells/"
# path to training target data ('None' if N/A)
  val_target_data: 'None'                # path to validation target data ('None' if N/A)
```

```
  test_target_data: 'None'          # path to testing target data
('None' if N/A)
  shuffle: 'True'                   # shuffle data [True/False]
  seed: 'None'                      # random seed for data shuffling
[integer value or 'None']
  train_split: 'None'               # percentage of data used for
training set [0.0 - 1.0 or 'None']
  val_split: 'None'                 # percentage of data used for
validation set [0.0 - 1.0 or 'None']
  test_split: 'None'                # percentage of data used for
testing set [0.0 - 1.0 or 'None']

State_Dicts:
  net_state_dict: 'None'            # network state dict (if
pretrained_network != 'None')
  optimizer_state_dict: 'None'      # optimizer state dict (if
pretrained_network != 'None')
  scheduler_state_dict: 'None'      # scheduler state dict (if
pretrained_network != 'None')
```

To evaluate or apply a pretrained model, you can also run: *spectrai_evaluate* and *spectrai_apply*, in which case you need to make sure that your Training_Options and State_Dicts parameters in the config file contain information about the pretrained network!

We also tried the Matlab GUI (requirement: Matlab v.2020b). For this, make sure you have the .mlapp file in your folder. Then go to the Matlab command window and setup the correct Python version:

```
>> pyversion('/usr/local/bin/python3.8')

>> pyenv

ans =

  PythonEnvironment with properties:

          Version: "3.8"
       Executable: "/usr/local/opt/python@3.8/bin/python3.8"
          Library: "/usr/local/Cellar/python@3.8/3.8.16/Frameworks/
Python.framework/Versions/3.8/lib/libpython3.8.dylib"
```

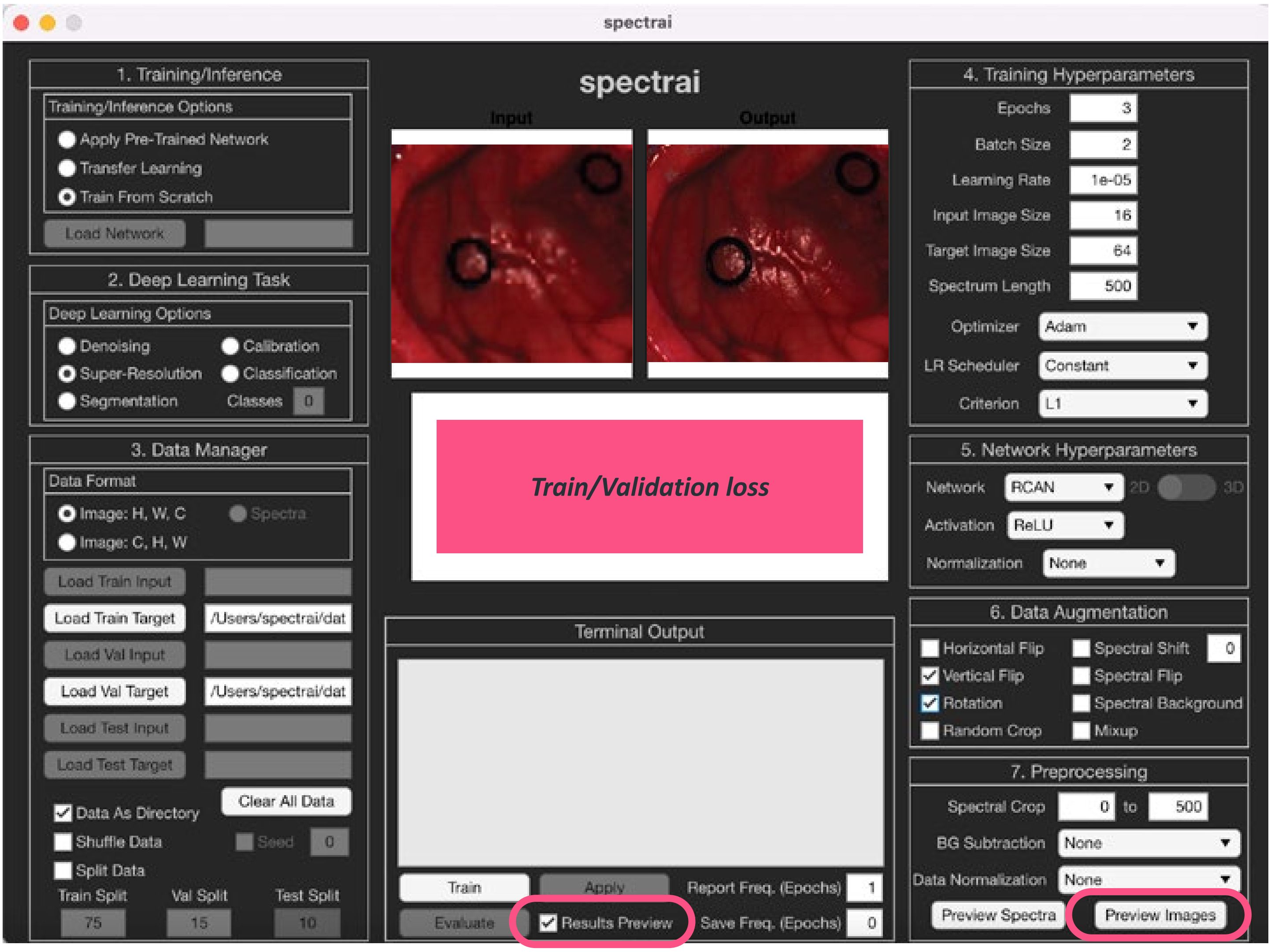
```
                        Home: "/usr/local/Cellar/python@3.8/3.8.16/Frameworks/
Python.framework/Versions/3.8"
                Status: NotLoaded
         ExecutionMode: InProcess
```

You can then move to the app directory and start it:

```
>> cd ~/spectrai/spectrai
>> spectrai
```

The app is a user-friendly GUI that allows you to select parameters instead of manually modifying the config file. It has a useful feature which disables options not suitable for the selected task and it can show the result of the pre-processing or network output and training curves. Nothing left but to try it on your own spectral dataset!

We thank **Conor C. Horgan** and **Mads Bergholt** for the realisation of **SpectrAI** and for the talk at **King's College London** two weeks ago which inspired this article!



SYBIL: A VALIDATED DEEP LEARNING MODEL TO PREDICT FUTURE LUNG CANCER RISK FROM A SINGLE LOW-DOSE CHEST COMPUTED TOMOGRAPHY

Peter G. Mikhael and Jeremy Wohlwend are second and third-year PhD students in Regina Barzilay's group at MIT. They speak to us about their paper proposing a novel AI tool for lung cancer risk prediction using low-dose CT scans, which the Journal of Clinical Oncology has just published.

Sybil: Predict Future Lung Cancer Risk

Early detection is key in improving the survival rate for **lung cancer**. Screening tests can help identify lung cancer early, even before symptoms appear. This work explores whether the low-dose CT scans used for screening today could also be used to **forecast the risk of someone getting cancer within a certain number of years**.

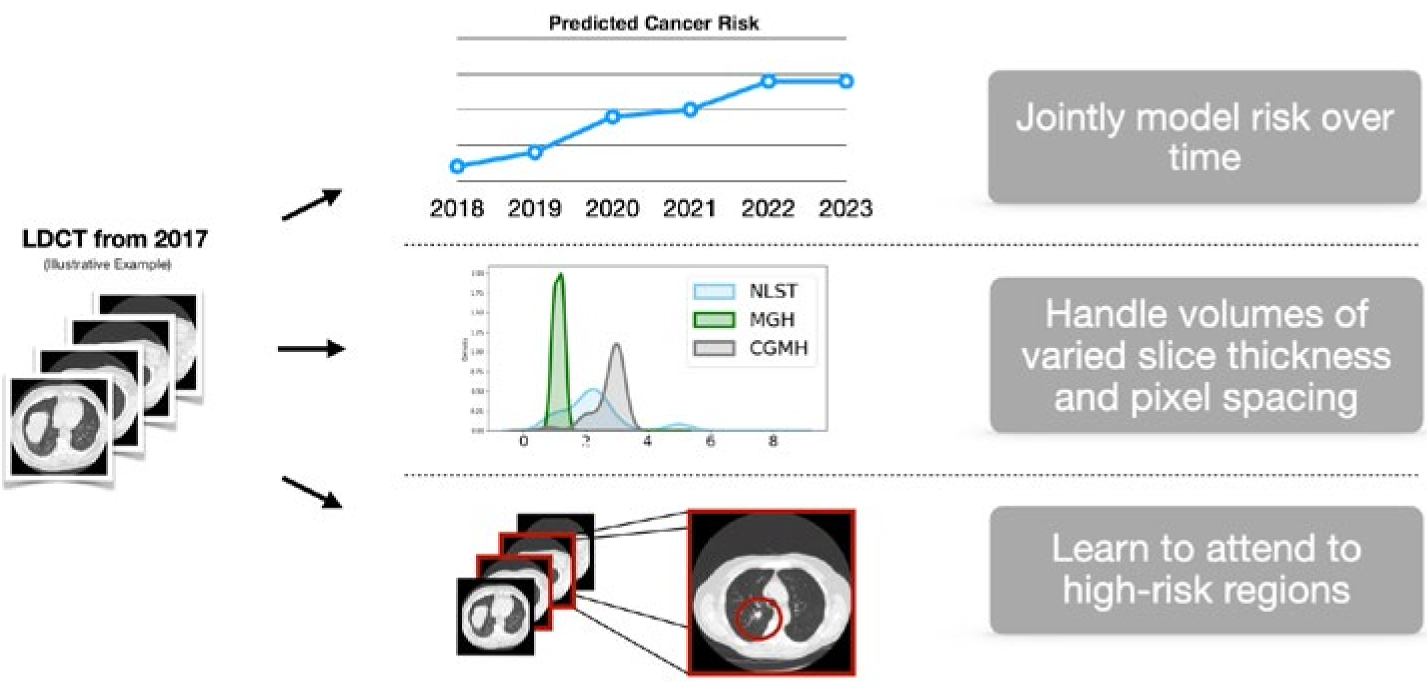
*"**Our model predicts whether or not cancer will develop within the next one to six years**,"* Peter tells us. *"We achieved this using data from a big trial that happened in the early 2000s, but we also wanted to validate it beyond that one trial, so we had data from our collaborators here in Boston at Massachusetts General Hospital (MGH) and the Chang Gung Memorial Hospital in Taiwan and were able to show that amongst those two populations, the model maintained its performance."*

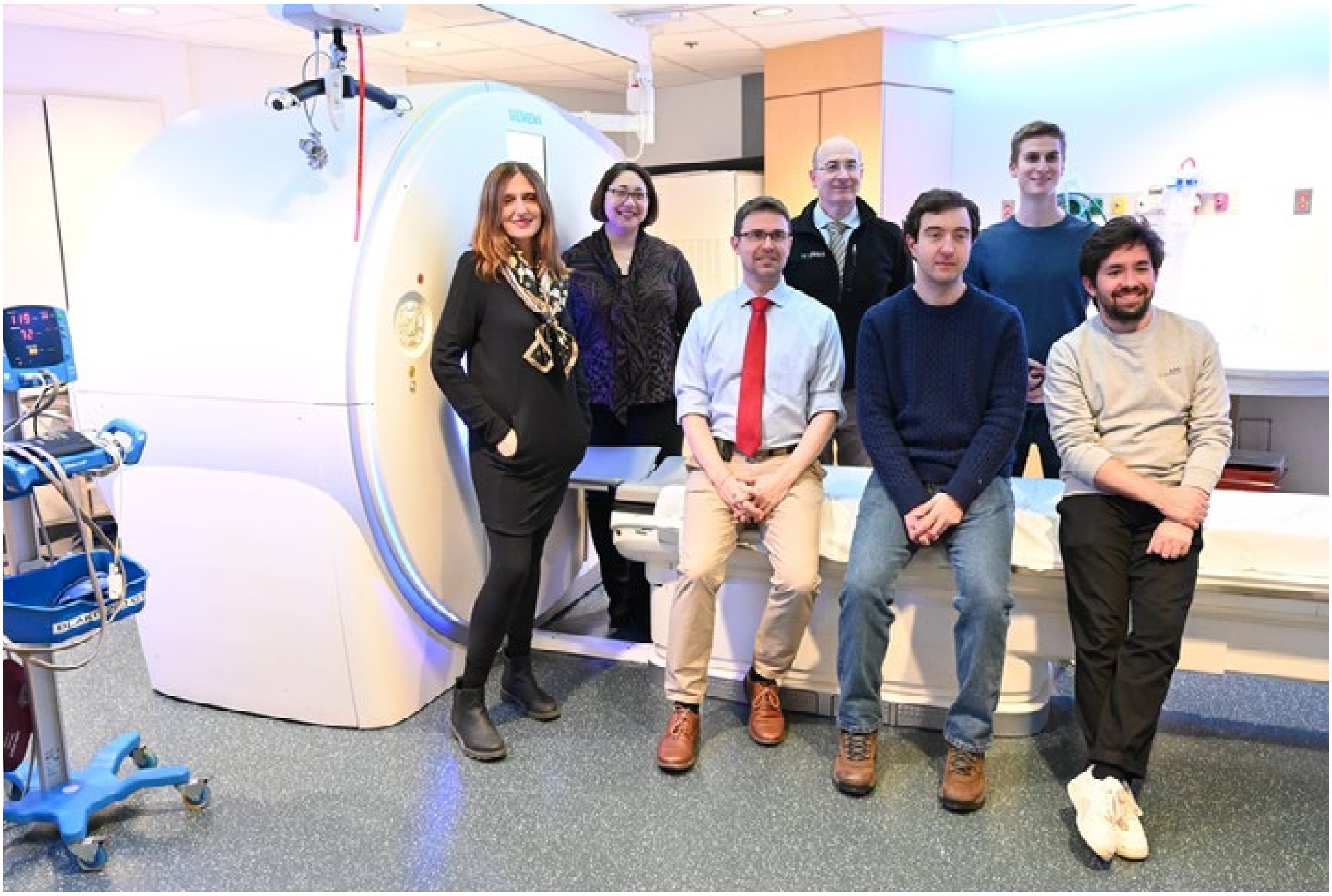
Low-dose CTs are typically recommended for people aged 50 years and over with a smoking history, but the number of people without any smoking history developing lung cancer is growing.

"One of our collaborators is looking at a cohort of firefighters, and in Taiwan, there's a big concern that cancer could be related to pollution," Jeremy points out. *"Ideally, screening should be something everybody does, but of course, that would be expensive. Hopefully, these technologies enable more efficient screening, meaning you can scan more people within the same budget. I don't think that will be the initial aim, but it's important if you want to expand to a broader population."*

Computational models that identify cancers in CT have already been developed, but to the best of the group's knowledge, their model is one of the first to forecast so far ahead. Radiologists are skilled at detecting cancers but **predicting what might happen in the future**, which is critical when looking at modifications to screening policies, is more complex.

The group felt confident their model would achieve its aims, having proven a similar concept in **breast screening with mammograms**, but were aware that





using CT this time around introduced new challenges.

"The amount of information you have per patient is more constrained for a mammogram, whereas CTs are big volumes," Peter explains. *"Also, mammography is much more standard. A lot of people get mammograms, so you have hundreds of thousands of data points. CT screening isn't as popular or well adopted."*

The group was inspired by a **Google AI** model that uses CT scans to predict lung cancer and lung nodule malignancy over a shorter term. Risk models have also been developed to decide who should be screened, considering clinical risk factors, including age and smoking history, and other correlated variables, such as education. These models can predict risk for up to six years but do not consider CT scans, limiting their performance.

Peter and Jeremy knew that **a combination of these two ideas could work**. By using high-quality 3D volumes, they were sure that **signs of cancer, if present, could be extracted**. With reasonable accuracy, it is possible to tell someone's age and smoking history from a CT alone.

The development of their deep learning model, **Sybil**, named after a prophetess in Greek legend and literature, has been four years in the making.

"If you consider the whole pipeline, it feels long, but for Peter and I, the active model development time was much shorter than that," Jeremy notes. *"Afterward, it was about **getting the right data** for validation,*

accessing it, managing the collaborations, and dealing with the review process. That takes a long time."

Early on, the group realized they needed doctors to tell them if cancer was present in the low-dose CT scans and where it was. It then took some time to complete the **manual task of labeling hundreds of scans for training**.

The team ruled out using clinical risk factors in the model and found that combining multiple CTs for a patient over time, as a doctor would do in a clinical setting, put too much strain on the system. In the end, the group's most significant challenge came down to one thing: **scale**.

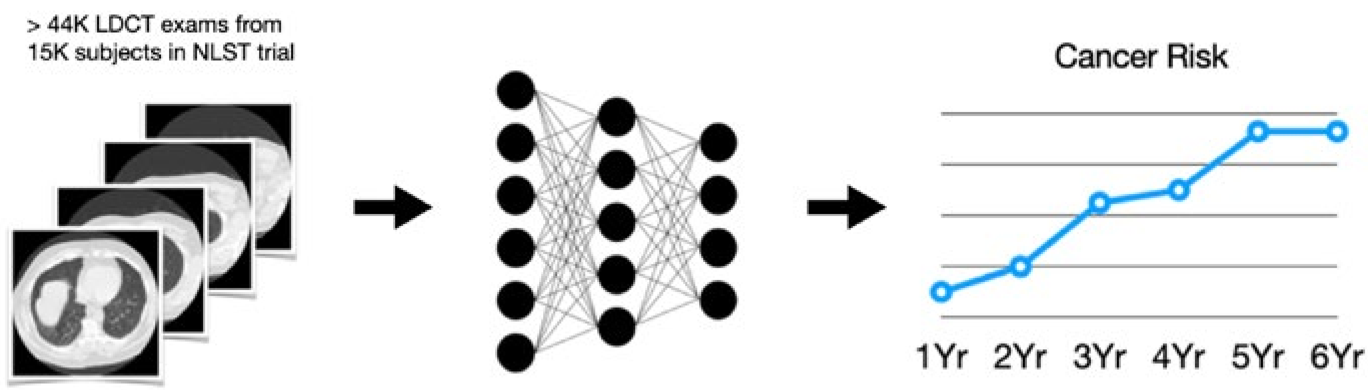
"When you work with super large images, you have potentially hundreds of slices, and with these huge volumes, ***things can get heavy computationally****,"* Jeremy explains. *"We hit many roadblocks with just being able to run experiments and iterate fast enough. We were lucky to have bigger, better machines with more state-of-the-art GPUs over time. We'd be in a much better situation if we were to start the project today, as our compute at the beginning was not even remotely close to where it is now."*

Nonetheless, with much engineering, **some clever work on neural architectures**, and a few trade-offs, the team has finally achieved its goal, but the work does not stop here. **Lecia V. Sequist** and **Florian J. Fintelmann** at MGH are at the forefront of the project's next stage. Analyses and studies are being run in their clinical setting to understand whether Sybil can maintain its performance enough to be used routinely in the real world.

"We don't know the answer yet, so they're trying to figure that out and, simultaneously, establish some protocols with other collaborators," Peter says. *"****We'd like more validation on patients not represented in our datasets****, which is necessary for it to become something used daily in the clinic."*

The group plans to make its **code and models available to the community** so that people can analyze them, research the topic, and hopefully produce even better models in the future.

"As a group, we're going more into chemistry and biologics, trying to model proteins and small molecules," Jeremy reveals. *"We're moving a little bit away from imaging and diagnostics and are trying to* ***use AI to accelerate therapeutic development****. I think that's the next big frontier for us!"*



Congrats, Doctor Stine!

Stine Hansen recently finished her PhD with the UiT Machine Learning Group and SFI Visual Intelligence at UiT The Arctic University of Norway. Her research aimed at developing machine learning-based models for data-efficient medical image volume segmentation that only require limited supervision. She now works as a researcher at UiT, where she will continue her research on machine learning and medical image analysis. Congrats, Doctor Stine!

Most current methods for machine learning-based medical image segmentation are supervised models that require large amounts of fully annotated images. However, obtaining such datasets in the medical domain can be difficult and expensive due to the specialized expertise and resources required to generate them. A wide-spread use of machine learning based models for medical image segmentation therefore requires the development of data-efficient algorithms that only require limited supervision.

When working in the limited supervision paradigm, exploiting the available information is key. During my PhD, I focused on addressing this challenge by developing new machine

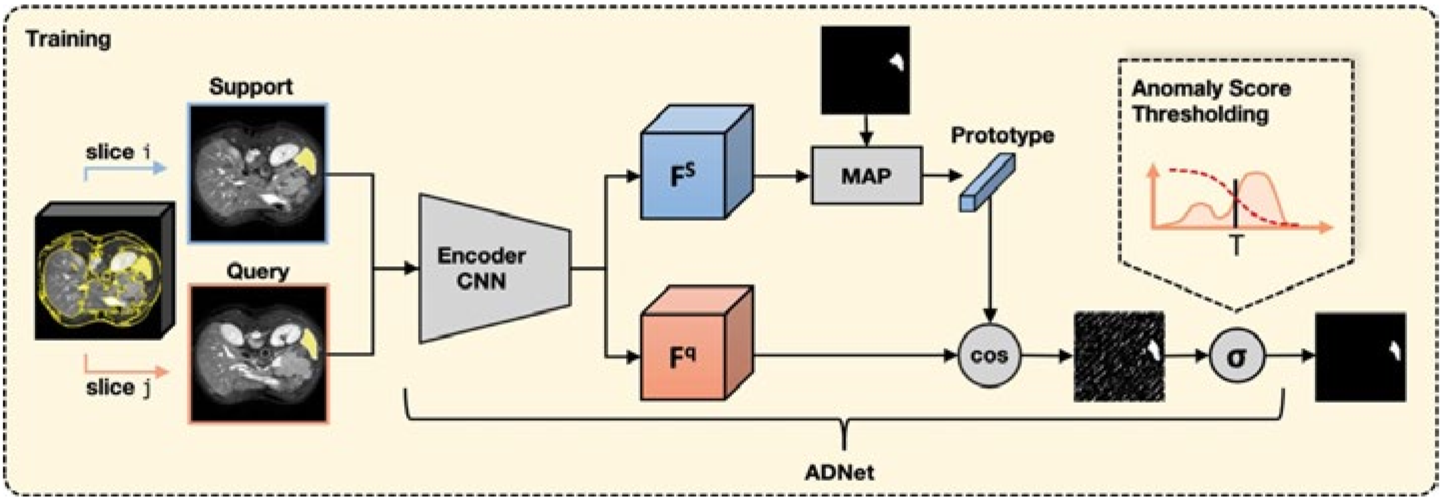


Figure 1: Anomaly detection-inspired network for few-shot medical image segmentation.

learning methodology that leverages automatically generated supervoxels in various ways to exploit the structural information in the images. Specifically, I explored few-shot learning-based solutions to organ segmentation and unsupervised approaches for lung tumor segmentation.

In the work on few-shot learning-based organ segmentation, supervoxels are used to generate pseudo-labels for self-supervised training. Additionally, to obtain a model that is robust to the typically large and inhomogeneous background class, a novel anomaly detection-inspired classifier is proposed to ease the modelling of the background (Figure 1). Further, to encourage the resulting segmentation maps to respect edges defined in the input space, a supervoxel-informed feature refinement module is proposed to refine the embedded feature vectors during inference (Figure 2). The module automatically identifies areas requiring refinement based on the predictive uncertainty.

The work on unsupervised tumor segmentation explores the opportunity of performing clustering on a population-level (as opposed to patient-level) in order to provide the algorithm with as much information as possible. To facilitate this population-level across-patient clustering, supervoxel representations are exploited to reduce the number of samples, and thereby the computational cost. Within the proposed segmentation framework, five variations of classical clustering algorithms are investigated to evaluate their sensitivity to tumor size and voxel noise, in addition to analyzing the type of segmentation mistakes they are prone to making and quantifying their associated benefit of the population-level clustering.

In summary, through the work of my thesis I have demonstrated that supervoxels are versatile tools for leveraging structural information in medical images when training segmentation models with limited supervision.

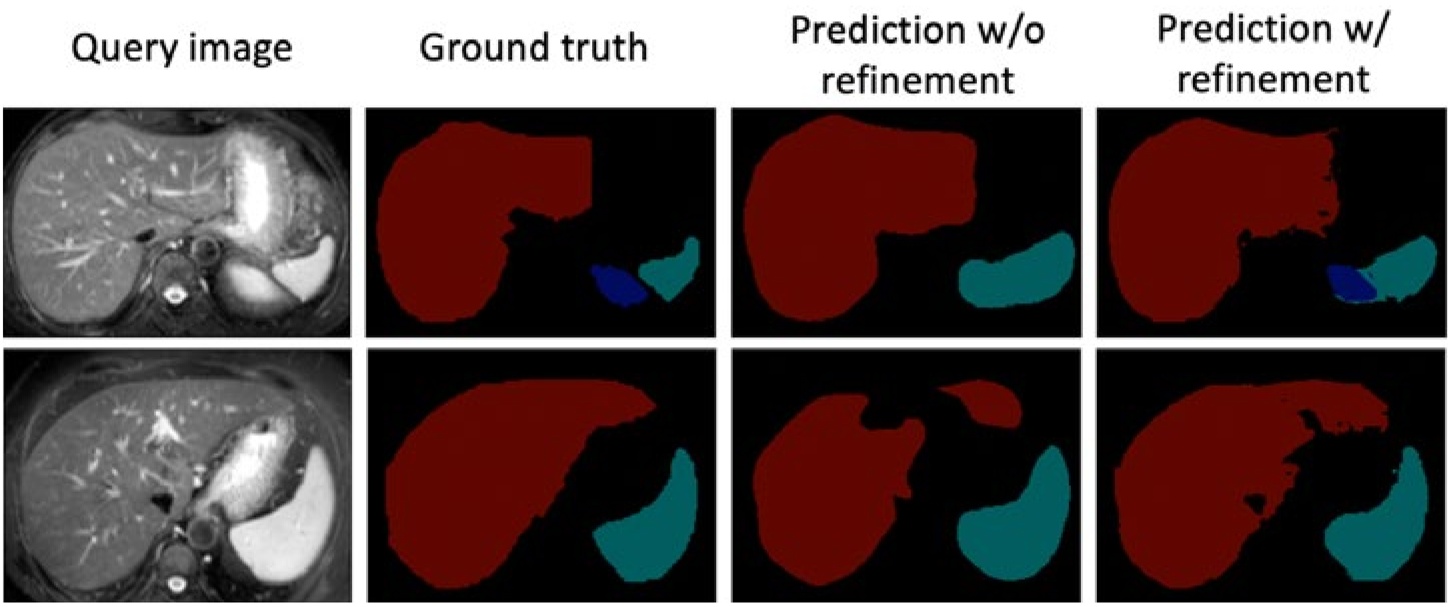


Figure 2: Few-shot organ segmentation with and without supervoxel-informed feature refinement. Red: Liver, Blue: Left kidney, Cyan: Spleen.

ROBUST NON-RIGID REGISTRATION CHALLENGE FOR EXPANSION MICROSCOPY (RNR-EXM)

Emma Besier, a Master of Engineering student at Harvard University, Ruihan Zhang, a fourth-year PhD student at the MIT Media Arts & Sciences Program in Edward Boyden's lab, and Donglai Wei, an Assistant Professor of Computer Science at Boston College, are co-organizers of an exciting ISBI 2023 challenge. Emma, Ruihan, and Donglai are here to tell us about it.

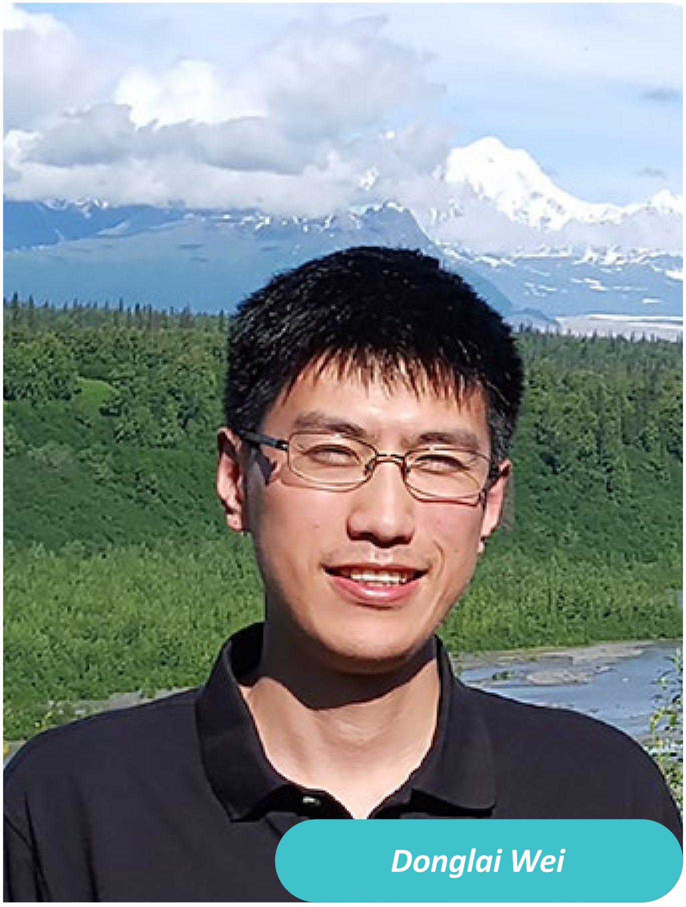
Emma Besier

Ruihan Zhang

Expansion microscopy is a technique to image small tissue samples that would not otherwise be visible on a microscopic level. **A gel polymer is** used to expand the sample, similar to how diapers expand when wet.

"Expansion microscopy is really novel, but it also presents some interesting challenges when it comes to registration," Emma tells us. *"The labs do **multi-round imaging**, where the same tissue sample is stained multiple times and sometimes over multiple days. When this happens, it can deform, lose staining intensity, and become noisier when you image it. We thought this would lend itself to a good challenge."*

The organizers have provided diverse data for participants to register generated from the **Boyden lab**, including C. elegans, zebrafish, and mouse data. These animals are considered model organisms in the biology community.



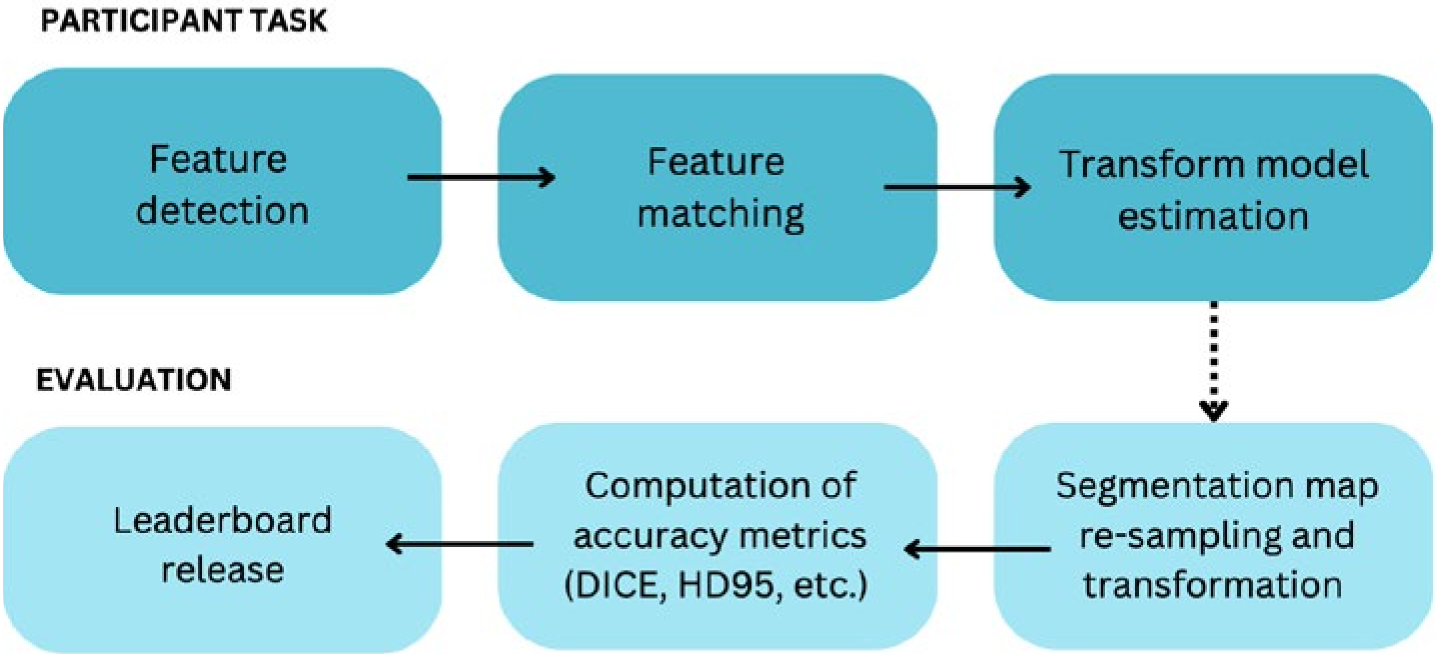
Donglai Wei

The project aims to **develop a benchmark data set for expansion microscopy image registration**, which could be used in tandem with developing new registration methods. There are already benchmark datasets for other medical imaging modalities, such as MRI and CT, but to the team's knowledge, none have been developed to date for expansion microscopy.

An ideal result from this challenge, Emma reveals, would be to see many **different approaches to registering the microscopy images**, including learning-based and traditional optimization-based approaches, as well as runtimes for each and a quantitative assessment of how they compare against each other. This work involves massive data, so accuracy alone is not enough.

"We're asking participants to submit runtimes on a purely honor-code basis," she reveals. *"The time it takes to register the images doesn't necessarily affect the registration quality. It's a small part of the experiment but can be super time-consuming. Five hundred volumes might need to be registered from all the different staining grounds for a single experiment. Finding methods to reduce the time it takes to do this step is important."*

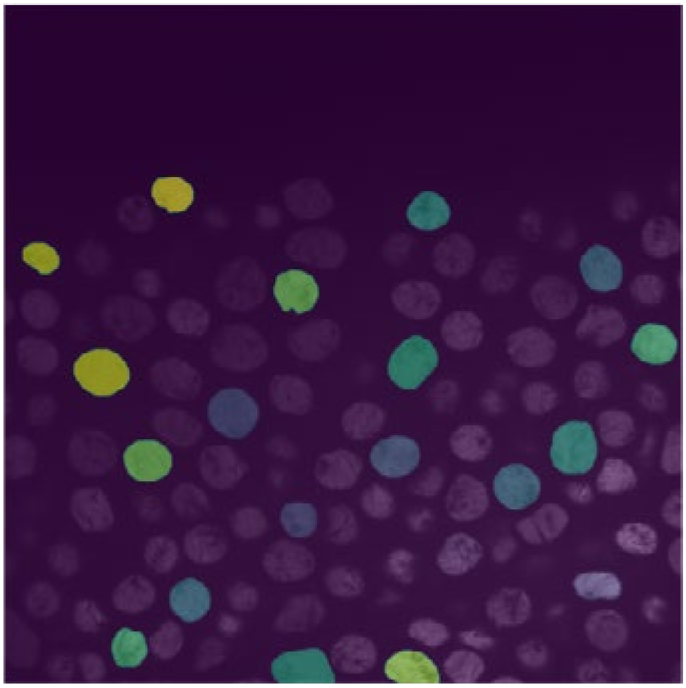
With the window for submission of results opening shortly, does Emma have any last-minute tips for participants?



"When you download the image data, ***look at it in 3D to have a clearer idea of what the offset and deformation actually look like****,"* she advises. *"It can be easier with volumes of this size to plot things in 2D in a Jupyter Notebook or something, but taking the time to overlay the volumes in a 3D space will help you to understand the deformation and how to approach registering it. The zebrafish data has mostly linear deformations, large rotations, and a large amount of offset in the Z-axis. The mouse data has the most nonlinear deformations, which is probably the most challenging. It also has the biggest difference in staining intensity and image quality."*

The Boyden lab routinely generates terabytes of data, and its papers are often cited, so it urgently needs registration to work perfectly. Registration is used for critical downstream tasks, such as **creating a 3D reconstruction of the entire brain**, so it is imperative to have a precise and almost exact registration of the image pairs for this volumetric analysis.

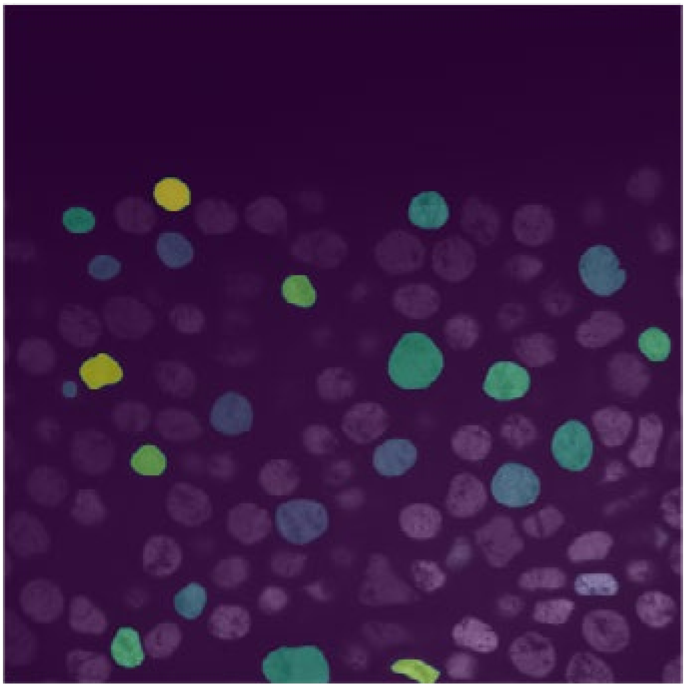
"Biologists traditionally only care about preparing a sample, imaging it once and that's it, but now we really want to get more information from these samples," Ruihan explains. *"We want to align two rounds of data together. This combinatorial encoding will be the future trend of biology. That's*



why precise registration is essential."

Donglai Wei completed his PhD at MIT and his postdoc at Harvard, which is how he linked up with Emma and Ruihan to work on this project. He has been contacted by people from different labs excited about this competition and hoping it will become an annual event so they can participate and donate their data in the future. They view it as an opportunity to create a bigger data set and a more challenging problem.

"Registration is a classic problem for microscopy, but we all want a deep learning revolution for this field," he tells us. *"The problem is a lack of data, but* ***ImageNet was not built in a day****. We only have a small data set, but we hope to make it more complete and diverse over the years. This project is only one small step towards* ***making expansion microscopy a more sure technology*** *– not only imaging wise but also processing wise to support neuroscientists' investigations."*



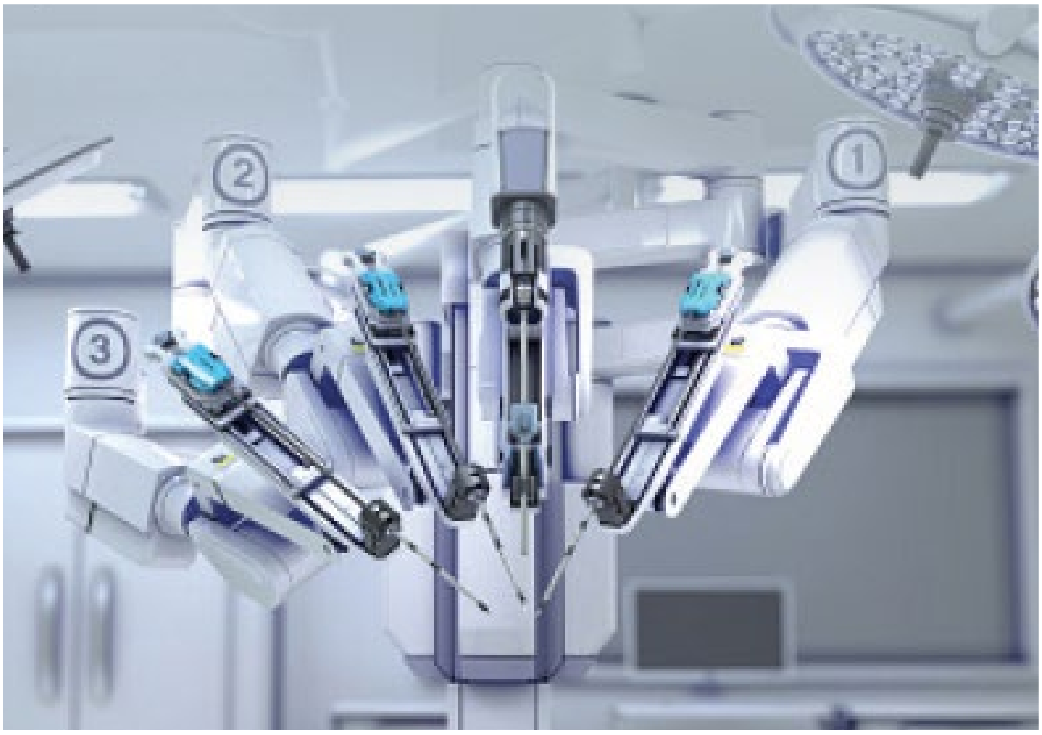
Ruihan adds:

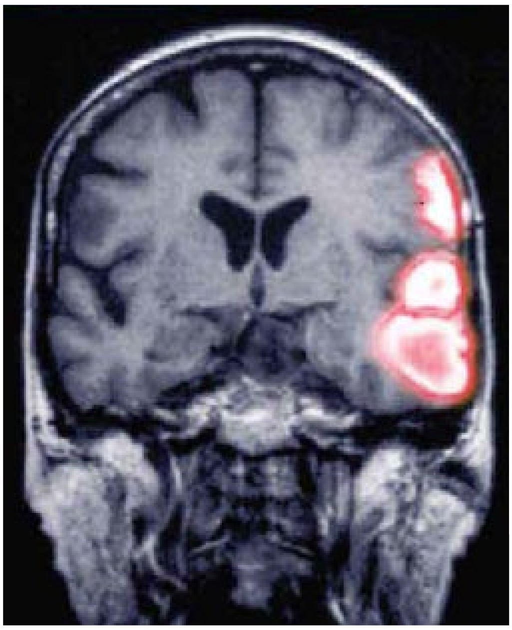
"If we solve this computational problem, more wet lab biologists will use our technology because they'll see how easy it is. It's an iterative process. If we make the accuracy 80%, then labs that can generate high-quality data sets can already use it. If we make it 90% accurate, people who prepare less good samples can use it. If we lower it, then everybody can use it. There's always an audience for our method."

The team want this challenge to facilitate a discussion around expansion microscopy registration and encourage the release of more open-source code for registering this type of data. Many labs have developed in-house methods, but the hope is that this will create something more communal.

Emma will be graduating in the spring. Meanwhile, she is working full-time on her thesis, of which this challenge has been a big part.

"I felt confident that I could do it, even though it would be a lot of work," she recalls. *"The results of this challenge will give important information to the* ***Boyden lab*** *and the* ***Visual Computing Group at Harvard****, two groups that I'm affiliated with. The kind of infrastructure we've set up, the evaluation tools, the website, everything can be used for future iterations of this when we're building up our microscopy dataset. I wanted to be a part of something that would have useful results for the lab but also create a long-term impact."*





RSIP VISION

IMPROVE YOUR VISION WITH

Computer Vision News

SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!