MARCH 2022

Computer Vision News

The Magazine of the Algorithm Community

MEDICAL IMAGING NEWS

Page 23

DILBERT

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editors:
Marica Muffoletto
Ioannis Valasakis

Publisher:
RSIP Vision

Contact us

Free subscription

Read Previous Magazines

Dear reader,

Welcome to the March issue of **Computer Vision News**!

Marica Muffoletto reviews **MediaPipe**, a free, open-source framework for machine learning solutions. Written in C++ and able to be deployed anywhere, MediaPipe has already built a huge number of demos and projects. Read all about it on page 4.

ICLR 2022 is coming up next month and we're getting in early on page 10 with a review of one of its most intriguing workshops - **Elements of Reasoning: Objects, Structure, and Causality**. Can object-centric representation learning and causal machine learning work together?

Just over a decade after its initial release, **Richard Szeliski** has published the second edition of his all-encompassing textbook **"Computer Vision: Algorithms and Applications"**. In our interview on page 16, he tells us what has changed (spoiler: a lot!) and how you can all download your very own free copy.

In **"Medical Imaging News"** this month, we interview surgeon and researcher **Pier Giulianotti**, who has performed more than **5,000 robotic surgeries**! On page 24, he tells us all about his pioneering work.

On page 28, Ioannis Valasakis reviews a topical deep learning framework paper called **NaroNet: Discovery of Tumor Microenvironment Elements from Highly Multiplexed Images**.

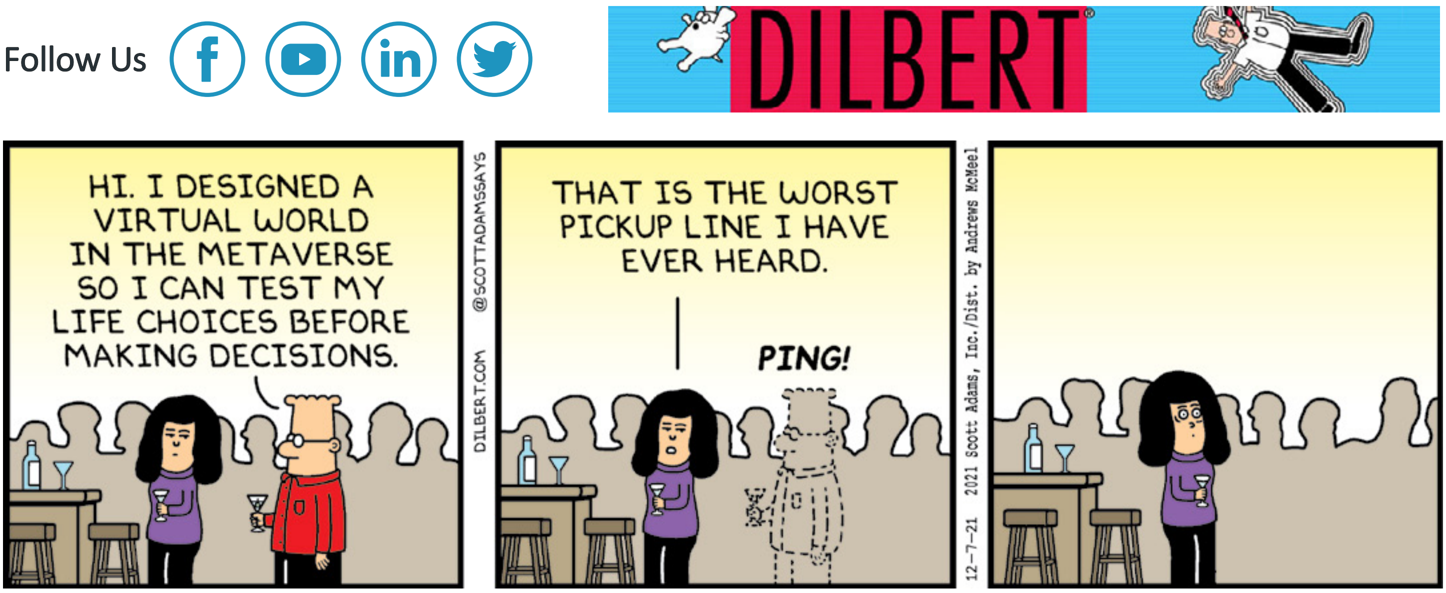
"That's been my dream!" Turn to page 38 to find out what has happened to make former MICCAI Society president **Wiro Niessen** and his research team at **Quantib** so excited about the impact they are going to have on the lives of people with cancer.

"I want to change medicine!" asserts **Bettina Baessler**, a brilliant radiologist and a research leader and educator in the field of radiomics. On page 44, she reveals all about how she is going to achieve this bold objective.

We hope you enjoy another busy issue of our magazine – please tell your friends and colleagues all about us!

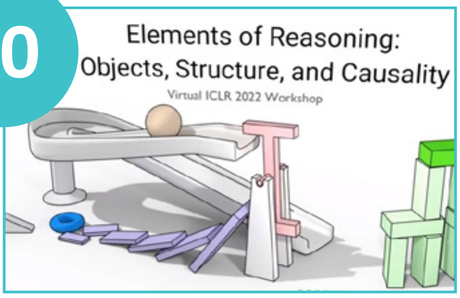
Ralph Anzarouth
Editor, **Computer Vision News**
Marketing Manager, **RSIP Vision**

Follow Us

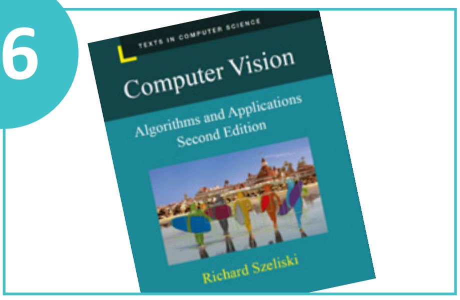


Computer Vision News

04

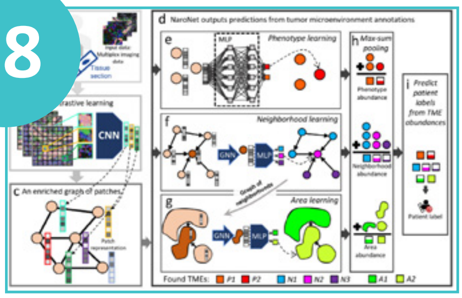

10


12


16


Medical Imaging News

24


28


38


44


INTRODUCTION TO MEDIAPIPE (1/2)

by Marica Muffoletto @maricaS8

Hi everyone, and welcome to a new full issue of **Computer Vision News** magazine and another review on a great computer vision tool which employs machine learning.

MediaPipe is an open-source framework (available here) for machine learning solutions. Currently still in the alpha stage, over the last 2 years since its first release MediaPipe built **a huge number of demos and projects** that demonstrate its use. MediaPipe is free, written in C++ and can be deployed to any platform, from web assembly to Android to MacOS.

MediaPipe ML pipelines use a wide range of algorithms, from classical computer vision software to state-of-the-art deep learning networks such as MobileNetV3, introduced by Google itself in 2017.

MediaPipe provides flexibility, through a modular architecture, and speed, which is warranted by the use of GPU acceleration and multi-threading. According to its users, it is in fact much faster than its competitors, making it an absolute world-class in the field.

What does it offer?

MediaPipe can be used for building multimodal (e.g. video, audio, any time series data), cross platform (i.e Android, iOS, web, edge devices) applied ML pipelines. It offers solutions for several computer vision problems including:

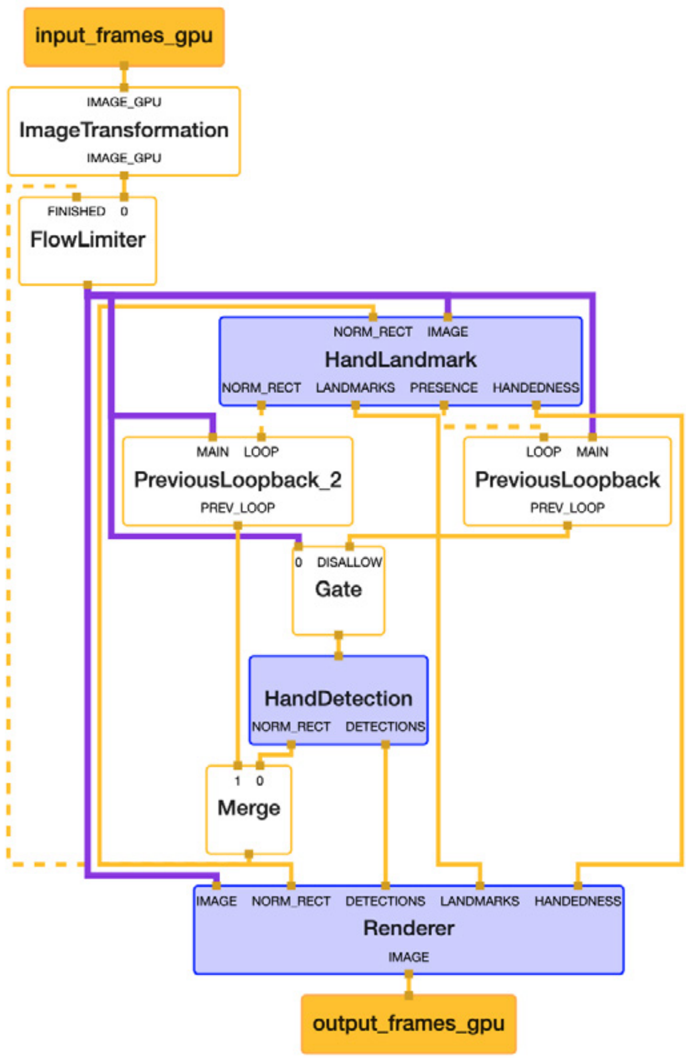
1. Face Detection ➔ solution that comes with 6 landmarks and multi-face support
2. Face Mesh ➔ algorithm that estimates 468 3D face landmarks in real-time even on mobile devices. It employs machine learning (ML) to infer the 3D surface geometry, requiring only a single camera input without the need for a dedicated depth sensor
3. Iris ➔ accurate iris estimation, able to track landmarks involving the iris, pupil and the eye contours using a single RGB camera, in real-time, without the need for specialized hardware
4. Hands ➔ high-fidelity hand and finger tracking solution. It employs machine learning to infer 21 3D landmarks of a hand from just a single frame
5. Pose ➔ it infers 33 3D landmarks and background segmentation mask on the whole body from RGB video frames

6. Holistic ➔ it combines all the previous tasks in real-time into a semantically consistent end-to-end solution requiring simultaneous inference of multiple, dependent neural networks.
7. Hair Segmentation
8. Object Detection
9. Box Tracking
10. Instant Motion Tracking
11. Objectron ➔ real-time 3D object detection solution for everyday objects. It detects objects in 2D images and estimates their poses through a machine learning model, trained on the Objectron dataset
12. KNIFT ➔ template-based feature matching tool

Most of these are offered for Android, iOS and C++ and some also for Python, JS and Coral.

How does it work?

Basically, **MediaPipe is a framework for Computer Vision and Deep Learning that builds perception pipelines**. MediaPipe works through graphs which help creating machine learning-based end-to-end tools. To get an idea of how such a pipeline functions, the best way is to load one of the graphs on the so-called visualiser (hosted at viz.mediapipe.dev) which lets you inspect graphs, modify them through the editor and experiment the result live. On the top right, you can choose which ones of the graphs to load. The graph on the right, for example, represents the Hand Tracking pipeline.



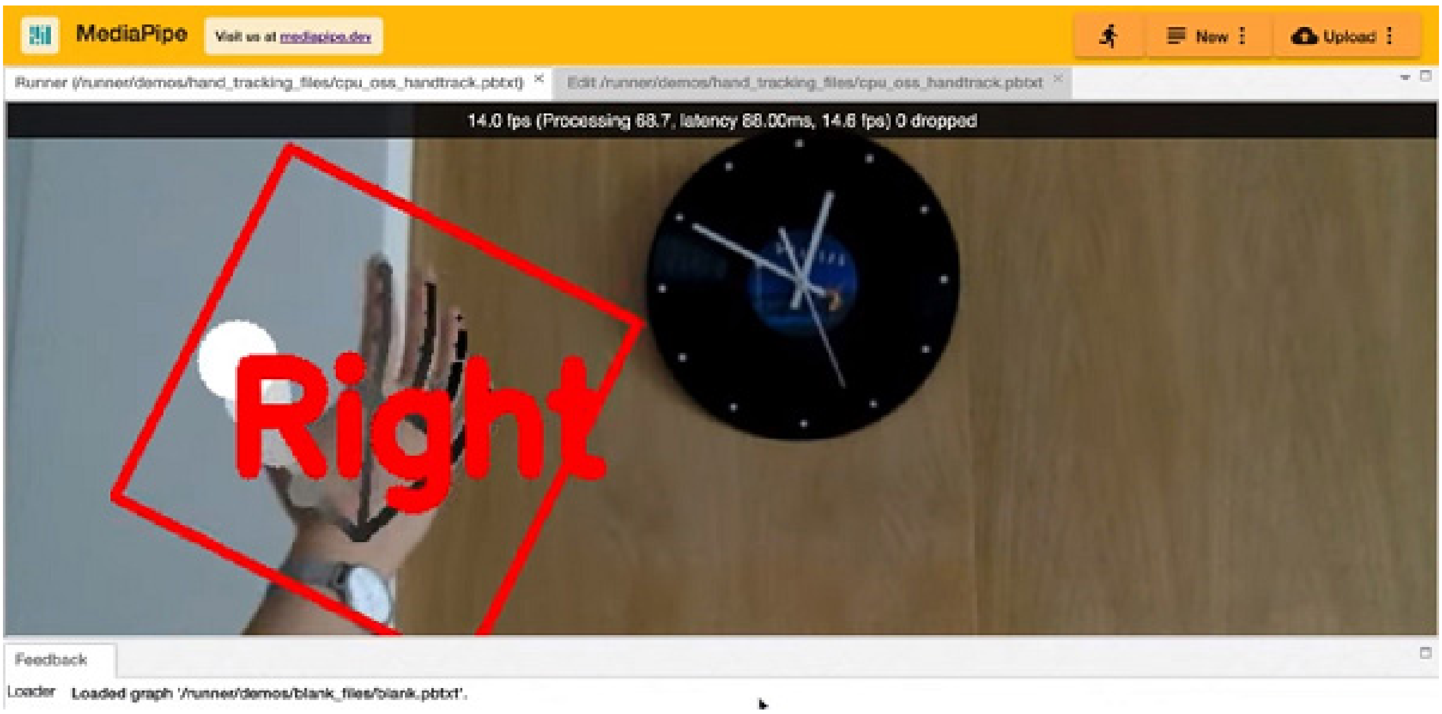
The purple boxes in this representation are the subgraphs which perform the main tasks, while all the other white boxes are calculators (node of the graph) performing necessary operations, such as managing input and output streams of images, image transformation and loops. A calculator may receive zero or more input streams and/or side packets and produces zero or more output streams and/or side packets (data containers).

These are the main nodes defined for Hand Detection, Hand Landmark and Annotation Rendering.

```
# Subgraph that detects hands
node {
  calculator: "HandDetectionSubgraph"
  input_stream: "hand_detection_input_video"
  output_stream: "DETECTIONS:palm_detections"
  output_stream: "NORM_RECT:hand_rect_from_palm_detections"
}

# Subgraph that localizes hand landmarks
node {
  calculator: "HandLandmarkSubgraph"
  input_stream: "IMAGE:throttled_input_video"
  input_stream: "NORM_RECT:hand_rect"
  output_stream: "LANDMARKS:hand_landmarks"
  output_stream: "NORM_RECT:hand_rect_from_landmarks"
  output_stream: "PRESENCE:hand_presence"
  output_stream: "HANDEDNESS:handedness"
}


# Subgraph that renders annotations and overlays them on top of the input
# images
node {
  calculator: "RendererSubgraph"
  input_stream: "IMAGE:throttled_input_video"
  input_stream: "LANDMARKS:hand_landmarks"
  input_stream: "NORM_RECT:hand_rect"
  input_stream: "DETECTIONS:palm_detections"
  input_stream: "HANDEDNESS:handedness"
  output_stream: "IMAGE:output_frames_gpu"
}
```



MediaPipe visualizer runner setting for Hand Tracking Pipeline

Use MediaPipe on Python

Installing MediaPipe tools on python can easily be done through pip.

```
python3 -m venv mp_env && source mp_env/bin/activate
pip install mediapipe
```

Once the environment is ready, we can import some basic libraries and mediapipe, and we can call the desired pipeline using the following lines.

```
import cv2
import math
import numpy as np
import mediapipe as mp
import glob

mp_hands = mp.solutions.hands
mp_drawing = mp.solutions.drawing_utils
mp_drawing_styles = mp.solutions.drawing_styles
help(mp_hands.Hands)
```

Running help on a MediaPipe module can give you useful insight on the class and how to use and modify its parameters. Running help on the hands pipeline gives you the list of parameters that can be modified to your convenience: *static_image_mode* (to treat the input images as static, rather than a video stream), *max_num_hands* to set the maximum number of hands to detect, *model_complexity* (ranging from 0 to 1), *min_detection_confidence* and *min_tracking_confidence* to vary the lower bound confidence for both the hand and landmark detection.

The remaining code allows to upload some static images and apply the technique to finally draw hand landmarks on the image itself, print the corresponding coordinates and draw a 3d plot of them.

```
filenames = glob.glob(basedir+"/*.jpg")
# Read images with OpenCV.
images = {name: cv2.imread(name) for name in filenames}
# Preview the images.
for name, image in images.items():
  cv2.imshow(name,image)
```

```
# Run MediaPipe Hands.
with mp_hands.Hands(
    static_image_mode=True,
    max_num_hands=2,
    min_detection_confidence=0.5) as hands:
 for name, image in images.items():
    # Convert the BGR image to RGB, flip the image around y-axis for correct
    # handedness output and process it with MediaPipe Hands.
    results = hands.process(cv2.flip(cv2.cvtColor(image, cv2.COLOR_BGR2RGB), 1))

   # Print handedness (left v.s. right hand).
    print(f'Handedness of {name}:')
    print(results.multi_handedness)

   if not results.multi_hand_landmarks:
      continue

   # Draw hand landmarks of each hand.
    print(f'Hand landmarks of {name}:')
    image_hight, image_width, _ = image.shape
    annotated_image = cv2.flip(image.copy(), 1)
    for hand_landmarks in results.multi_hand_landmarks:

     # Print index finger tip coordinates.
      print(
         f'Index finger tip coordinate: (',
          f'{hand_landmarks.landmark[mp_hands.HandLandmark.INDEX_FINGER_TIP].x * im
          f'{hand_landmarks.landmark[mp_hands.HandLandmark.INDEX_FINGER_TIP].y * im
     )
      mp_drawing.draw_landmarks(
         annotated_image,
          hand_landmarks,
          mp_hands.HAND_CONNECTIONS,
          mp_drawing_styles.get_default_hand_landmarks_style(),
          mp_drawing_styles.get_default_hand_connections_style())
    cv2.imshow("Landmarks", cv2.flip(annotated_image,1))
```

```python
# Run MediaPipe Hands and plot 3d hands world landmarks.
with mp_hands.Hands(

    static_image_mode=True,
    max_num_hands=2,
    min_detection_confidence=0.7) as hands:


 for name, image in images.items():
    # Convert the BGR image to RGB and process it with MediaPipe Hands.
    results = hands.process(cv2.cvtColor(image, cv2.COLOR_BGR2RGB))
    # Draw hand world landmarks.
    print(f'Hand world landmarks of {name}:')
    if not results.multi_hand_world_landmarks:
       continue
    for hand_world_landmarks in results.multi_hand_world_landmarks:
mp_drawing.plot_landmarks(
        hand_world_landmarks, mp_hands.HAND_CONNECTIONS, azimuth=5)
```

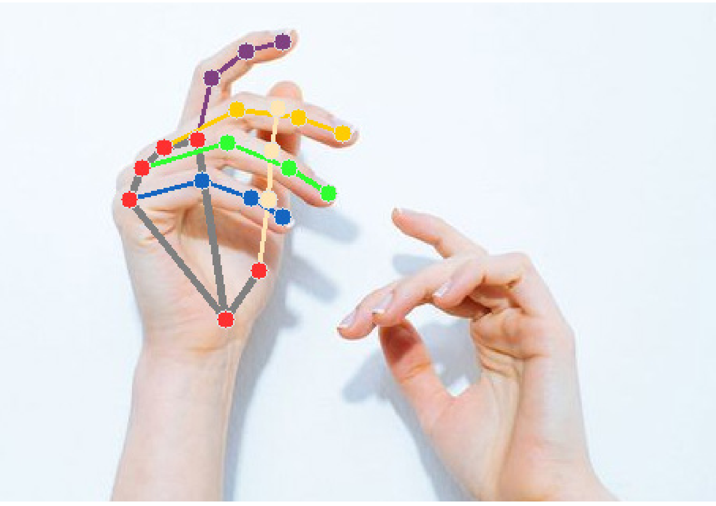


Figure 1: Result of Hand Tracking pipeline when using parameter min_detection_confidence set to 0.7

Figure 2: Result of Hand Tracking pipeline when using parameter min_detection_confidence set to 0.5

We are at the end of this first chapter on MediaPipe. We gave a basic introduction to this platform, and we looked at one of their machine learning pipelines in their Visualizer software and through the interaction with Python. In the second chapter of this series, we will further explore its functions and code base, while looking at the interaction with Javascript and mobile solutions.

I look forward to discovering more together, and to your feedback and questions on MediaPipe ☺

ELEMENTS OF REASONING: OBJECTS, STRUCTURE, AND CAUSALITY – VIRTUAL ICLR 2022 WORKSHOP

Wilka Carvalho is a PhD candidate in computer science at the University of Michigan, under the supervision of Satinder Singh and Honglak Lee. As a co-organizer of the first Objects, Structure, and Causality workshop at this year's Virtual ICLR, he is here to tell us more about the topic and what we can expect from the event in April.

Object-centric representation learning and **causal machine learning** have similar goals, but the two fields have been working independently. Both are motivated by the idea that structured representations of the world will enable agents to have more robust behavior and to generalize better, but they focus on different things.

Causality assumes structure and then shows how robust an agent can be. In contrast, **object-centric learning starts from an unstructured input** and aims to infer a useful decomposition into meaningful factors. Both fields will be represented at the Objects, Structure, and Causality workshop, which Wilka hopes will help foster collaboration.

"You can make good inferences about causality if you know the causal factors in the world, and one important causal factor is an object," he explains. *"We are getting to a point where we can at least infer objects in simple scenes, but there's no state of the art that merges these. In causality, nobody knows how to go from what you see with your eyes to the causal factors.* ***We hope we can get an end-to-end system that can discover the causal factors and then discover causality on top of that****. That will be an important part of future AI."*

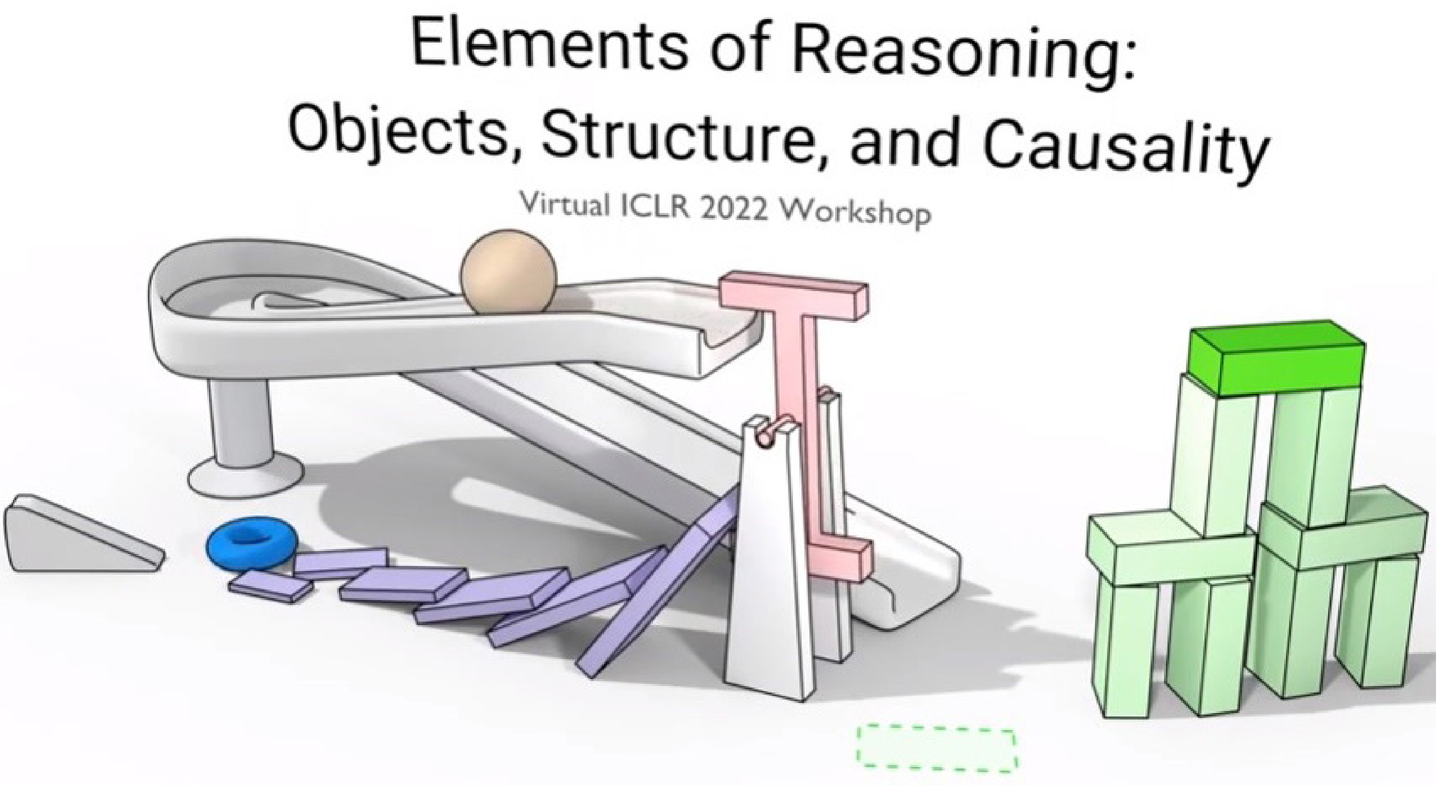
Outside of this, Wilka works in deep reinforcement learning, which he laughs makes him *"a bit of an outsider"* as one of the workshop organizers. His work involves developing networks that choose actions in virtual environments. Currently, he is working on a network that tries to find objects in a virtual kitchen. Part of that is understanding that objects have shared relations across experiences.

"I am so fascinated by the human ability to discover the pieces of the world and to leverage that to adapt our behavior," he tells us. *"For example, I'm in Boston now to give a talk at Harvard, and even though the kitchen here in this apartment is different, I have no problem generalizing my behavior to this setting. I know the pieces, and I know roughly how they relate, so I can recognize them in new settings. Last night, I needed to find a fork, so I looked around this completely new kitchen I'd never been in before and was able to find one. AI is not quite there yet."*

Reinforcement learning defines a framework for learning behavior that maximizes reward. Using the fork example, Wilka's reward for finding the fork was that he got to eat. That reward guided his behavior. **The next step is training agents to learn behavior guided by reward**.

As well as bringing together object-centric and causal researchers, the workshop will play host to speakers from a range of fields and perspectives, including core causality people like **Bernhard Schölkopf**, Director of the Max Planck Institute for Intelligent Systems and professor at ETH Zürich, and big object people like **Nikolaus Kriegeskorte**, a professor at Columbia University. There are also speakers from the worlds of machine learning, neuroscience, and psychology, including **Alison Gopnik**, a professor of psychology and affiliate professor of philosophy at the University of California at Berkeley.

"We're very excited by all the different perspectives," Wilka adds. *"Integrating different expertise on the same problem is such an important part of making progress. With our collection of speakers and participants, I do not doubt that we are poised to do that!"*



Congrats, Doctor!

Karen Tatarian recently defended with distinction her PhD at Sorbonne University while receiving a Marie Skłodowska-Curie ITN fellowship funded by the European Commission's H2020 projects.

While completing her PhD in Robotics and AI, Karen was also working as a robotics engineer and researcher at SoftBank Robotics in Paris, France.

Prior to her PhD, Karen completed her Bachelor of Science in Physics and Masters in Mechanical Engineering from the American University of Beirut. While her prior work in robotics focused on control and automation for various mobile robots as well as arm robots and drones, Karen was curious to better understand how the human fits in the equation as technologies and robotics are evolving at a rapid pace.

Since then, Karen has been on a quest to make human-centered socially intelligent robots and products a reality. Her research interests are in Robotics, Human-Robot Interaction (HRI), Machine Learning, and Reinforcement Learning (RL).

Congrats, Doctor Karen!

Overview:

By merely observing humans, one can directly infer that no social interaction takes place without cues, whether verbal or nonverbal, that allow others to interpret behaviors and reasonably estimate intentions. These powerful social signals and nonverbal behaviors are complex and multi-modal, which means they are made of different combinations of modalities and cues like gestures, gaze behavior, and proxemics (e.g., management of space and environment). Thus, **for a robot to be perceived as a socially-intelligent agent by humans**, it is expected to be able to hold a successful social interaction, adapt to the social environment, and exhibit appropriate multi-modal behavior. In my thesis, I first investigate how one of these modalities can help adapt another one, then explore the effects of the modalities when performed multi-modally on behavioral interaction outcomes and perception of the robot's social intelligence, and finally present an architecture using **reinforcement learning** for the robot to learn to combine its multi-modal behaviors with a reward function based on the multi-modal social signals of the human in an interaction.

Using human behaviors for robot adaptation in group settings:

Modalities are coupled in nature and for one to adapt to the changes in the environment it may need to rely on other sensory modalities. For instance, in my first thesis work, the robot would autonomously adapt its gaze pattern to social interaction changes, i.e groups being formed around it, based on its proxemics, which was used to estimate the roles of participants in group formations around the robot, such as active speaker, bystander, or overhearer (see Figure 1). A pilot study looked into group formations made of a robot and two users. Results showed that participants stood closer to the adaptive robot and ranked it higher in perceived adaptability and perceived sociability as well as feeling attended to by the robot in comparison to a robot that switched its gaze attention based on new sensory det ection [1].

Understanding the effects of the robot's multi-modal behavior:

Multi-modal behaviors made up of gaze mechanisms, which are turn-taking, turn-yielding, floor-holding, and joint attention, social gestures, which are emblem, deictic, and beat gestures, proxemics, through social navigation, and social dialogue were implemented autonomously and studied by extracting one modality in each condition and looking into behavioral outcomes and subjective measurements. An extract of the system is shown in Figure 2 and the code can be found in my github/KarenTatarian. The data collection included 105 participants in a seven minutes interaction alone with the robot to investigate behavioral outcomes including but are not limited to distances of the users, speaking time, greetings performed, as well as backchannels. The study showed the extent of which each modality within the multi-modal behavior

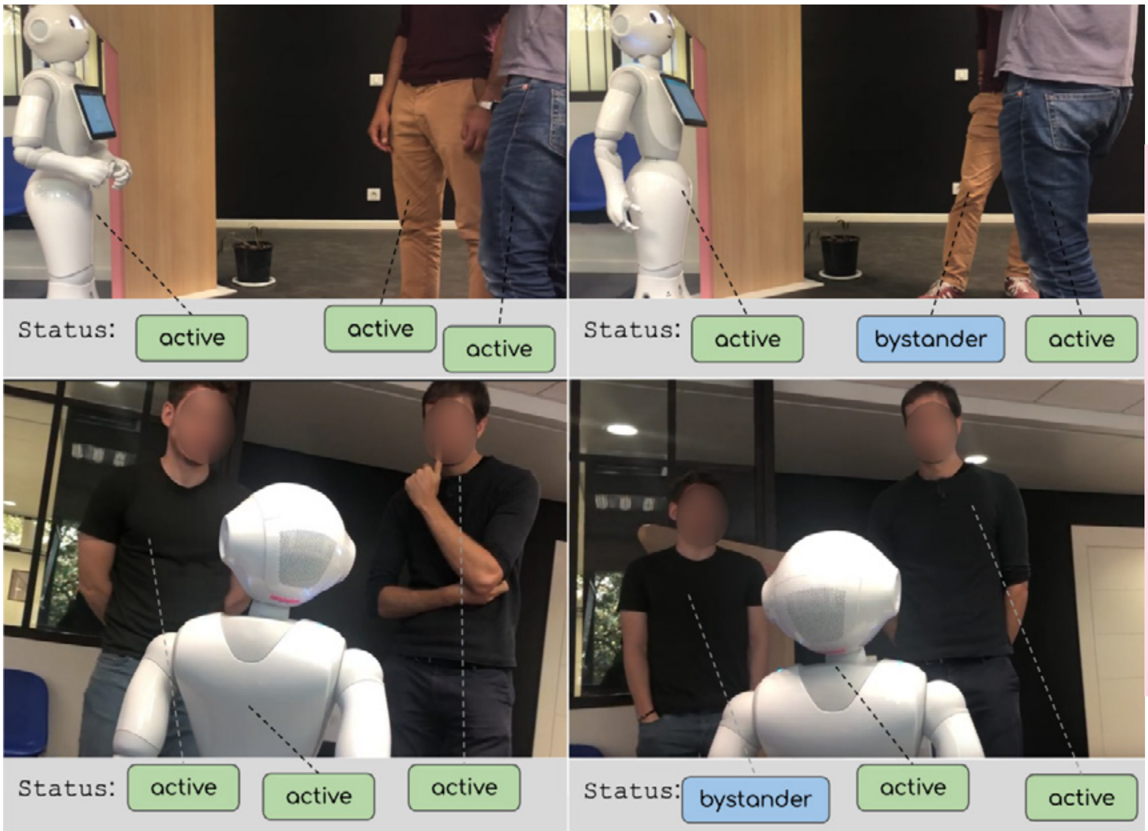


Figure 1: Overview of two sets of interactions between the robot and participants forming a small group. The statuses represent each conversational role for each participant estimated by robot in different contexts and times during the interaction

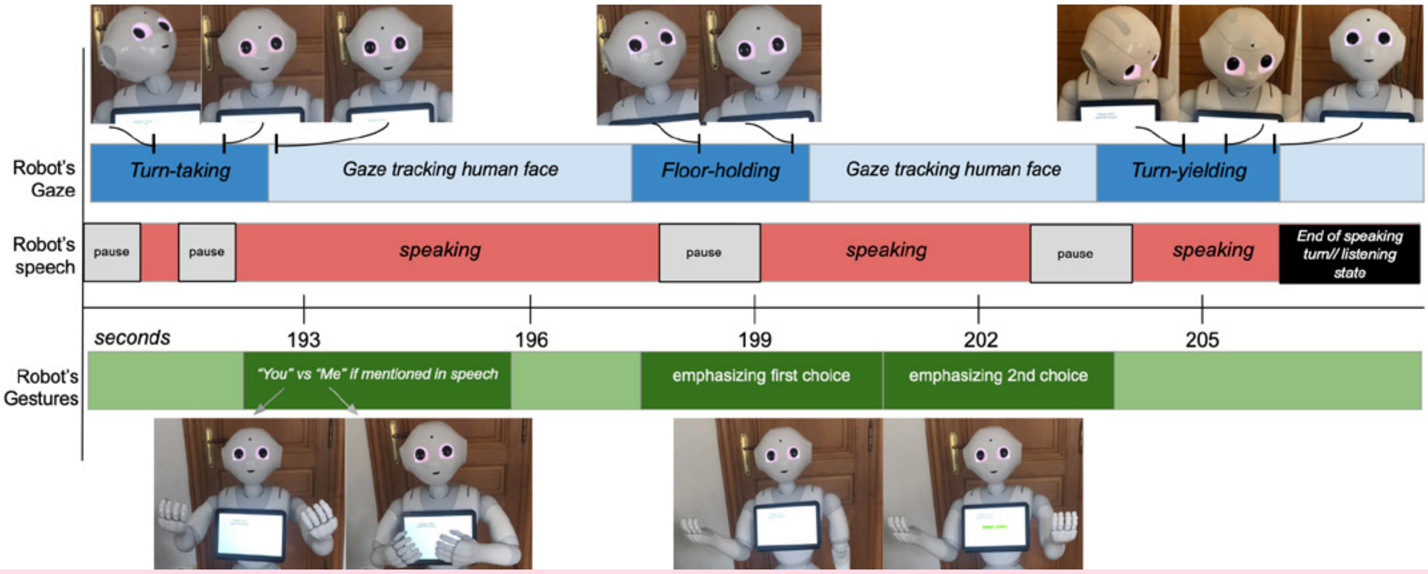


Figure 2: Sample of time-line including speech, gaze mechanisms (turn-taking, floor-holding, turn-yielding), and social gestures (deictic gestures: "You" vs "Me" if mentioned in speech, beat gesture: emphasizing the two choices user needs to select from)

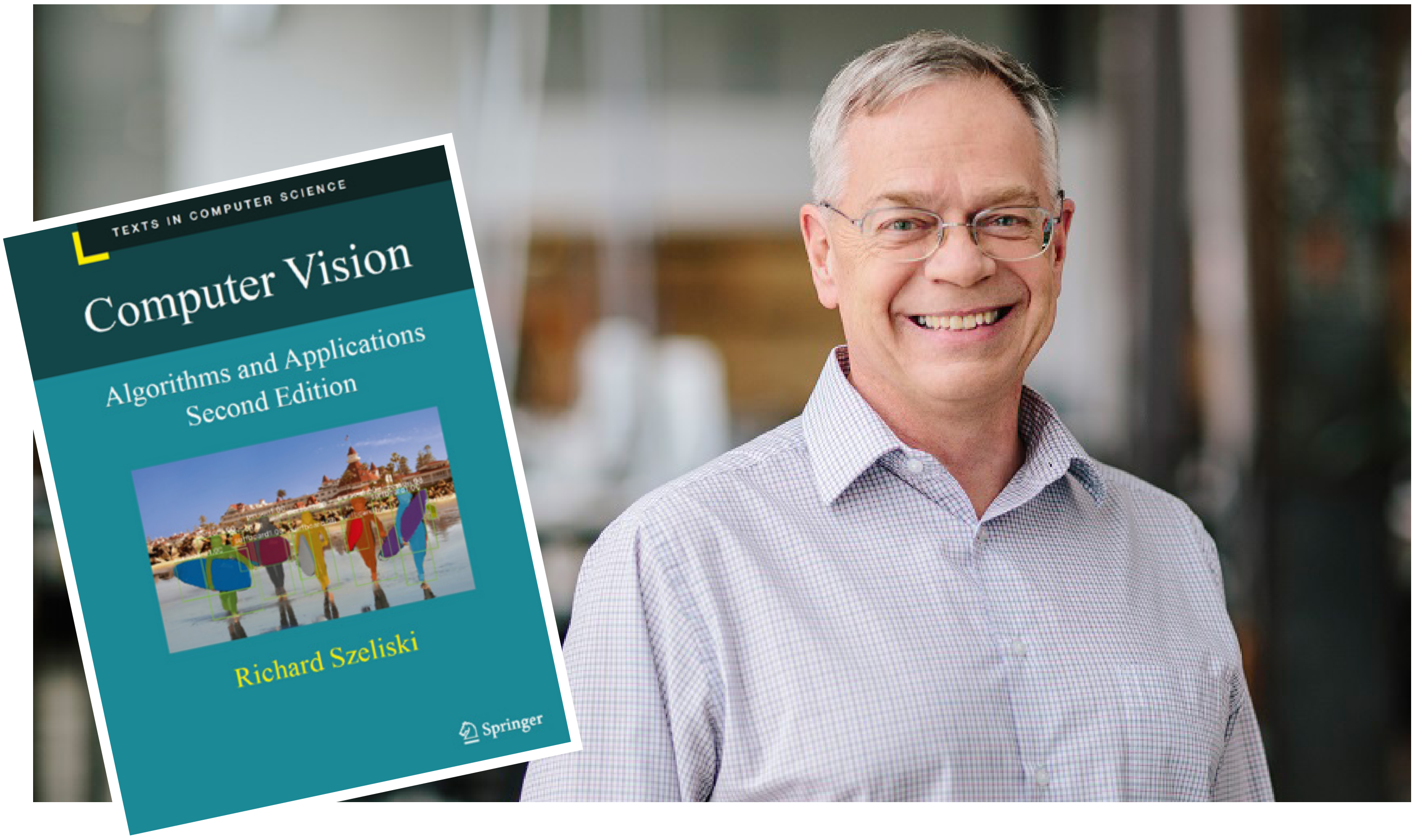
allows the robot to influence how close the human stands, how they address the robot, whether they take its suggestions or not, and how they greet and end the interaction by mirroring nonverbal behaviors of the robot [2].

Adaptation and Personalization for Human-Robot Interaction (HRI):

Human users expect the agents, robots, and technologies they are interacting with to also adapt to them as this in turn improves their usability. In HCI, an adaptive system does not necessarily have the agent learn new behaviors, but rather decide what to adapt, by for instance combining different behaviors, and when or when not to make these modifications. RL may provide a possible solution for achieving adaptation in HRI as a way for the robot to evaluate its behavior. However, as I have mentioned human behaviors and social signals are complex, dynamic, and continuous in nature. Trying to discretize them would lead to large state-spaces and loss of information. In addition, training RL models for HRI is costly and as the COVID-19 pandemic has shown us, it is sometimes extremely challenging.

To answer the latter, I used all rich data-sets collected in previous work to build a simulation set-up and environment for HRI to train RL models for such use cases on. Moreover, to address the adaptation problem using RL, I looked into the multi-modal social signals of the human to formulate the reward signal. It was then used to adapt the robot's multi-modal behavior creating various possible combinations, which are made of gaze, gestures, proxemics, and emotional expressions, **with the goal to increase the robot's social intelligence and influence**. This reward function was designed to reflect the complexity and dynamics that take place during HRI. The results allowed us to further investigate what combinations of modalities making up the robot's behavior would the agents choose. These findings are crucial to unlock **the advancement of the social intelligence of future technologies in adapting to humans, learning from them, and communicating with them beyond just verbal means**.

COMPUTER VISION: ALGORITHMS AND APPLICATIONS



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

Richard, how did you come to write the book?

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

What made you decide to publish a second edition?

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

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

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

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

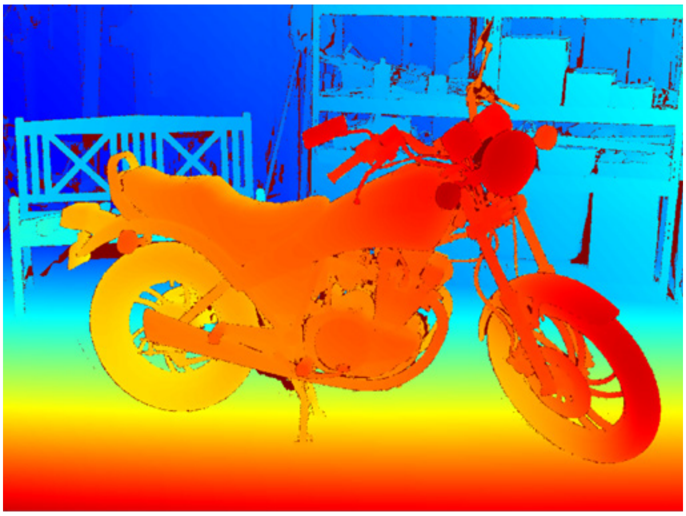
My primary motivation has always been to share information. There was a particular way of teaching the material that I developed with Steve Seitz, and we were both very excited about it. Universities like Carnegie Mellon would take our class notes and slide sets and build their classes around them. I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

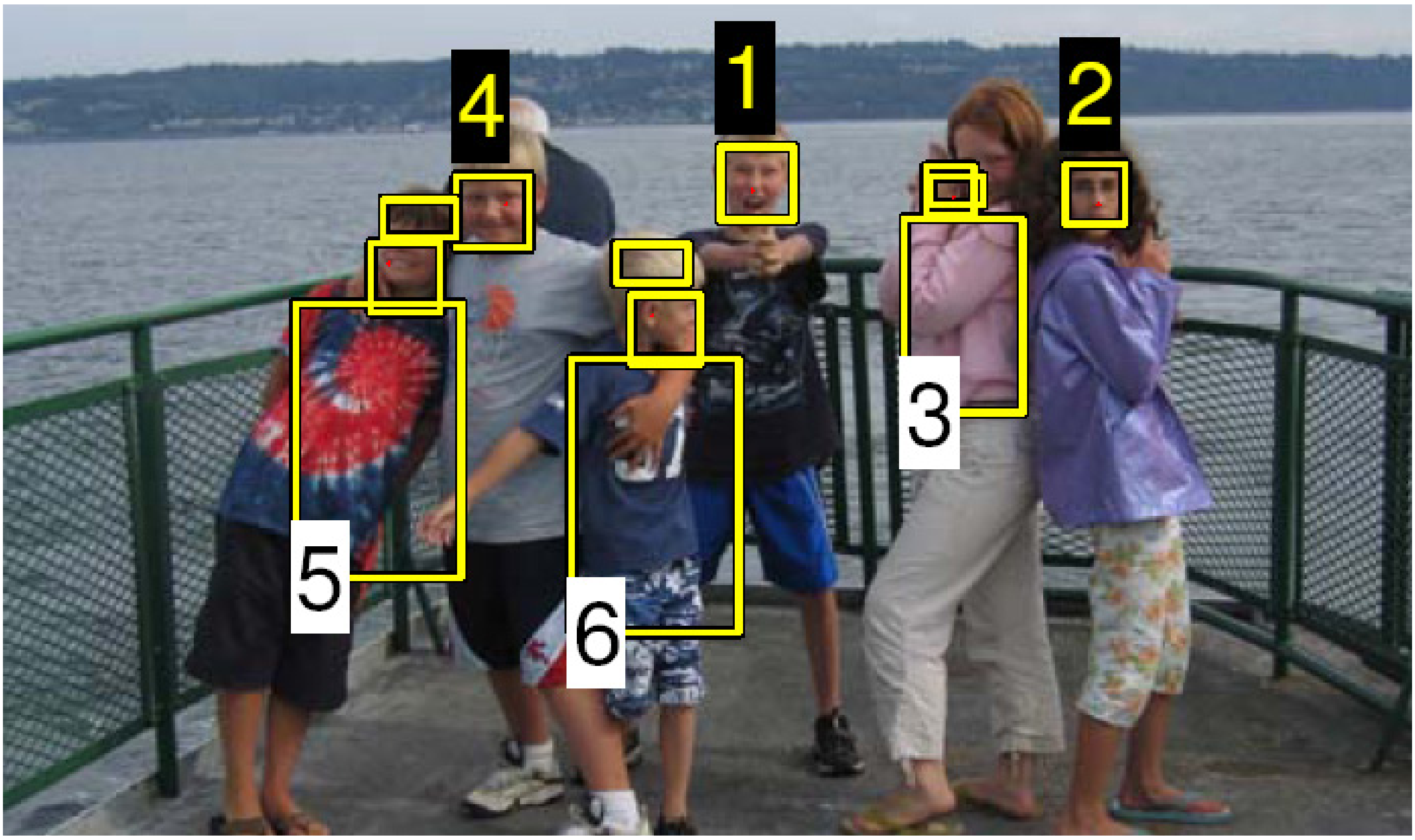
You've had 16,000 downloads already this year!

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

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

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





Have you ever received a suggestion that gave you a real eureka moment?

The help I get is at a variety of levels. I have long-time collaborators like Sing Bing Kang and Daniel Scharstein, who read chapters or co-author things. I was fortunate that in the class I was teaching at the University of Washington, an undergraduate called Matt Deitke would ask excellent questions. It became apparent there were parts of deep learning, such as LSTMs and transformers, that he understood a lot better than I did. He gave me pages worth of comments, and at some point I said, *"Would you be willing to write a section of the book?"* He agreed, and there's a whole chapter on the more advanced topics of deep learning that he wrote. That was my biggest aha moment!

Do you know who the people downloading the book in their thousands are?

It's used as a textbook for classes, so lots are undergraduate and graduate students. It's a reference book that most computer vision engineers and researchers would have on the side. It's just there if they want to know a bit more about a particular topic or which papers to read. It's a perfect starting point for people already doing computer vision who want to get into a subject they don't quite know yet because it's peppered with references and citations.

Do you have any other books in you?

Some students, like undergraduate students, find all the references distracting, so a shorter book with fewer references would make the ideal textbook for them. But I have never had the energy to write it! I'm hoping someday a professor will step up and say, *"Okay, let's do that!"*

Computer vision is such a broad field. Do you have any tips for people new to the community who may be finding it hard to navigate?

If you're a student, one of the best things you can do is sign up for a class at the university or talk to the professor who does computer vision and see if you can do a research internship. Most of us who teach the class believe you can easily shape a complete degree around just computer vision. There are so many different topics. If you want to understand the whole thing, you have to take many classes. There's no way to learn computer vision without putting in a lot of effort. You can take introductory classes, online courses, and do some of the great tutorials in deep learning associated with PyTorch and TensorFlow. Dive Into Deep Learning by Zhang and co-authors at Amazon has inline code snippets. That's an excellent way to teach yourself.

Looking back over your career to date, can you pick one thing that has particularly impressed you?

Boy, that's hard because I would go to conferences most years and come away with pure delight at something that happened! My PhD was in computer vision, but I've always loved computer graphics. The 3D graphics that Pixar produces now came out of people like Ed Catmull, who was publishing papers in the '80s on computer graphics. I was still working on my master's degree at the University of British Columbia, and we drove down to Seattle in 1980, where the SIGGRAPH conference was being held. They showed this short video called Vol Libre, flying around a fractal terrain. It was just amazing because we'd never seen anything like that!

You have spent so much of your career at the University of Washington. What has kept you there so long?

The computer science department and the Graphics and Imaging Lab (GRAIL) are extremely collegial. The professors and

students work so well together. When I moved to Seattle in 1995 to work at Microsoft Research, they had a connection with the university and encouraged us to work with people there and co-supervise students. I met people who became lifelong collaborators, like David Salesin, Steve Seitz, and Brian Curless. For me, the biggest joy in doing computer research has been working with smart people and up-and-coming students. Just that sheer joy of coming out of a meeting with ideas and excitement of things I hadn't thought about before. That's what makes the University of Washington very special.

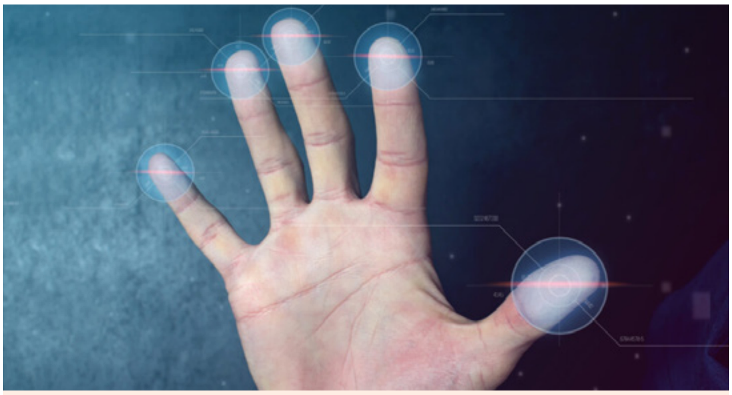
What's your message for the community?

I'm looking forward to seeing all my friends and the younger people doing great research in person once the conferences restart. I've retired now, so I'm in a position where I can look back over the last 40 years, and I've had a wonderful career. The field has exploded. We can do things now that we couldn't back then. The conferences were about 300 people when I started, and now they're over 5,000. Computer vision has become useful in practice. I love it because it combines visual elements with deep mathematics. I'm so excited to share what I know and love about the field with others!

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

AI Unmasks Anonymous Chess Players, Posing Privacy Risks

What if a chess player's game discloses patterns that make it unique, just enough for AI to catch the clues and guess who the player is? This is probably not the worst privacy problem in the world, but combined with many other privacy threats that are growing rapidly, it is something we should probably know about. Researchers from U of Toronto developed a system and tested it by seeing how well it distinguishes one player from another on a chess online platform. Apparently, the system looked for the best match and identified the mystery player 86% of the time! Read More

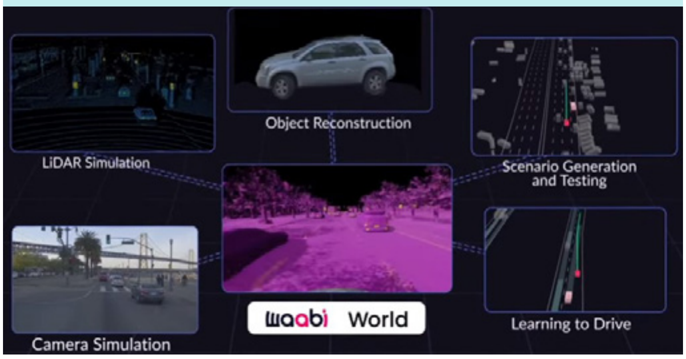


Contactless Fingerprinting Maturation Allows the Unification of Biometric Capture Using Smartphones

Contactless fingerprinting, as a subfield of biometric development, was becoming very popular even before the Covid-19 pandemic started. Of course, the adoption and further development of contactless technologies only **accelerated during the pandemic**. Until recently, for some strange reason, placing your finger(s) on a physical reader was considered the only proper way acquiring or verifying a fingerprint. The development of AI solutions for contactless fingerprint capture might require the gradual replacement of a whole generation of dedicated acquisition devices with **only software and smartphones**. Read More

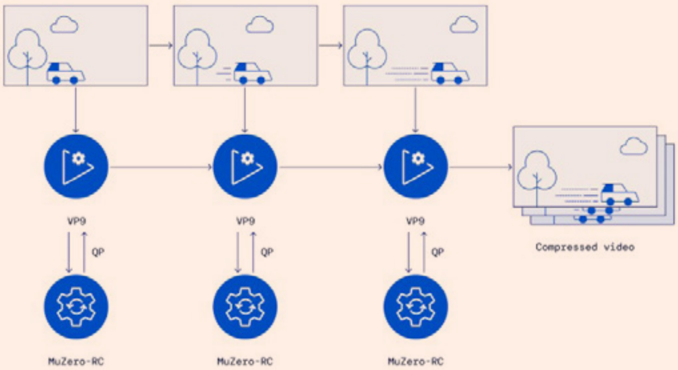
Waabi World: the "Ultimate School for Self-Driving Vehicles"!

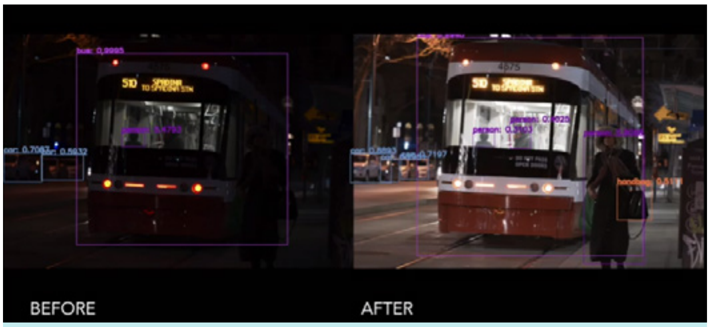
Remember the great interview that Raquel Urtasun gave Computer Vision News only four months ago about her big plans for the **autonomous driving industry**? She said **"We are going to change the world!"** At the head of Waabi, Raquel is not wasting her time: the startup just unveiled a simulator named **Waabi World** that they call "**revolutionary**". In their words, Waabi World is "*the most scalable, highest fidelity closed-loop simulator ever and the key to rapidly and safely commercializing self-driving technology.*" As great fans of Raquel, we bet she's doing it right once again! **Watch the Video**



DeepMind's MuZero: Collaborating with YouTube to Optimize Video Compression

MuZero is DeepMind's successor of famous **AlphaGo** and less famous **AlphaZero**. These AI agents are not only good at winning games: now, in pursuit of **DeepMind**'s mission to solve intelligence, MuZero (which is based on **Reinforcement Learning**) has taken on real-world tasks by optimizing video on **YouTube**, which it compresses by an average 4% with no loss of video quality. Consider the huge time and bandwidth savings that MuZero can generate on the whole activity around zillions of YouTube vids! And you know that YouTube belongs to **Google**, the sister company to **Alphabet**'s DeepMind! **Listen to the Podcast**

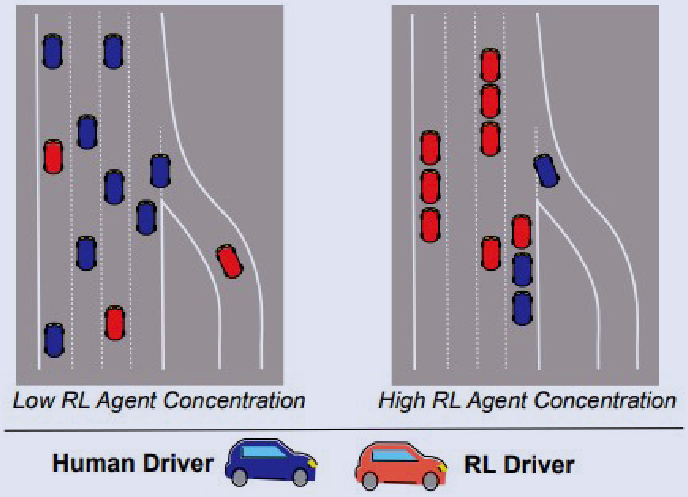




An end to dark, blurry images? Thank AI!

The folks at Visionary.ai have developed an **AI-based image signal processor (ISP)** which achieves cutting-edge image quality in low light. Check the video and see by yourself! Apparently, the startup has also developed a denoiser which removes blur and noise, and they claim to decrease missed detections by over 50% when compared with classical denoisers. Their AI ISP also works in real-time. Visionary.ai say they are already working with consumer electronics manufacturers, and are now starting to enter the medical imaging market to help improve detection accuracy." **Watch the Video**

Choices, Risks, and Reward Reports: Charting Public Policy for Reinforcement Learning Systems



Reinforcement Learning (RL) is considered by many the most promising path to artificial general intelligence. Its use is growing in the real world, not without **concrete dangers**. Imagine an autonomous vehicle looking for the most rewarding (easiest) way to succeed at a task: it may come by optimizing defensive driving in a way that disrupts traffic and makes it less safe for other road users. It's not me saying this: it's a new study just published by the **Center for Long Term Cybersecurity** at **Berkeley**. Luckily enough, the authors recommend a solution: RL models should be accompanied by "**Reward Reports**". Read More

COMPUTER VISION EVENTS

IVRHA 2022 Nashville, Tennessee March 3-4	LSI Emerging MedTech Summit Dana Point, CA March 15-18	SAGES 2022 Denver, CO March 16-19
AAOS 2022 Chicago, IL MEET US THERE 22-26 March	ISBI 2022 Kolkata, India March 28-31	ICLR 2022 Virtual 25-29 April
World Summit AI Americas 2022 Montréal, Canada and online 4-5 May	TechEx North America S.Clara, CA 11-12 May	
Robotics and AI 2022 Prague, Czech Republic 13-14 May		
Int. Conf. and Expo on Robotics and AI London, UK 16-18 May		

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.

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

it takes less than 1 minute!

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!

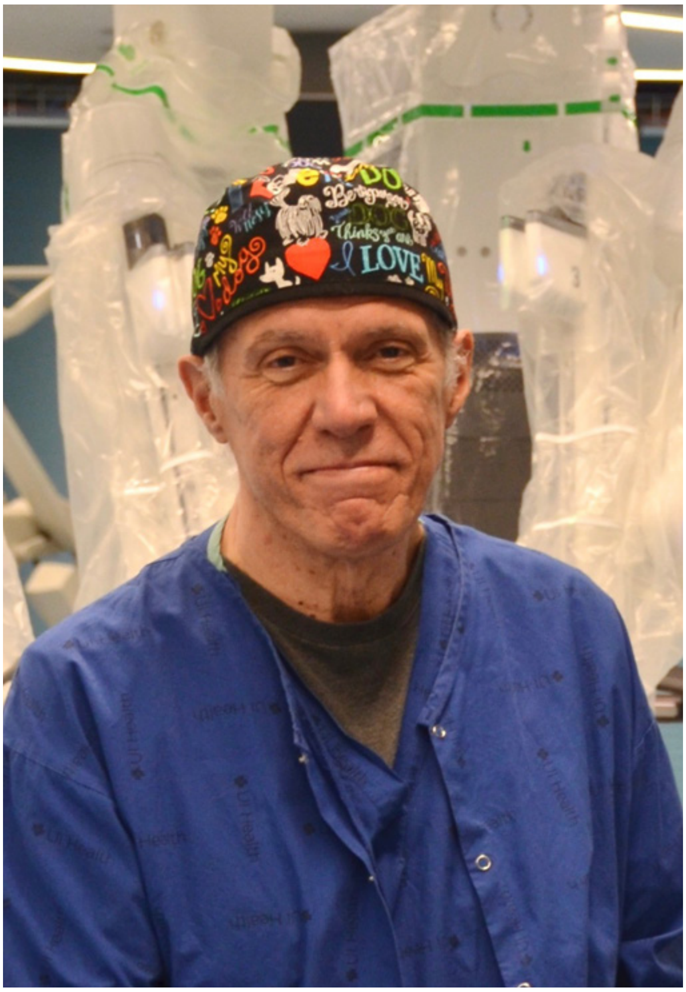
MEDICAL IMAGING NEWS

MARCH 2022

SURGICAL ROBOTICS

EXCLUSIVE INTERVIEW WITH PIER GIULIANOTTI

PAGE 24



Pier Giulianotti is a Full Professor and Chief of General Surgery at the University of Illinois in Chicago, the first center in the world to conduct robotic training for surgical residents. Pier speaks to us about this pioneering work and what the future holds for the field.

Professor Giulianotti, can you tell us about your work?

My work is mainly based on advancing techniques in minimally invasive surgery, including artificial intelligence and robotics. The idea is that we can perform operations better, minimizing complications and improving outcomes in terms of faster recovery and convalescence. It is an ongoing and never-ending process of improving science – a revolutionary step that opens up the field for unlimited possibilities.

What is the current status of minimally invasive surgery?

That is a very broad question. The penetration of minimally invasive surgery in different disciplines and specialties depends on the technical challenges of that specialty, but it is almost 100% in some indications. For example, the number of minimally invasive prostatectomies due to prostate cancer is probably more than 80% now. The concept is more difficult in other fields, such as major abdominal surgery or transplants, so the number of cases performed with a minimally invasive approach is much lower.

What makes some fields more suited to this kind of approach than others?

That is a good question. The prostate, for example, has a particular anatomical location deep into a narrow field into the pelvis. Robotics is the favored application here because the instruments do not need to move around much. For multi-quadrant surgery, which may require working inside all quadrants of the abdomen, instruments must have an extensive range of motion, which is more challenging for robotics.

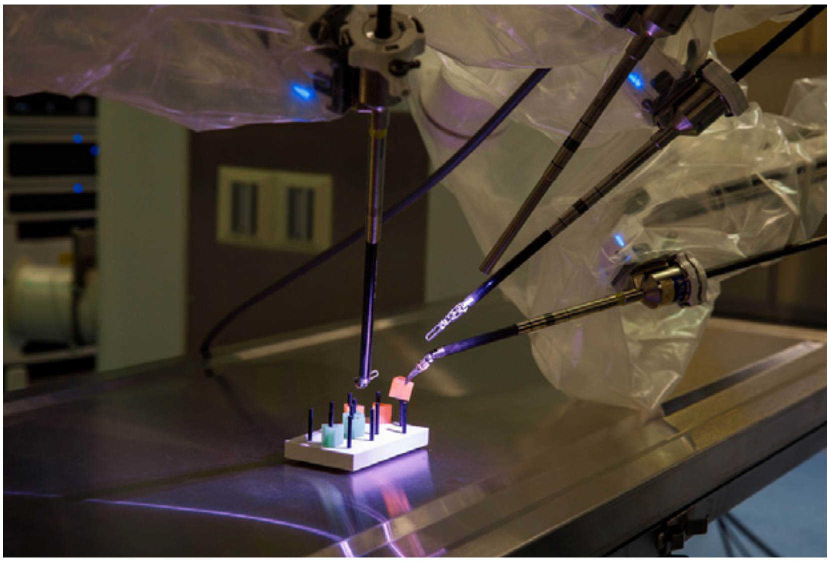
Is this because tracking and navigation among the organs is a real challenge?

There is not yet AI-guided navigation. Tracking and navigation are so far based on human guidance. The human mind recognizes landmarks or anatomy and then addresses the direction of the instruments. In the future, the use of AI will mean we will be able to recognize anatomical landmarks and automatically manage the instruments in the proper location and at the appropriate angles of approach to the target anatomy, avoiding major collisions with anatomical structures.

You have already performed thousands of robotic surgeries – what is the main breakthrough you have noticed since your first one?

It is probably in the range of 5,000 robotic surgeries now. Some of those were major operations; some were smaller. When I got started in August or September 2000 in Grosseto, Italy, the platform was a kind of prototype. We were struggling to make it function. It

"Instead of spending trillions of dollars on developing nuclear bombs, maybe we can spend it on improving the quality of our world to allow people to have a better quality of life, including better surgery and better outcomes!"



was like having a Formula One car still in development! If you have a good pilot, that can compensate for the state of the vehicle. I was a good pilot, so I could compensate for any problems in the surgery. It was in the infancy of the project. Now, the system is much more refined. It is much simpler than it was 20 years ago. It is easier to guide and perform even in a sophisticated operation.

What new developments in AI would make your work easier?

From a mechanical standpoint, the robot is very well refined, and I do not think we can achieve more sophistication in terms of mechanics right now, but we can achieve much more in terms of interaction with AI. The first step is elaborating virtual images to help the surgeon interpret a patient's anatomy better and adopt the right strategy to reach the target and avoid collisions. Think of it like playing chess. The human mind can do very well, and there are probably some champions who could defeat the computer. Still, in general, the computer is much better at playing chess because it can perform millions of operations in a fraction of the time. It is the same for surgery. The computer can calculate in milliseconds the risk of collision and damaging important structures while you are navigating to reach your target. It can tell you, no, this is not the right direction, you have to change approach, or instead of coming from the left to reach a tumor in the middle of the liver, it is better you pass to the right or from the bottom, or the back. AI can do this better than the human mind.

Will you need real-time imaging and great surgical intelligence software to understand what is happening?

Absolutely. The next important development is to overlap and integrate virtual images into real-time surgery. The surgeon can decide to overlap or replace the standard images with the virtual images elaborated by AI. This could help them make critical decisions, such as how close they are to a tumor to ensure the operation is within safe margins while sparing healthy liver or parenchyma.

You founded the International School of Robotic Surgery in Chicago and are very active in the Clinical Robotic Surgery Association (CRSA). What can you tell us about those two organizations?

We have opened up a new lab at the school in Chicago called the Surgical Innovation Training Laboratory (SITL). It is a very modern facility with a lot of equipment and connected with supercomputing centers like the Argonne National Laboratory. We offer training courses and are working on research to apply updated software to improve imaging and navigation.

The CRSA was founded many years ago in Chicago and brings together robotic surgeons who all share the same vision that the future improvement of medicine is connected with the correct application of robotics and AI. They are committed to clinical research, developing the best techniques in multiple operations, and training and teaching younger generations to improve the application in the surgical ward.

That is fascinating! Do you have a final message for the surgical robotics community?

Keep going! Be optimistic that science is making continuous progress in this field. Our operations will be safer and more precise and minimize the suffering, pain, and negative consequences for our patients. We have to believe that. Humanity will define the priorities. Instead of spending trillions of dollars on developing nuclear bombs, maybe we can spend it on improving the quality of our world to allow people to have a better quality of life, including better surgery and better outcomes.

Naronet: Discovery of Tumor Microenvironment Elements from Highly Multiplexed Images

IOANNIS VALASAKIS, KING'S COLLEGE LONDON

@WIZOFE

I hope everyone had an amazing time last month and tried out some coding examples from our last issue! As always, feel free to reach out if you have more ideas, examples, requests. Thank you kindly for sending always nice words and suggestions for our future issues.

Let always be kind to people around us, educate and be patient! Keep up the amazing work you are doing in your life, professional and academic world, and most of all: enjoy it ☺

Review

This month the review article is a very relevant deep learning framework paper, called **NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images**, published in **Medical Image Analysis** by Elsevier and authored by **Daniel Jiménez-Sánchez, Mikel Ariz, Hang Chang, Xavier Matias-Guiu, Carlos E. de Andrea** and **Carlos Ortiz-de-Solórzano**.

We will try to do our best to review and explain to you in much less detail every article we find interesting from a scientific or exploratory perspective. If you want to learn, go deeper, and understand, even more, remember to always check the original references, article, and have a good discussion about it.

Feel free to share our issue with your friends as well, as this may help evolve to new collaborations, conversations, and disputes (scientific disputes are always the best!)

Introduction

The histopathology and phenotype of a tumor guide its diagnosis, prognosis, and help to predict its response to anticancer treatments. Automating these tasks using machine learning (ML) is the goal of a novel field known as computational pathology. For instance, WSDL has been effectively used for tumor subtyping.

SCA emerged in the context of the research for novel cancer biomarkers. SCA methods build topological networks containing cell phenotype interactions. They apply graph-based clustering to assign groups of cells to different neighborhoods. Since SCA methods use the cell as the basic unit of tissue representation, they provide a high level of interpretability.

But you will ask what exactly is this paper trying to do? NaroNet is a **multilevel, interpretable deep learning ensemble**. It learns the most relevant TMEs from **multiplex immunostained tissue sections**. It is a computer model that predicts TMEs and predicts clinically relevant parameters and is based on synthetic sets of multiplex images.

Methods and approaches

NaroNet integrates novel and state-of-the-art ML approaches. The development of **patch contrastive learning (PCL), a self-supervised learning algorithm** that encodes high-dimensional pixel information into enriched patch-embeddings. The paper explores the following important elements and is structured as follows: first, the synthetic and real datasets used are explained, followed by the proposed methodology. The next section contains the experiments used to test the performance of NaroNet and reports the results obtained. The following section provides an in-depth analysis of the proposed methods. Finally, the results, conclusions and future ideas are explored.

Let's see a few of them! First, let's talk about the data.

Synthetic patient cohorts. An in-house developed multiplex immunostained tissue simulator was used to create patient cohorts. Each patient of the cohort was represented by an 800x800 multiplex image.

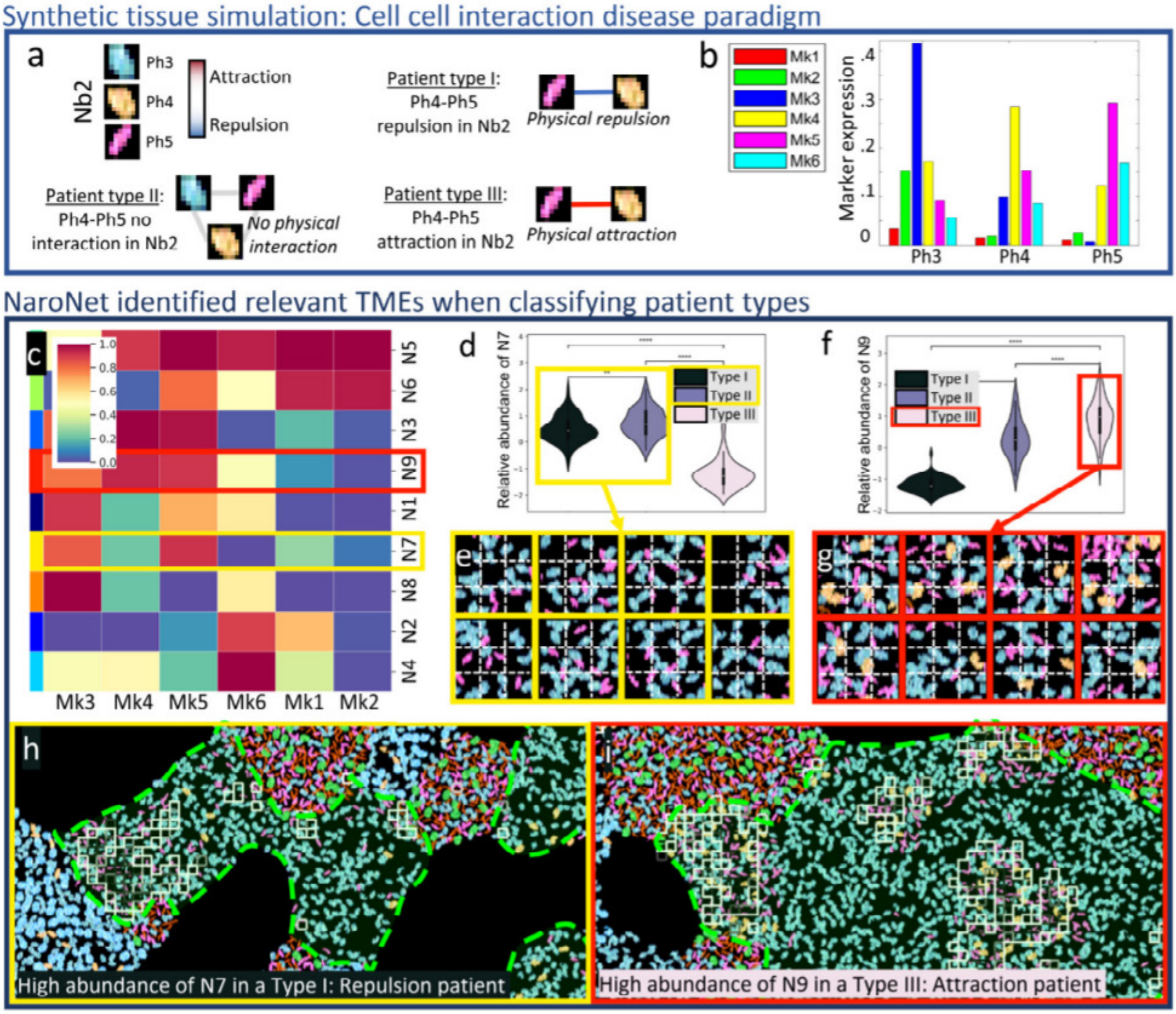
7 patient cohorts were simulated. Each cohort contained 240 patients, distributed in 3 groups (type I, II, and III) of 80 patients each with different disease paradigms inspired by real scenarios. **Phenotype Marker Intensity (PMI)** was used to determine the relative intensity of Mk6 marker expression in each group of patients. Four neighborhoods are defined based on the relative abundance of the phenotypes. Nb3 was set to 15% (moderately present), whereas in PMI2 the relative abundance of Ph6 was set to 0.25% (rarely present).

Phenotype Frequency (PF) was set to 0% (type I), 30% (type II), and 60% (type III) % and in relevant ratios for the other frequencies (you can read in more detail on the paper)

Neighborhood-Neighborhood Interactions (NNI) was designed to simulate different interactions between cellular neighborhoods, related to patient type. Nb2 and Nb3 repel (type I), show no interaction (type II), or attract (type III). Next, tissue sections from twelve high-grade endometrial carcinomas were stained with a seven-color multiplex panel targeting key elements of the immune environment and 336 with a size of 1876x1404x7 pixel images were obtained.

The image dataset is publicly available, and the tissue sections were stained with a 35-plex antibody panel.

Let's see an image describing the experiment!

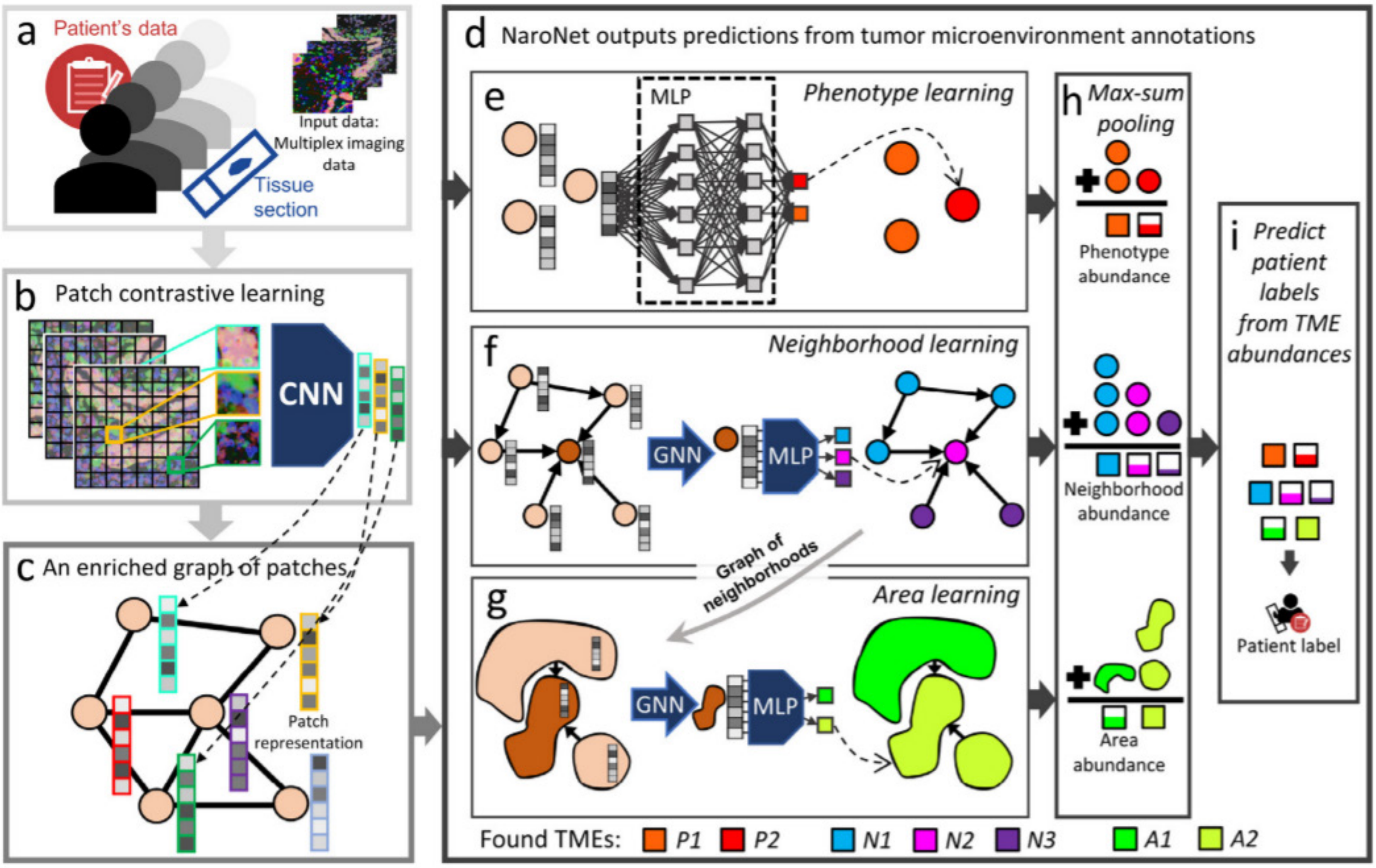


In this figure, you can see the graphical description of the CCI1 experiment. The ground truth on (a) define each patient type which after the processing described earlier leads to classification with the squares showing patches assigned to the learned neighborhood (N9) located in the GT neighborhood Nb2.

The goal of the first step of the pipeline is to convert each high dimensional multiplex image of the cohort into a more manageable list of low-dimensional embedding vectors. To this end, each image is divided into patches - our basic units of representation of the local tissue microenvironment, or phenotype - and each patch is converted by the PCL module.

The PCL module is trained iteratively. In each iteration, the PCL module is unsupervised trained to learn embeddings of a random set of patches. The choice of the image patch size S L is critical as it determines the extent to which biological structures can be captured.

A multi-layer perceptron (MLP) maps each representation to a 128-dimensional vector to create similar embeddings for patches contained in the same crop. A graph is then created that contains all the embedded patches of each tissue/image capturing cellular neighborhoods. This graph is G = (Z, A) where Z is a matrix that contains all the embeddings of the image. A is an adjacency matrix that contains the connectivity between patches.

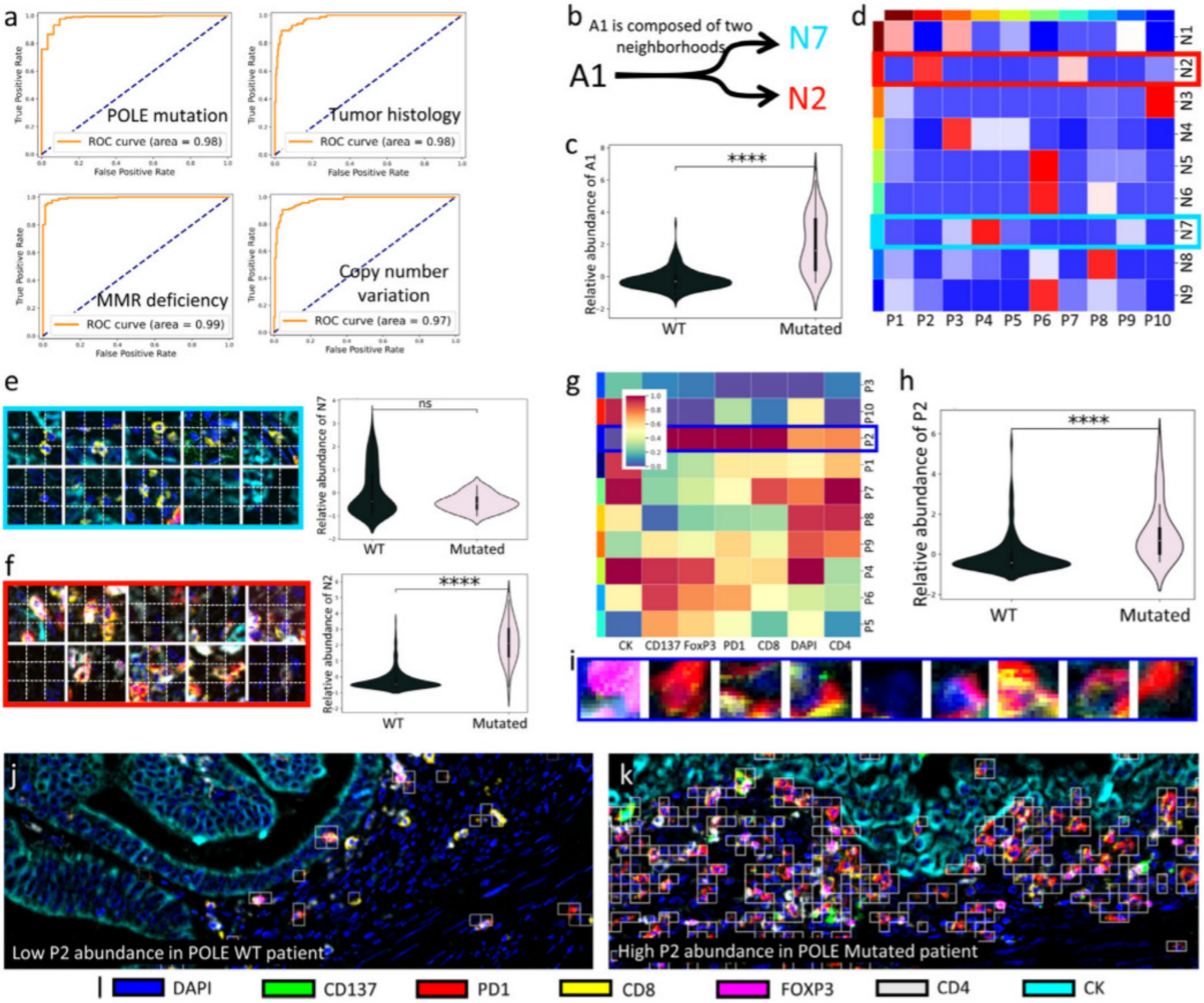


In the previous figure, you can see the process of **training NaroNet**. It learns a mapping that relates patient information with patient labels. The architecture of NaroNet is divided into two consecutive networks G which is an ensemble of three parallel networks that assigns nodes to distinct P, N, A values. The second section, f2, assigns the patient's predictions from the learned TMEs. To learn the specific tumor microenvironment, the three neural networks were used with the previously provided data and later pooled to obtain the abundance of each TME.

Phenotype learning, neighborhood learning (with a graph neural network, GNN) and area learning were utilized as extra elements in the NaroNet.

NaroNet's classification accuracy (and 95% confidence interval) and interpretability were calculated as the intersection of the most relevant extracted TME and the ground truth of each synthetic experiment. The network parameters and architecture variations are optimally selected by an architecture search algorithm. It identifies elements of the tumor landscape that relate to a specific predictive task. The patient's predictions are made solely using the relative abundance of TMEs. The model is evaluated with the entire patient cohort, and new prediction probabilities are obtained. If the null hypothesis is accepted, the extracted TME is considered to have a predictive value.

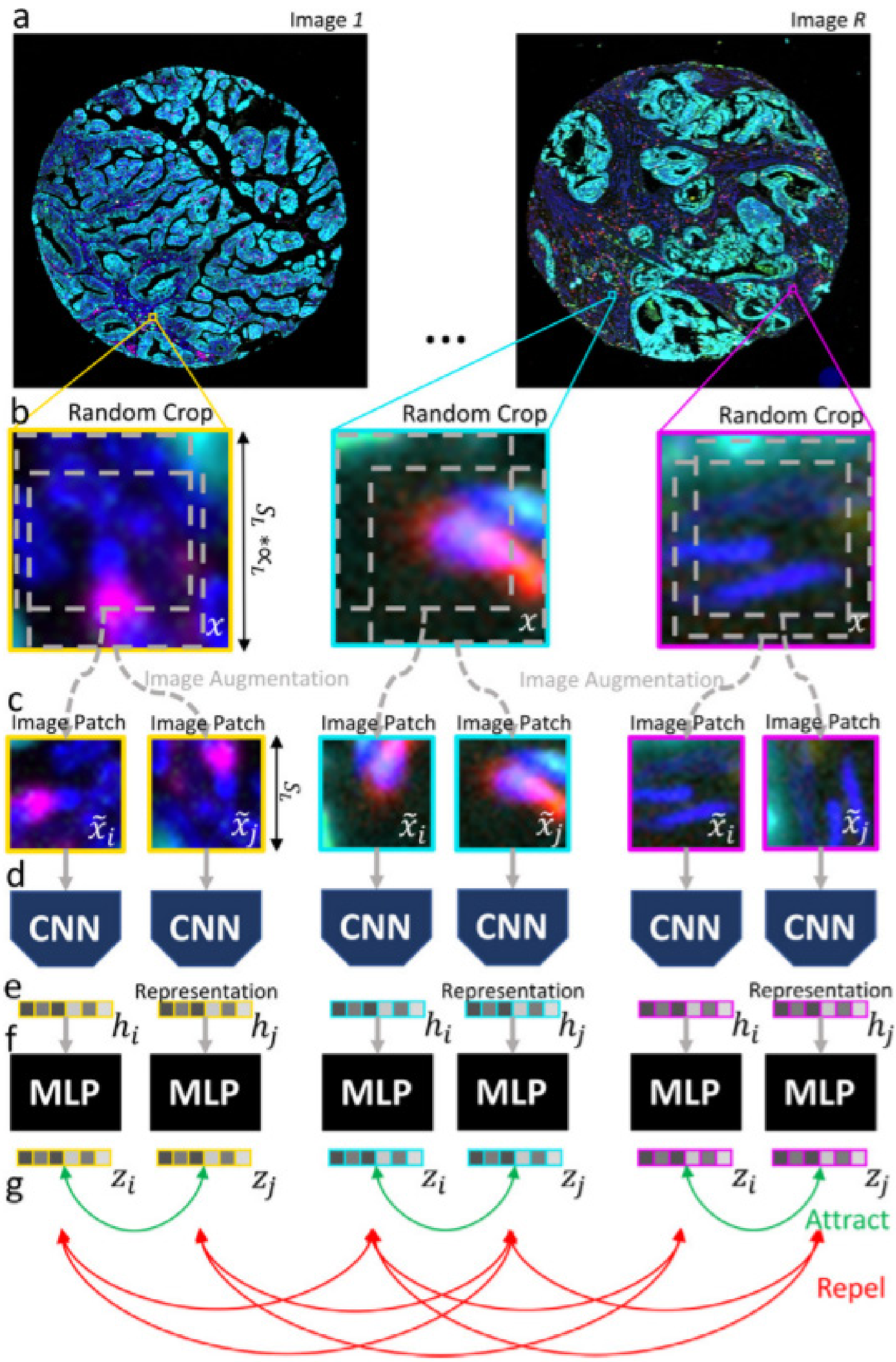
In the following figure, you can see the association of high-grade endometrial carcinomas with patient-level labels. The ROC curves show the classification performance of NaroNet for the four tissue characteristics learned, the neighborhood composition of area A1 and patches assigned to a specific phenotype. Again, see more details of this in the original article!



More synthetic experiments were performed on the network, and it was measured on endometrial carcinomas. The network unsupervised learned 26 TMEs. The area A1 (p-value: $2.56x10^{-9}$) seemed the most predictive TME when making patient predictions. A1 is significantly more associated with tumors harboring POLE mutations. To illustrate the individual interpretability of the results, show two examples of images in which phenotype P2 was the most relevant TME selected by NaroNet.

Summarizing

NaroNet is the first WSDL method fully adapted to multiplex imaging. Two state-of-the-art WSDL methods were used to classify H&E tissue sections, adapted for the analysis of multiple images. ClAM is based on a two-step strategy. In the first step, the image is divided into image patches (i.e., hundreds of cells) which are fed to a ResNet50 pre-trained on ImageNet. In the second step, attention scores are assigned to patch representations. NaroNet as an end-to-end deep learning framework proves this hypothesis true. It accurately performs patient predictions from local phenotypes, neighborhoods, and areas. One of the major bottlenecks in developing high-performance machine learning classifiers for computational pathology is



the low number of available labeled tissue images. A data-efficient contrastive learning loss pre-processing step was proposed and seem to be efficient.

A visualization of the patch contrastive learning is shown on the following image with a step-by-step illustration of the patch contrastive strategy using random crops from the images and CNN which are fed as image patches in a series of CNNs.

NaroNet can learn relevant TMEs local phenotypes, cell-interaction neighborhoods, and neighborhood interaction areas - even when their presence in the tissue is rare.

NaroNet can achieve more accurate predictions while providing inherent interpretability of the reason behind those predictions.

Overall, the network comprises an ensemble of networks that unsupervised identifies and annotates relevant TMEs that drive patient outcomes with clinical predictions which are directly based on the annotations of TMEs.

Let’s catch up

Let’s meet next month with a surprise article! See you soon ☺

TOTAL SHOULDER ARTHROPLASTY (TSA) PLANNING BASED ON SHOULDER MRI

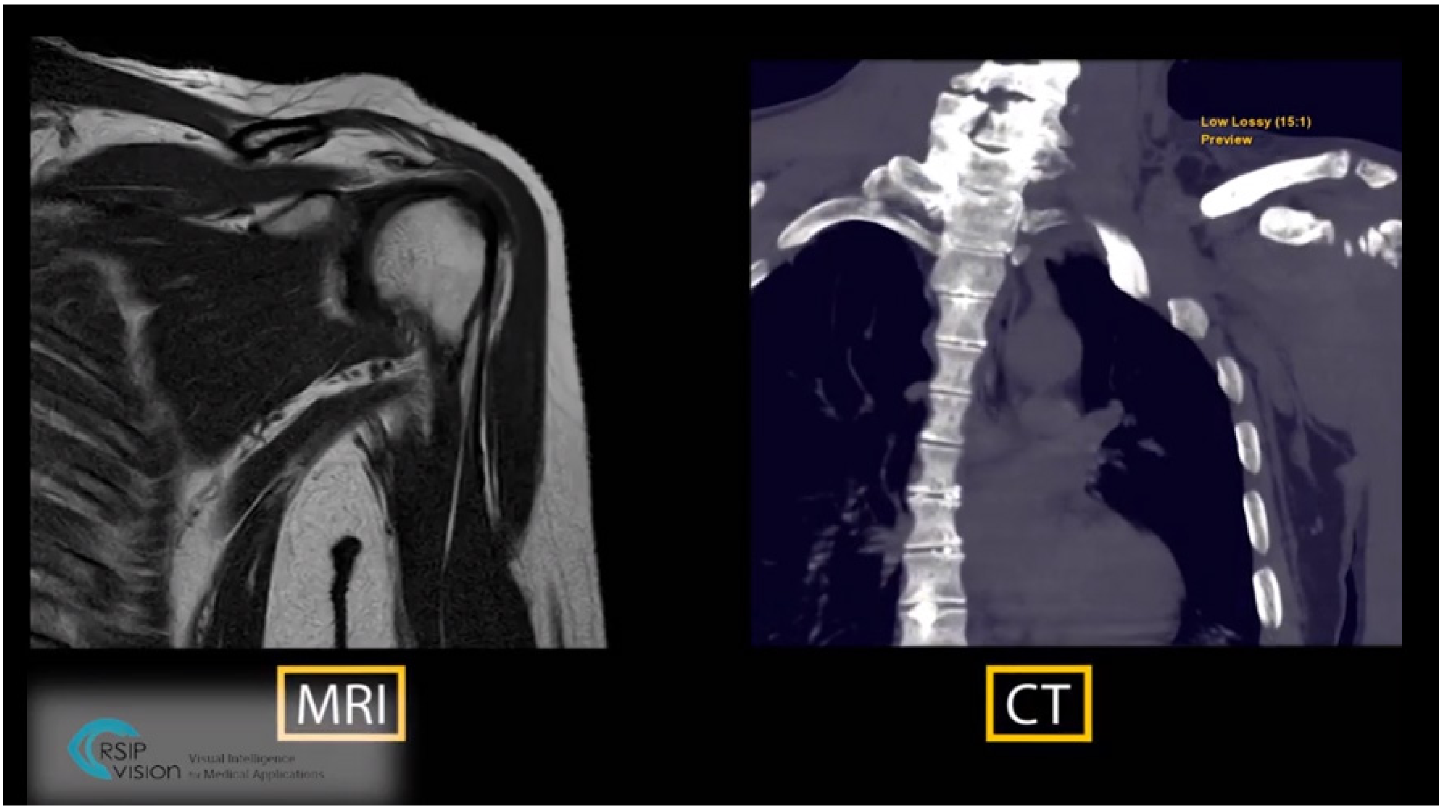
Shoulder pain due to osteoarthritis or trauma is common, specifically in the elderly. To rule out soft tissue damage - tendon, muscle, or labrum tear - a **shoulder MRI scan** is conducted. Once soft-tissue damage is eliminated and bone damage is diagnosed, the physician may decide that **TSA, or shoulder replacement**, is required.

Current practice dictates that for adequate planning, a shoulder CT scan is required. This scan is used to reconstruct **a 3D model of the shoulder bones - the humerus and scapula**. From the model, the surgeon can extract accurate dimensions of the shoulder bone and fit a model of the implant to verify sizing. Additionally, this model can be used to determine whether **standard or reverse TSA** is needed.

RSIP Vision applies **advanced Artificial Intelligence (AI) and computer vision** algorithms to reduce radiation exposure to the patient, time to treatment, and procedural costs.

Our new TSA planning tool accurately segments the shoulder bones - scapula and humerus - from a shoulder MRI scan. This tool is a pre-trained **fully convolutional network (FCN)**, trained on real shoulder MRI scans. The output of this network has one main detriment - it has a relatively low resolution compared to CT-based models and does not allow proper planning.

To overcome this challenge, we developed



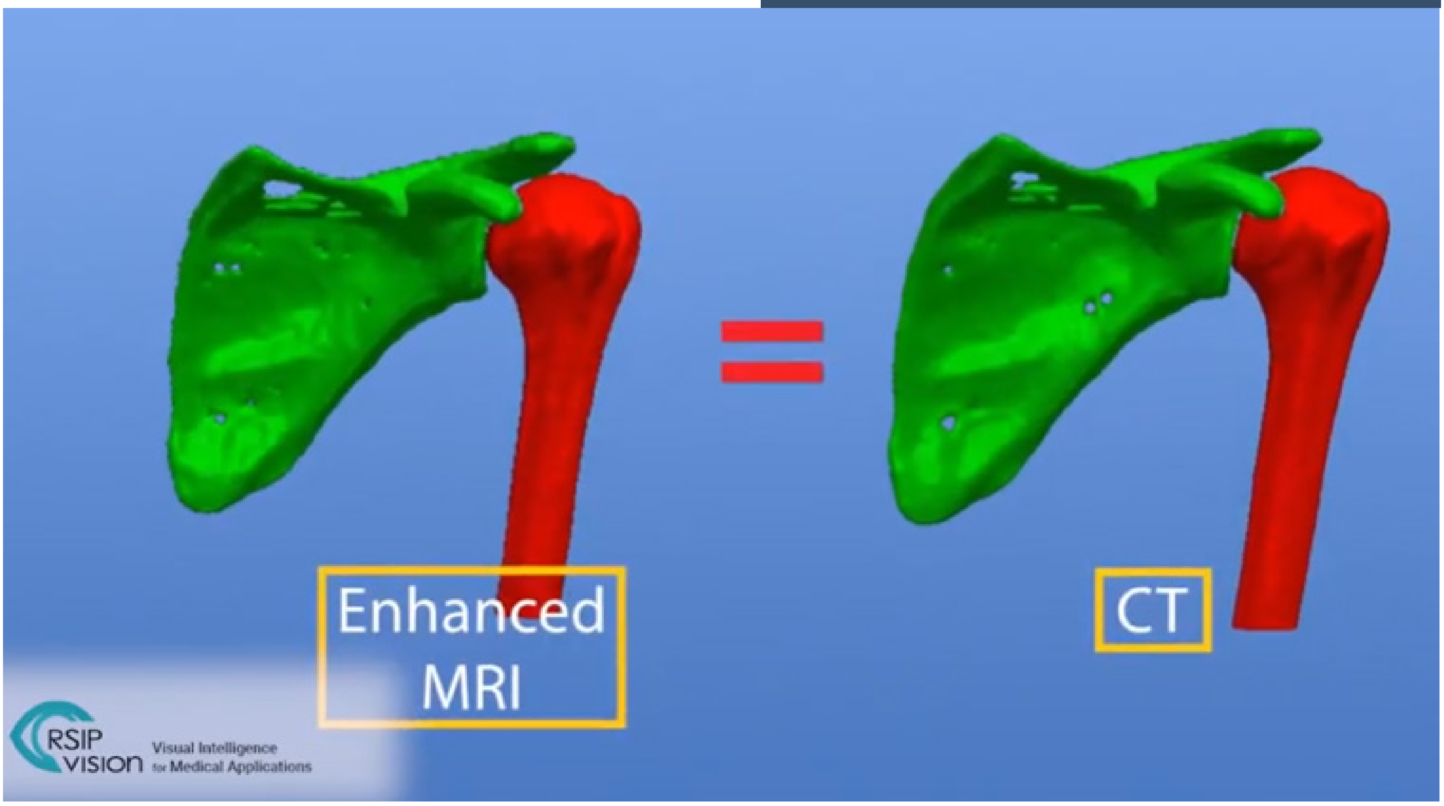
another FCN, trained to enhance the model's resolution. The input to this network is the humerus and scapula segmentation file from the MRI scan, and the output is **an enhanced 3D reconstructed model of the shoulder bones, with superior resolution**.

Both networks are based on **U-Net architecture**. The input to the network is the MRI scan or the output of the previous network - the shoulder segmentation. The output in the final layer is a 3-channels probability map of the same shape, indicating how likely each voxel is to be a part of the background, humerus, or scapula.

This final model's resolution is **equivalent to that of a CT-based model**. It can be used to measure the bones and assess compatibility of specific implants. **This model was created with state of the art AI to obviate the need for a CT scan**. The main benefits of this tool is the ability to use existing imaging modalities rather than acquiring additional scans. Furthermore, a CT scan is contraindicated for some patients, such as pregnant women, and this tool can be used as it does not require ionizing radiation.

Our team of computer-vision experts is currently working on expanding the use of this tool for **other anatomies** (e.g. knee) and **soft tissue** (e.g. tendons). This may change the concept of orthopedic surgical planning - from CT to MRI based planning.

RSIP Vision has a proven track record in computer vision solutions for medical devices, specifically in orthopedic devices. Our talented engineers and clinicians can add the TSA planning tool from MRI to your portfolio, efficiently and swiftly. For more information, contact us today!



Congrats, Doctor!

Mara Graziani recently completed her Ph.D. at the University of Geneva and the University of Applied Sciences of Western Switzerland (Hes-so Valais). She researches interpretability techniques for deep learning models used in healthcare. She will now embrace a new postdoctoral adventure at IBM Research Zurich and the Zurich University of Applied Sciences (ZHAW) trying to apply interpretability on systems biology. Congrats, Doctor Mara!

The models that we work with in **machine learning** are approximations based on assumptions that are never exactly true. Even if wrong, some models may be useful.

I find it rather compelling to understand when a model is useful for the clinical application. In this field, mistakes come at a high cost, and understanding the limitations of a model is fundamental. If pitfalls are uncovered, **new models can be built to be more reliable and trustworthy** than the already existing ones.

Being the end-users of our developments, physicians should be part of the model evaluation process, though they often ignore the processes of feature extraction, model selection and training. If they were explained which features are used by the model to obtain a certain output, **they would then be able to predict eventual failures and unexpected outcomes**.

So can we make Deep Learning models for medical image classification more understandable to physicians? How can this analysis be used to improve model generalization?

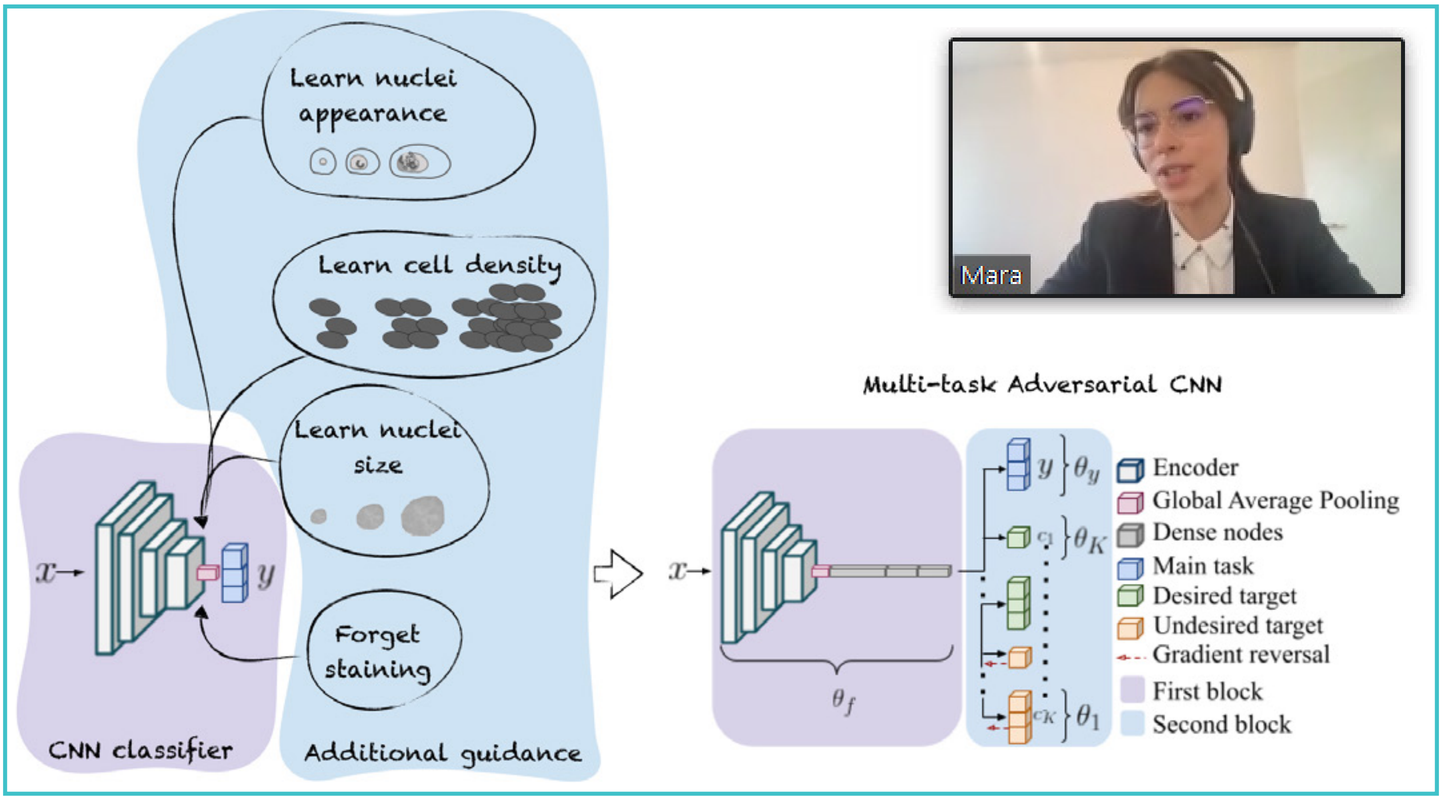
Physician-friendly Explanations

We proposed visual and concept-based explanations that are understandable by pathologists, focusing on the application of breast cancer detection in digital scans of tissue samples. Our work on **Sharp-LIME** uses nuclei contours to generate neat heatmaps that highlight the relevance of individual nuclei to the prediction. With Regression Concept Vectors I have generated explanations in terms of nuclei morphology and appearance.

These works improved the understandability of the explanations by physicians and generated new insights on the CNN mechanics for cancer detection.

Guiding CNNs to learn interpretable features

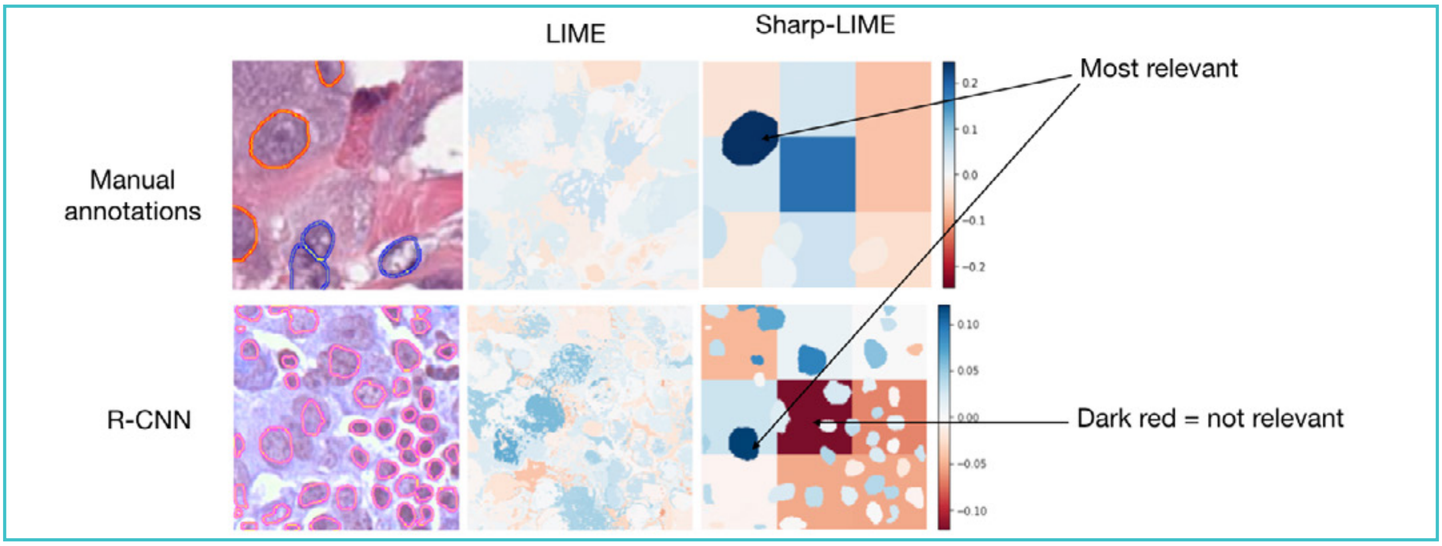
CNNs were shown to be biased towards



learning texture differences rather than shapes and scale. **These biases may affect the generalization** if new scanning protocols or equipment are adopted by the laboratory. To counter this limitation, we developed **a multi-task adversarial architecture that encourages the learning of clinical features**. This model has an additional branch for each clinical feature that is important to predict tumorous regions, e.g. nuclei density, size and appearance. The shift across data from multiple institutions is handled by an additional adversarial branch that removes the domain information from the internal features. **This architecture outperforms conventional CNN-based approaches for the prediction of tumor vs. normal regions in histology slides**.

Acknowledgement

This research was supported by the EU Horizon 2020 projects PROCESS, AI4Media and ExaMode.



Wiro Niessen is a man of many talents. He is a Professor of Biomedical Image Analysis at Erasmus MC in Rotterdam and Delft University of Technology. In the Netherlands, he works for Health-RI, an initiative to facilitate a health data infrastructure to enable optimal use of health data for research and intervention. Our readers also know him as the president of MICCAI from 2016 to 2019. Wiro speaks to us today about Quantib, a leading radiology artificial intelligence company, recently acquired by RadNet, where he is Founder and Scientific Lead.

Quantib was founded in 2012 as a spin-off from Erasmus MC. At the time, Wiro's research group was working on multiple projects, including looking to develop quantitative imaging biomarkers to make radiology more objective. Several companies were interested in their work, and eventually, Quantib was founded with **General Electric Healthcare** as a launching customer.

Quantib's first product analyzed **MR brain images**, quantifying atrophy or white matter lesions to support the diagnosis and prognosis of neurodegenerative diseases such as dementia. It received a very positive response from clinicians.

"It's such a sense of achievement when you get feedback that these tools are helping clinicians in their clinical practice," Wiro says proudly.

Compared to working in a university lab, where people are driven by curiosity, a commercial company needs to **build a product that people want and are willing to pay for**. It must be able to fulfill a proven clinical need and should be able to demonstrate how well it works, that it is of a certain quality, and what its limitations are. In a short space of time, it can go **from research and proof of concept to a product that people use**, with FDA approval and CE marking.

"What we brought from the academic side was the AI knowledge and a lot of talent," Wiro explains.

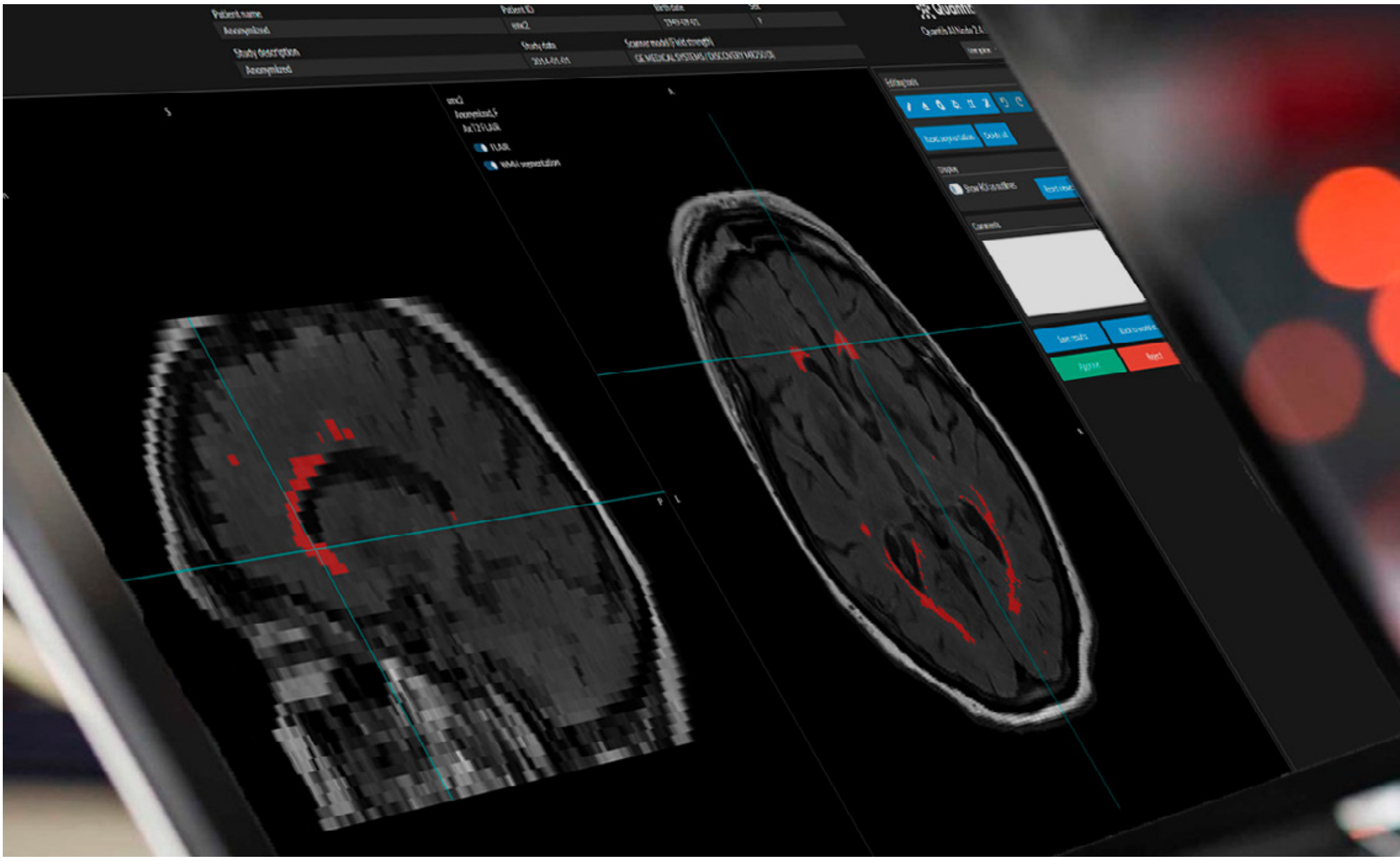
"My research group was already within the hospital and we were well aware of where the potential was for AI in clinical practice, but we could not do it alone.

“It’s a fantastic time to work in this domain!”

When we started Quantib, we hired a new CEO, and we brought in expertise in quality, marketing and sales, and business intelligence. I stayed with the research and development side, and other people took over the business side. ***You have to collaborate with people from different backgrounds, experiences, and cultures to make something like this a success!****”*

Quantib has developed several products over the years. After the MR brain application, it investigated areas of urgent need where technological solutions were still missing. It identified **prostate cancer** as an interesting application. **MRI has a non-invasive role in prostate cancer in monitoring and diagnosing patients and determining which patients need additional diagnostics**. For example,

“… A combination of high-quality imaging and radiologists supported by AI is the way forward for improving cancer diagnostics and screening.”



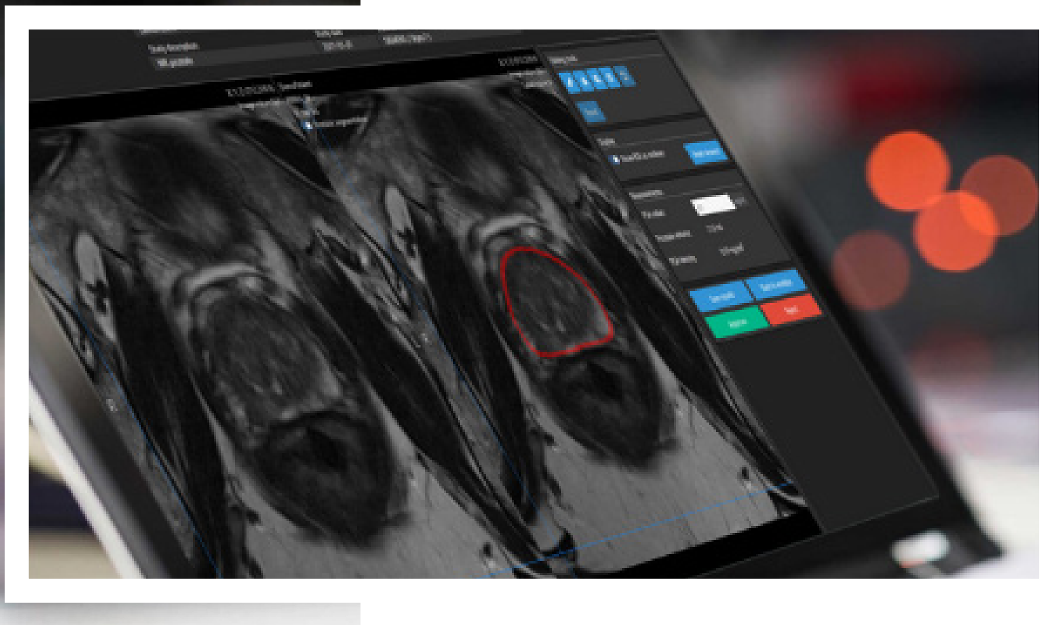
"With such a large user base, we'll have access to much more data!"

some people will need a biopsy, while others will not, and **MRI is invaluable in making those decisions**.

"It's complex imaging data to analyze even for an expert radiologist," Wiro points out.

"We started to develop AI products that could ***assist the radiologist in this analysis****. We got in touch with* ***RadNet*** *as it is the largest provider of radiological services in the US. They were interested in providing* ***high-quality imaging, particularly for breast cancer, lung cancer, and prostate cancer****, and were already active in AI for breast cancer. The idea being that* ***a combination of high-quality imaging and radiologists supported by AI is the way forward for improving cancer diagnostics and screening****."*

RadNet and Quantib have a symbiotic relationship. One is providing imaging services, with access to clinical expertise and plenty of data, but lacking the assistance of AI. The other is an AI company. Together, they can create a proactive rather than reactive approach, with **imaging performed earlier so that disease can be detected early, before it has progressed**, opening up the possibility for better treatments.



“That’s been my dream!”

With Quantib now aligned under the “**RadNet family**”, Wiro anticipates it having a significant impact on people’s lives.

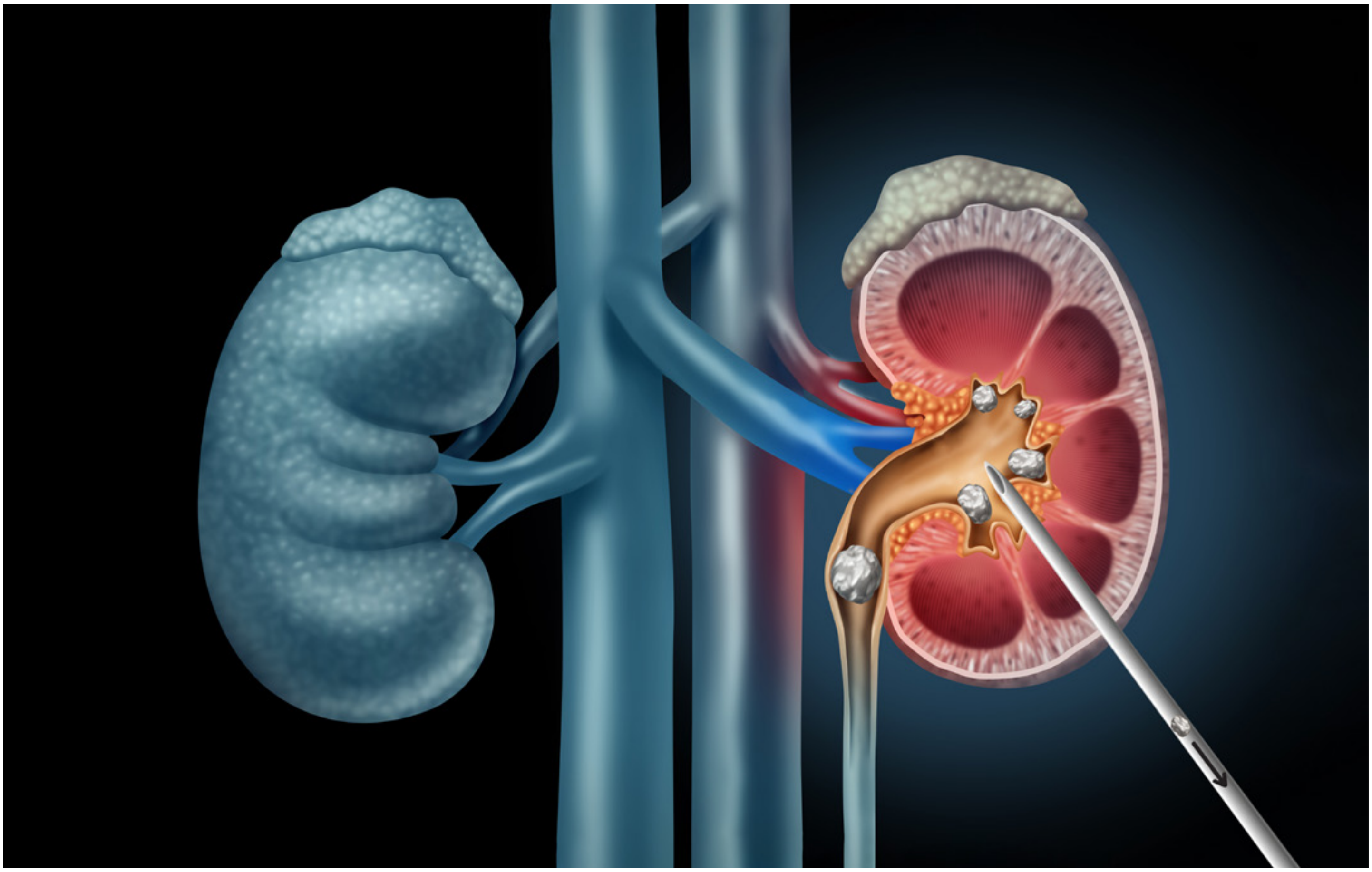
“*We’re going to deliver high-quality imaging for the main cancer types, supported by AI, to reduce the overall burden of disease,*” he affirms.

“***That’s been my dream!*** *The number of people who will profit from the software developed at Quantib will increase enormously because of all the imaging centers run by RadNet. I’m even more excited about continuous product development.* ***With such a large user base, we’ll have access to much more data! We’ll be able to retrain and optimize our algorithms, gather feedback from the clinical end-user, and repeat this cycle over and over to improve the quality of the end product.***”

Quantib is now 40 people with talents ranging from **advanced AI and deep learning** to customer needs and quality.

“*Individually, they’re very good people, but* ***our uniqueness is in the team,***” Wiro smiles.

“*Constant improvements in technology have made it possible to have* ***a massive impact on healthcare****. It’s wonderful to see the community of medical imaging, medical image analysis, and AI impacting the health and healthcare of the future.* ***It’s a fantastic time to work in this domain!***”



PCNL - PLANNING AND REAL-TIME NAVIGATION

Urolithiasis, or kidney stones, is a common pathology affecting nearly 10% of the population in the USA. **Percutaneous Nephrolithotomy (PCNL)** is a minimally invasive procedure intended to remove stones from the kidney and proximal ureter which are less amiable to other endourological modalities, specifically large stones or in the presence of abnormal anatomy. This procedure requires **real-time ultrasound and/or x-ray guidance** and is usually preceded with a **CT scan** for diagnosis.

Gaining access is the first and most challenging part of the procedure. The tract needs to be created in the exact direction and depth of the kidney's collecting system and in close proximity to the stone. This becomes more challenging in cases where multiple stones need to be targeted, and it is preferred to avoid the need for a second tract. Hence, **obtaining a precise tract is essential for a successful and safe procedure**.

To achieve the desired accuracy, the surgeon uses fluoroscopy, ultrasound, or both, to **position the needle in the correct angle** and to verify the needle's depth. Both modalities suffer from inherent disadvantages: ultrasound has poor image quality and fluoroscopy does not allow depth estimation. Another issue is the **radiation exposure with fluoroscopy**. Excess radiation is harmful to the patient

and the operating-room staff as it increases the risk for future disease, so a minimal radiation dosage is required.

To address these challenges, **RSIP Vision presents a set of AI tools**:

Procedure Planning: The pre-op scan contains both high resolution and 3D information. **RSIP Vision**'s dedicated **deep-learning algorithms can detect and segment the kidney stones**, as well as surrounding anatomic structures. Using this output, a detailed surgical plan can be developed - from point of incision, to needle angle and depth penetration, and even further planning of specific methods for stone removal. The latter may include insertion of 3D models of standard needles and lithotripsy and suction tubes to view potential pathways and directions.

Image Registration: A combination of **advanced Artificial Intelligence and classic computer-vision algorithms** are trained to accurately register the fluoroscopy and ultrasound images with the pre-op scan, overlaying the detailed planning map on the real-time images. This tool provides needle position and direction within the 3D information of the pre-op scan, adding a depth dimension in real-time. This method may obviate the use of fluoroscopy, as it maximizes the use of ultrasound, thus reducing radiation exposure during the treatment.

Needle Tracking: State-of-the-art tracking modules follow the needle and provide an accurate estimation of the needle position within the images. This provides **continuous assessment** of the needle position and allows correction if needed.

Combining these tools provides the surgeon with **real-time, precise information about the stone location and the needle position throughout the procedure**, reducing risk of complications and speeding up the surgery. This set of tools can be introduced into **robotic systems**, thus improving PCNL outcomes even more by using AI and the robot's precision.

RSIP Vision has a proven track-record of implementation of segmentation, registration, and tracking algorithms. Our experienced clinicians and engineers collaborate to provide a quick and accurate solution to any clinical need, using advanced AI technology.

Learn more about their AI solutions for urology.

Bettina Baessler is a Professor of Clinical Radiology with a focus on cardiovascular imaging and artificial intelligence as well as the Head of Cardiovascular Imaging section at the University Hospital Würzburg. Recently, she received the Most Effective Radiology Educator award in the EuroMinnies 2022.

Congrats! Awesome job, Bettina!

Bettina, can you tell us about your work?

I have many different aspects of work. My clinical work is focused on cardiovascular imaging. I'm leading the section of cardiovascular imaging here at the radiology department. I also do research. This research also focuses on cardiovascular imaging, but my main focus is on artificial intelligence; not only in cardiac applications but also in many different applications, like oncological imaging, workflow aspects... I'm also involved in teaching activities. I have my own company, besides my work at the hospital, Lernrad GmbH, which also focuses on teaching activities on an online learning platform. I'm introducing many digital teaching formats here and at the university hospital as well.

I see that your activities are very complex. Which part do you prefer?

I like all of them! [*laughs*]

If you like all of them, it almost does not feel like work.

No, actually not. It's a lot of work, but I really enjoy it. And I really enjoy the combination of all these aspects. That's why I'm in academic radiology and not a private practice.

I will change my question: which aspect do you think is the most useful?

[*laughs*] That's a difficult question! Actually all of them. Clinical work is helping people. It is helping our patients. It is introducing value in the radiology chain. Of course, research also tries to add at least some value to the community and to bring new thoughts and new aspects into our lives as radiologists. I really try to solve real problems. I don't like doing research without any clinical meaning behind it.

"I want to change Medicine!"

I always try to solve clinically relevant problems and to translate my research into clinical practice. It doesn't always work but it does sometimes.

What do you do when it doesn't work the way you want?

[*laughs*] Then I try something else! Being creative is really what research is about. Tolerating frustration is maybe the most important aspect of research! [*laughs*]

Is this something taught at school or did you have to figure it out yourself?

Oh, you have to find it by yourself. There is a trial and error principle. You get frustrated and then you have to overcome this. [*laughs*] Try again and try again.

Can you give some expert tips to our readers?

The thing I learned is that it's not about me. If a funding agency rejects my grant proposal, it's not because I am a bad person or a bad researcher. It's because of many, many different potential aspects; political aspects, they didn't like the topic, whatever. There are so many reasons. This is the most important point I learned: I don't have to take it personally.

When did you realize you wanted to become an educator?

Very early - I already started teaching when I was in school. I was teaching others who weren't doing so well in special topics: music, German language, English language... I started doing this very early and during my medical education as well. I was not excited about the teaching at university and I always thought that I have to do it better one day. That's also why I entered academia and worked at a university hospital for my entire career so far... and at different hospitals as well. I then started to teach very early during my residency and overhauled the entire curriculum and introduced digital teaching formats. I was very lucky that I was able to do that, so that as a resident, I had the opportunity to really do what I wanted to do.

"I always try to solve clinically relevant problems and to translate my research into clinical practice."

Do you have any tips for teachers and professors?

What I learned was really to try many different things. Watch out for what is suited for you, because not every teacher is the same. Everyone has their own skill sets and attitudes toward how to teach people. They will be successful in many different ways. For me, it worked nicely to try out many different things. I had feedback from my students on what worked and what maybe didn't work so well. I could condense out of all these many experiments what is the best way for me as a teacher to interact with my students. I'm very interactive. Not all people are like that!

I am sure that you also learn from your students. Can you share one "a-ha!" moment when you learned from your students?

I have to think about that!

We have time.

For sure, I have many of these moments. Usually, students ask so many questions. But some of them may have some fear and don't want to interact: if you get them to the point where they ask questions and interact with you as a teacher, there is always something to learn. There are always different questions, things you haven't thought about yet. I always leave the room with some new thoughts. I can't remember a concrete situation. It's small steps of lifelong learning and thinking differently about things you have taught.

Have you ever had a question from a student without knowing how to answer?

Yes, this happens often. I have no problem saying I don't know. Maybe we can look it up together, or you look it up and tell me! I have no problem admitting that I don't know everything [*laughs*].

Maybe you can teach us something now!

OK! One important aspect of artificial intelligence, to me, is that I don't like calling it *artificial intelligence*. Why is that? That is because we have very narrow artificial intelligence tools, as far as what we are doing in routine with the FDA and CE mark... detecting brain hemorrhage, detecting lung nodules, detecting something... These tools are focused on very specific clinical tasks. This is not very intelligent at all! We could also reflect what intelligence means. This is maybe a philosophical question...

We like philosophical questions!

I don't know what real intelligence is. Why do we call it artificial intelligence when we actually don't know what intelligence means? All these algorithms have been trained on these specific aspects. They are far from being really intelligent as we are as humans, able to interact and integrate so much different information. What I prefer over the term artificial intelligence in the current situation is the term "*augmented intelligence*". What it is doing is augmenting us.

"This is not very intelligent at all!"

The initials are the same, so we can still call it AI...

[*laughs*] Exactly! I never thought about that...

Since we are asking philosophical questions: what is the meaning of life, Bettina?

The meaning of life is to do something that makes life better.

What is the next thing that will make your life better?

I will grow my group here, in Würzburg. I started recently, about five months ago. My group is still in the very early phase. I'd like to grow this group to then make life better and better and better. [*laughs*]

I'm sure you are also making the patients' lives better. That is probably the main goal. Is that true?

Yes, that is true.

Can you think of any specific case where this happened?

Yes! That happens often. Maybe we get a diagnosis that we have never encountered before. For example, we have lots of cardiac patients here. As cardiovascular radiologists, we do some minimally invasive aortic valve implantation pre-procedural planning with CT. We often detect incidental findings on these scans. For example, what we have quite often is Renal Cell Carcinoma on these scans, or at least suspicion of RCC. By detecting this in an incidental way, we can actually save lives. This is something I really like about my job.

I hope that you can save many more in the future! You have done a lot of interesting things in a relatively short time. What makes you run?

[*laughs*] I ask this myself actually! I have so many ideas and interests. I really like to do many different things at once. [*laughs*] I like them so much! It really motivates me. I'm quite efficient.

Can you give us an example of how these different activities converge together?

Yes, they converge! That's very interesting. For example, when I started teaching many people told me, *"Don't get involved with teaching, especially as a woman. Don't enter your career doing teaching because people who do this don't have such a big career..."* In the end, I said to myself, *"Why not teaching?"*. This gets me to where I want to be. I want to be a good teacher because this is how to get to the best people in my field, to get known by the young people. It's fun actually. So I made my way! The interesting point is that teaching is one of the major reasons that I moved to this position. They found that my teaching was so different from other teachers. That's how, in the end, everything finally came together.

You will have a brilliant career. What is the biggest dream that you have for your career?

I want to change Medicine! That's also why I founded the Diversity Commission at the German Röntgen Society, which I am leading because I really want to change the way we are working right now and the way women are present in medicine in general - and especially in radiology. There is much going on right now, so I think it's the right moment to introduce change. That's what I try to do as a leader, to encourage young women in their careers, also to change the way we are working in terms of different aspects, like work-life balance and new work concepts. In radiology, we are so well-suited to working from home. We can actually do our reporting from home. We don't have to sit all day at the office at the university hospital. Currently, medicine is far behind all other movements taking place, in industry for example, where home offices are a hot topic. We are very much behind. I want to push this forward. That is my dream.

I give you now 30 seconds to speak to the world and tell them how to change!

We have to get more flexible. We should leave this strong hierarchical order we are facing in medicine. We have to build diverse teams because we will be so much more efficient! Also, it will be much more fun to

work in a diverse team! This is proven by so much science that we don't have to discuss it. Why hinder women, for example, on their career path because they might get pregnant? Offer them different possibilities to combine work and family. They will be happy and very motivated! They won't drop off the career path as they do right now. This is one major message I want to convey.

We have the same opinion with your message. At RSIP Vision, we make a point not to hinder women. We have this program "Women in Science", in which we interview a different female scientist in every magazine. We are very honored to have you on board this time. What is the next step for you?

I try to build a diverse group. Currently, this is a quite homogenous team at Würzburg. I have an open postdoc position, and I'm trying to find someone who is very different. I'm trying to get there.

Do you accept bald, Italian, non-scientist guys?

[*laughs*] Well, non-scientist could be a problem...

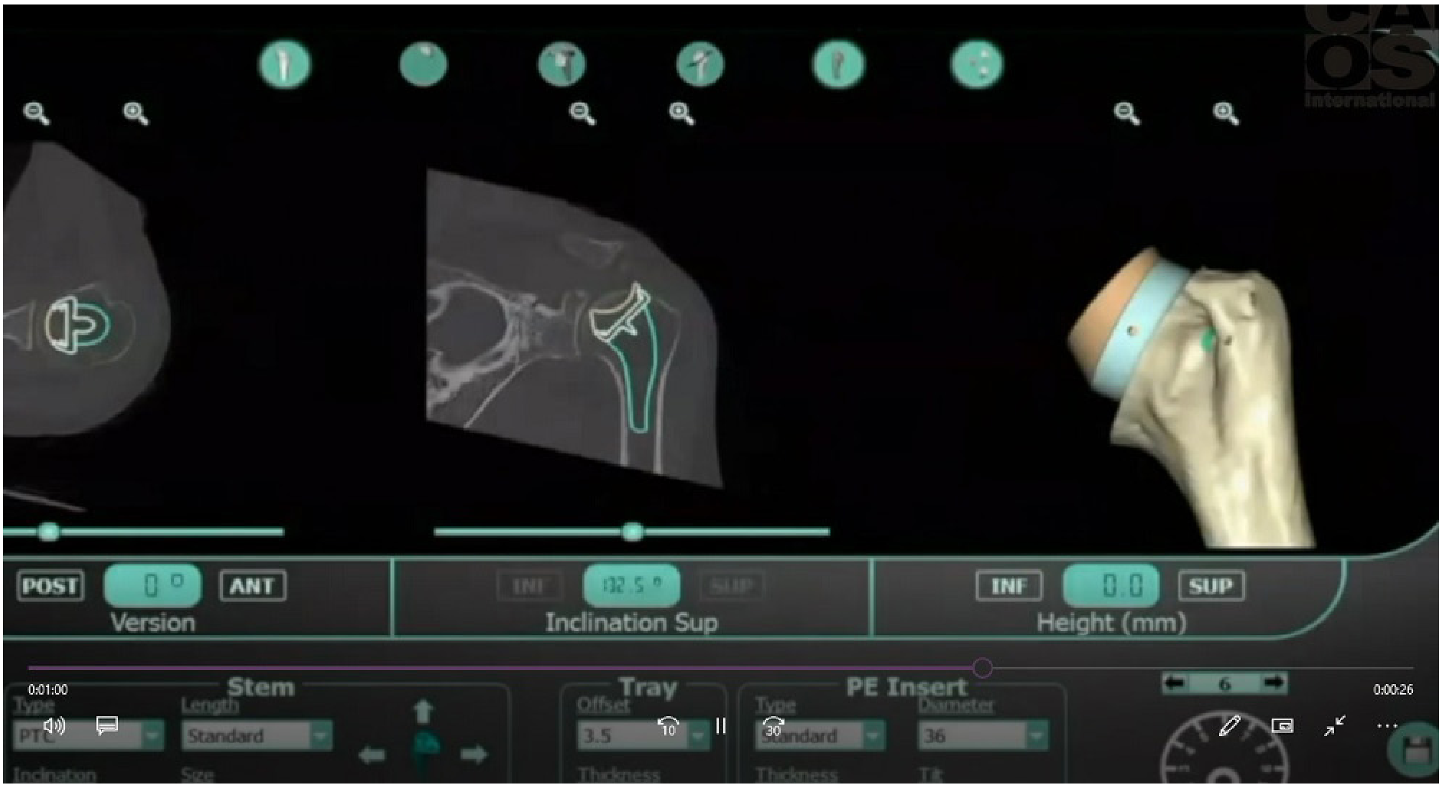
CAOS 2022

International Conference on Computer Assisted Orthopaedic Surgery

Guillaume Dardenne is a Research Engineer in the Laboratory of Medical Information Processing (LaTIM) in Brest, France. With Eric Stindel, an orthopaedic surgeon, and former director of LaTIM, he is co-organizing the upcoming international conference on Computer Assisted Orthopaedic Surgery (CAOS). They both manage the followknee.com project. Guillaume speaks to us about what to expect at the event later this year.

After a brief interlude due to the pandemic, **CAOS** is back, bringing around **300 surgeons, engineers, researchers, and people from industry** together to meet and exchange ideas related to computer-assisted orthopaedic surgery. For its 20th meeting in June this year, it is heading to France for the first time since 2010.

"We have all these different people with different interests coming together in one place to discuss new technology and how it can be applied ***to improve real clinical outcomes in orthopaedics,***" Guillaume tells us.



Several workshops and focus groups will be organized to discuss new and emerging technologies before the main program begins. The conference, held in English, will discuss the latest research and innovations, including **knee arthroplasties, hip solutions**, and more. There will be a technical exhibit for discussions with industry.

Although the event is being held in person, it will make use of hybrid technologies to reach more people. The deadline for submitting papers closed last month. Awards will be presented for the best technical and clinical presentations and posters.

The field of **computer-assisted orthopaedic surgery** has seen many innovations recently, with the development and application of new technologies. As a result, CAOS has a major new focus this year: **data**.

*"With so much new technology, we see different kinds of data which could be used to improve each phase of the **surgical workflow**,"* Guillaume explains. *"These changes will lead to better treatment for patients **preoperatively, intraoperatively, and postoperatively**. And we can better recognize the surgical operating theatre to detect potentially bad gestures to improve the overall surgery."*

With image and language running two parallel paths today, there has been talk of AI tools that can fuse audio and visual data. Is this something Guillaume can see happening?

"Yes, I think this is probably the next generation for surgery and more specifically for orthopaedic surgery," he responds. *"Not only image and language, but there are a lot of other data which could be used to propose treatment better."*

CAOS takes place in **Brest, France**, from 8-11 June, and Guillaume is looking forward to seeing you there.

"We will be delighted to welcome you all!"

NEURO-IMAGING, NEURO-ONCOLOGY, NEURO-SCIENCE SUMMER SCHOOL

Alessandro Crimi is the Research Group Leader of the computer vision and neuroimaging group at the Sano Centre for Computational Medicine in Kraków, Poland. Today, he speaks to us as co-organizer of the first international Neuro-Imaging, Neuro-Oncology, Neuro-Science Summer School (N3S2 for short) on Lipari Island in Italy in July. Alessandro is joined by Anna Maria Trawińska, a PhD in social sciences and Sano's PR and Communications Manager.

Neuroscience is Alessandro's field of research. A broad area covering everything from clinical neuroscience to fundamental topics like conscience and how the human mind is built. The **Neuro Summer School** grew out of a need for the community to reconnect as we all recover from the worst of the pandemic.

"After two years, like many people, I was tired of Zoom meetings," Alessandro tells us.

*"We want to restart and get people working on **neuro-imaging, neuro-oncology** and **neuroscience** together again. We're looking forward to welcoming PhD students, postdocs, and faculty. Industry researchers are welcome too. The idea is to make it as diverse as possible. We'll have lectures, workshops, and tutorials covering topics from neurology to psychology. Geographically, in theory, we're open to everyone."*

Anna Maria adds:

*"**Networking is one of the keywords in neuroscience**. That's why after two years of sitting in our home offices, we're keen to create an interdisciplinary space for people to network and learn together again. We also want people to rest together because the best ideas come when we're feeling fresh."*

The event is a joint venture between **Sano** and the **University of Pennsylvania**, where **Spyridon Bakas** is the lead organizer. As for


Anna Maria Trawińska


Alessandro Crimi

speakers, the team has found a diverse set of well-known names, including "hardcore" neuroscience people like **Karl Friston**, and people from the worlds of neuro-oncology and neuroimaging, more familiar to the communities of **MICCAI** and **CVPR**. Some speakers are famous psychologists who have moved into computer science, so the subject matter will be scientific and technological, and medical imaging content will be high.

"You've seen already that the world is dominated by machine learning, and many models of machine learning come from neuroscience," Alessandro tells us.

"There's an intersection, and this will continue. We're not done. ***There's a good marriage for the future between machine learning and neuroscience****, as well as growing interest in neurotech. Start-ups are working on things controlled by the mind, like video games controlled by VR attached to the brain. Neurotech is one of the next revolutions."*

The team has picked the perfect location for work and play: **Lipari Island in Italy**, one of the Aeolian Islands, and somewhere Italian-born Alessandro used to visit as a child.

"People in Lipari say that even the cats are on vacation there!" he laughs.

"These islands all used to be volcanoes. Some are completely dead; some are still active. There is one that is full of sulphuric geysers. Another island is an exploded volcano, and there is a nice lagoon there where we hope to do a social event. We'll also have a boat trip around a real active volcano. Aside from these very picturesque things, there will be plenty of nice Mediterranean food."

"Don't forget the dolphins!" Anna Maria interjects.

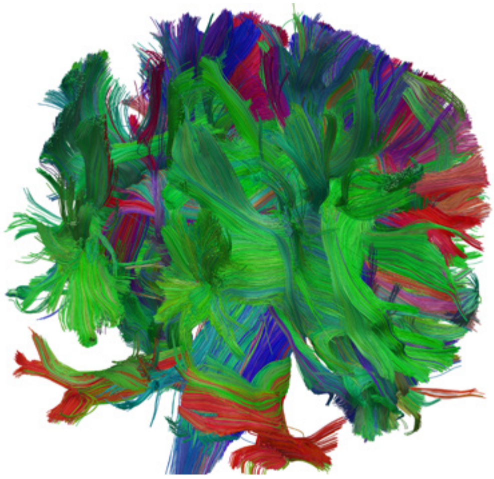
"Yes, some of these islands are not very populated, so there are a lot of dolphins and turtles," Alessandro responds.

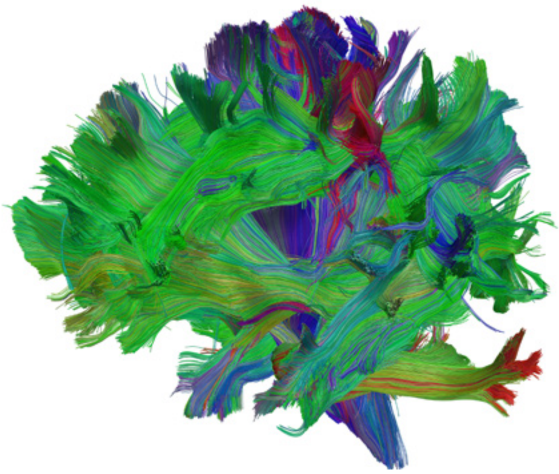
"It's kind of like a wildlife preservation area - they like when people come and visit, but they don't want mass tourism."

Alessandro's neuroimaging group, the **Brain and More Lab**, is one of six research groups at Sano. It is focused mostly on the clinical side, including image analysis for brain tumors, Alzheimer's, and Parkinson's, as well as considering the future of neurotech in everyday life. The team organizes events for MICCAI, including the **Brain Lesion (BrainLes) workshop** and the **Brain Tumor Segmentation (BraTS) Challenge**, which have been running for several years now - as our readers know well.

"We're at this new institute in Kraków which is relatively young," he reveals.

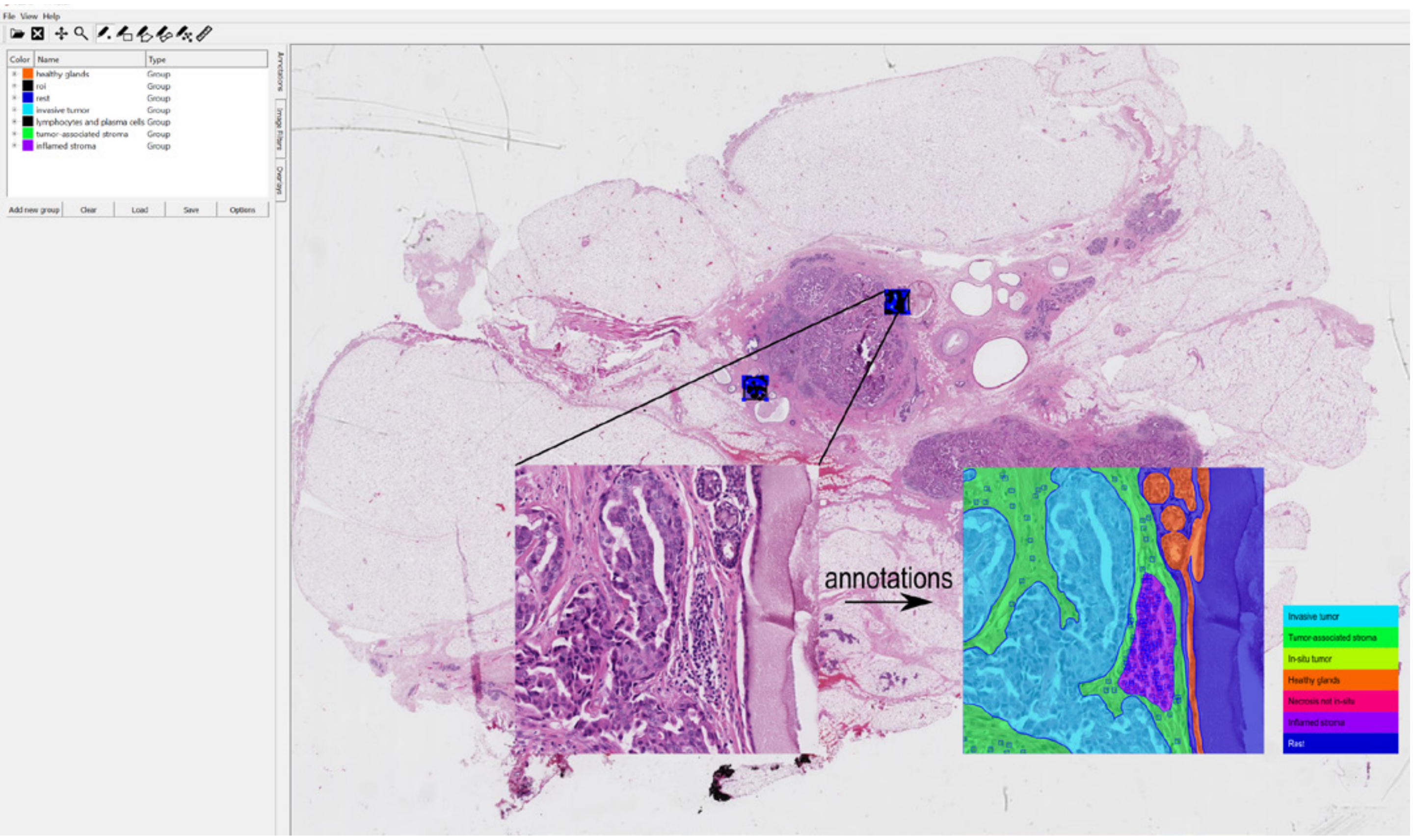
"We're trying to make things as modern and as progressive as possible!"





The Neuro Summer School runs from 4th-9th July 2022.

Francesco Ciompi is Assistant Professor of Computational Pathology in the Department of Pathology at Radboud University Medical Center, and part of the Diagnostic Image Analysis Group. Mart van Rijthoven is a PhD student under his supervision, Witali Aswolinskiy is a postdoc in his group working on whole-slide image classification, and Leslie Tessier is a resident pathologist and PhD student in the same research group, under the supervison of Prof. Jeroen van der Laak. All four are co-organizing the first TIGER (Tumor InfiltratinG lymphocytes in breast cancER) Grand Challenge, which runs until the end of April. They are here to tell us what it is all about.



Breast cancer is the most common form of cancer worldwide and is the leading cause of cancer-related death in women. But **not all breast cancers are the same**. Some patients will have a better prognosis than others. The key question for clinicians is how best to treat each patient to ensure **the most favorable outcome**.

In **classifying breast cancer**, three main tests are carried out for the hormonal receptors, estrogen receptor (**ER**) and progesterone receptor (**PR**), and the human epidermal growth factor receptor (**Her2**).

"If the tumor is positive to any of these tests, it gets a label," Leslie tells us. *"If it's negative to all three tests, it's called a triple negative. **Her2 positive** and **triple-***

***negative** results have the worst prognosis."*

This challenge focuses on a specific target: **tumor-infiltrating lymphocytes (TILs)**. Lymphocytes are immune system cells able to kill tumor cells under certain circumstances. Patients with a higher concentration of lymphocytes in the tumor region seen on histopathology slides tend to have a better prognosis. The way this is determined currently is by pathologists looking at these cells and coming up with **a personal and subjective assessment**. However, different pathologists will come up with different results. Even the same pathologist can make different assessments of the same case over time, so **TILs assessments are not used routinely in clinics**.

There are two main goals of this challenge. One **is to develop AI models to automate TILs assessments**. The second is to **validate this assessment on an extensive test dataset**. Ultimately, the aim is to make TILs assessments **more objective and reproducible** to enable their use more confidently in clinical practice.

The training dataset has been released in two formats – one is oriented towards the computational pathology community, while one is more compatible with the type of work that computer vision people do.

*"We've released all the detections in **COCO format**,"* Mart tells us. *"The COCO format is widely used in many benchmarks to check if detection models work better than other models. Also, we normally train with **huge whole-slide images**, often 100,000 by 100,000 pixels, but on ImageNet and other databases, you generally see smaller images. We have cropped the regions out of these whole-slide images and cropped the masks so that people don't have to deal with these big annotations."*

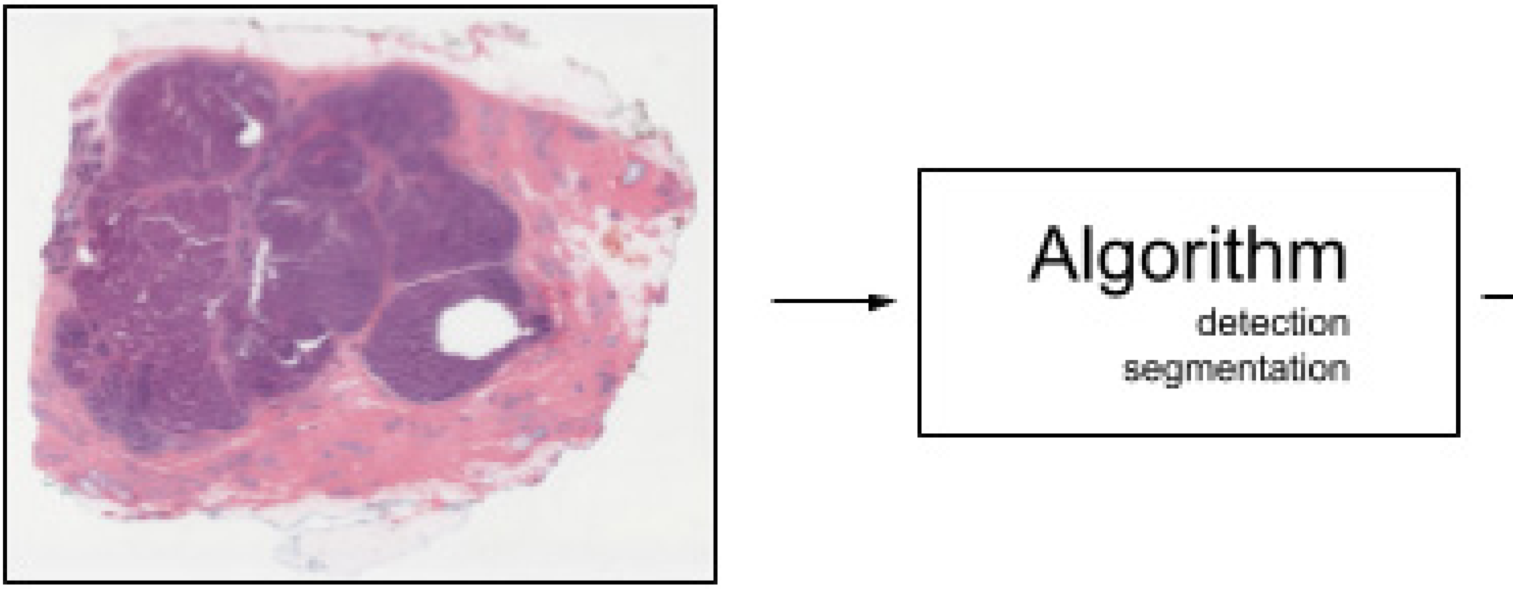
The team created **specific libraries** (e.g., the WholeSlideData Python package) to make it easy


Francesco Ciompi


Leslie Tessier


Mart van Rijthoven


Witali Aswolinskiy



for participants to navigate and connect standard **convolutional neural networks or deep learning** approaches to the data.

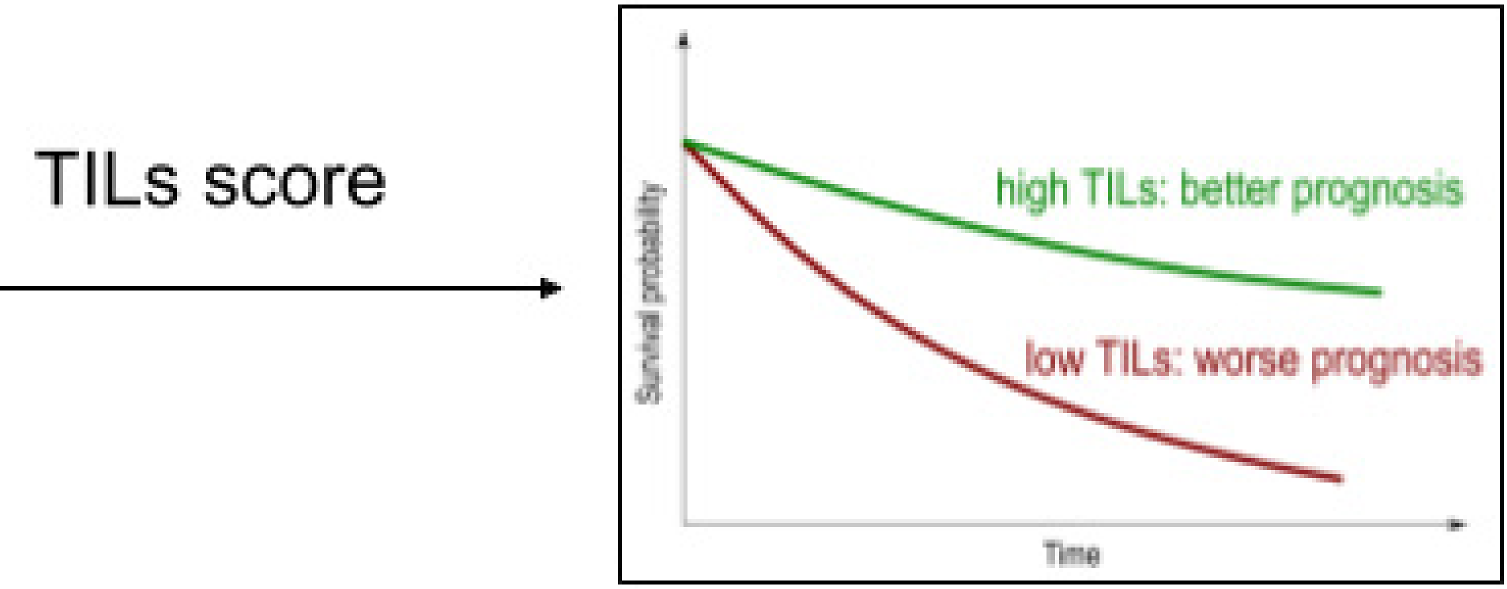
"From the start, we designed this challenge to be inclusive," Francesco points out. *"We wanted scientists and researchers in medical image analysis or computational pathology to take part, but also the wider computer vision community.* ***Clinical application is the kind of goal that we can only reach if we do it collaboratively****!"*

TIGER is a complex challenge with multiple components. Two leaderboards evaluate different aspects of what the participants develop. One looks at the performance of algorithms designed to segment tissue parts, including the tumor and the connective tissue associated with it, and algorithms designed to detect lymphocytes. The second leaderboard considers the TIL score. Participants have to develop ideas for converting what they have segmented and detected into a biomarker. The team will run this biomarker on a large set of independent test slides of breast cancer patients whose outcome is known. Then they will perform a survival analysis to **see if the TIL score can predict recurrence**. Recurrence is the most crucial factor to know clinically in terms of treatment. It is the question that every cancer patient asks their oncologist: **Will the cancer return, or am I cured**?

Until recently, chemotherapy and hormone therapy have been the primary treatments for breast cancer patients. Now, there is growing interest in immunotherapy. Treatment plans might change over time, but either way, **there will likely still be a role for TILs**.

"For me, the goal of this challenge is perfectly in line with my current research work," Witali tells us. *"One of my projects is about* ***trying to predict whether chemotherapy will work for breast cancer patients****, and TILs are a biomarker for that. When I tried to develop and discover novel biomarkers, I ended up with something that correlates with the TILs."*

Francesco, Witali, Mart and Leslie are



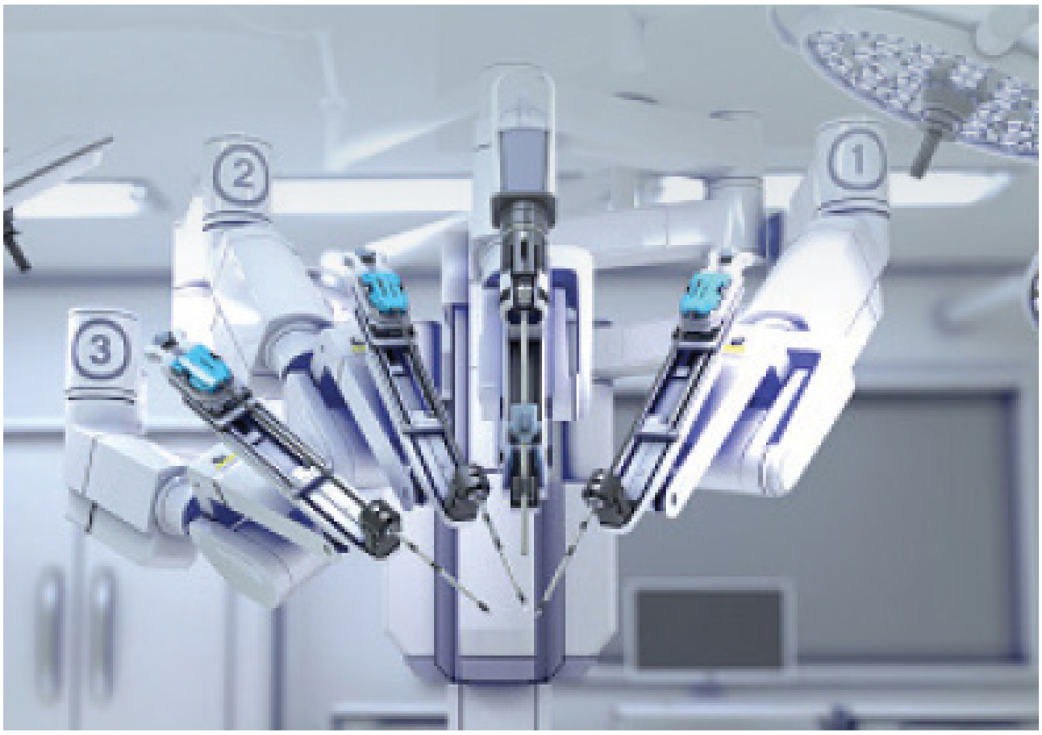
keen to stress that they are part of **a wider multidisciplinary team of organizers**. They are on the technical side, but pathologists have annotated the data, engineers have built and maintained the web-based platform, and multiple clinical centers have provided data.

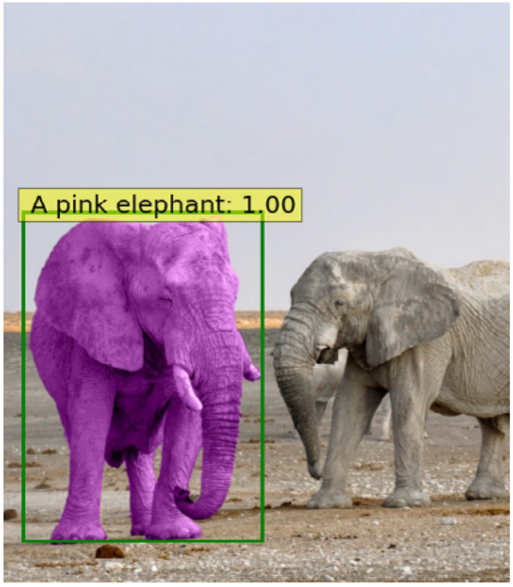
The concept for the challenge originated from a collaboration between Francesco's research group and the **International Immuno-Oncology Working Group**, a multidisciplinary group of clinicians led by **Roberto Salgado**, a pathologist working in Belgium. The TIGER challenge is a close collaboration between Roberto's group, the **Radboud University Medical Centre**, and **Amazon AWS**.

Amazon AWS is sponsoring the challenge by giving access to computing power to run the algorithms and offering a prize for the top three methods in both leaderboards. In total, there is **$13,000** in AWS credits available to the winners.

"This is just the first step," Francesco adds. *"There are more questions we can address in future with follow-up challenges. For us, it's imperative first to show that AI can drive the quantification of TILs, then that TILs have a prognostic value on an independent test set. This challenge is the first step to eventually applying them in clinical practice."*

Visit the challenge website to sign up and download the training dataset.




A pink elephant: 1.00