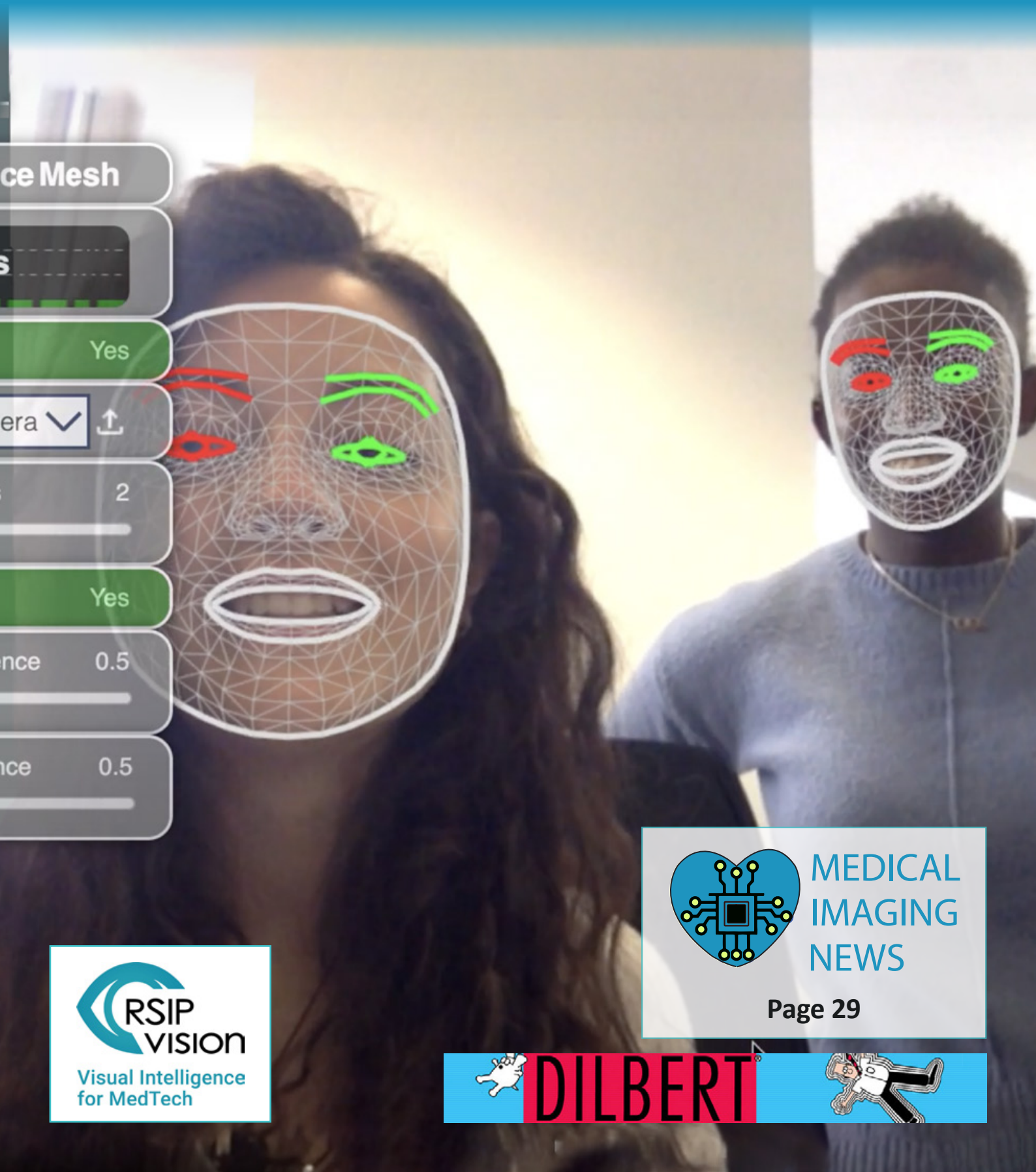
MAY 2022

Computer Vision News

The Magazine of the Algorithm Community



MEDICAL IMAGING NEWS
Page 29

DILBERT

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

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editors:
Marica Muffoletto
Ioannis Valasakis

Publisher:
RSIP Vision

Follow Us

Dear reader,

Welcome to the May issue of **Computer Vision News** and our supplement **Medical Imaging News**!

The 2022 conference season is already well underway, and we are preparing for **CVPR2022** next month, which will be back as an in-person event in **New Orleans**. This month, we are previewing with **Danna Gurari** another of its exciting workshops – **VizWiz** – which will educate and empower researchers to **support the technological needs of people with vision impairments**. Read all about it on page 4.

Turn to page 8 for the second part of **Marica Muffoletto**'s report on 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.

ISBI returned in March with a physical presence in Kolkata, India. On page 30, we have a review of **Best Paper Award winner Vatsala Sharma's work**, which proposes a new system for low-dose PET-CT image enhancement in the case of limited data availability.

Pallavi Tiwari, a brilliant professor at Case Western Reserve University, tells us about her innovative work using **AI and machine learning in the treatment of brain tumors**. Learn how it is helping patients avoid unnecessary surgery in our interview on page 34.

Also, this month, **Ioannis Valasakis** has reviewed some awesome research on **Vision Transformers**, which introduces readers to what they are and how they are being used as a state-of-the-art approach in **Medical Computer Vision**. He discusses the pros and cons in depth on page 40 .

We hope you enjoy another busy issue of Computer Vision News – please share our link with your friends and colleagues so they can subscribe for free!

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

Computer Vision News

Medical Imaging News

2022 VIZWIZ GRAND CHALLENGE WORKSHOP

Danna Gurari is an Assistant Professor at the University of Colorado Boulder and the co-organizer of the VizWiz Grand Challenge Workshop. VizWiz has the dual aims of educating researchers about the technological needs of people with vision impairments and empowering them to improve algorithms to meet these needs. Danna speaks to us about the challenge and upcoming workshop.

People with visual impairments use applications on their mobile phones or smart glasses to learn more about their visual surroundings. **Does this shirt match my pants? Is this yogurt strawberry-flavored? What type of pills are these?** These applications are primarily human-based, but this workshop and the challenges running now aim to get the AI community involved in designing algorithms that will automatically help blind or partially sighted people in their day-to-day lives.

Now in its fourth edition, **the VizWiz workshop** was focused initially on the immediate task of visual question answering but has evolved over the years.

"*We discovered there are many other problems entangled in solving the problem of answering a visual question,*" Danna explains.

"*Problems such as helping the photographer capture a good image of the content of interest. If you want to know the expiration date of food, for example, the food needs to be oriented to see the expiration date. Given that it can be difficult for people to take a good image, we're now putting out video-based challenges.*"

Other use cases include providing a description of images and predicting when and why answers to visual questions might differ.

The datasets generated by the challenges so far have caught the attention of some

big industry players who are keen to use them to help shape their products to meet customers' needs, including **Facebook**, **Google**, **Microsoft**, **Amazon**, and **IBM**. Is this the dream scenario of scientific work having a practical impact on real people in the real world?

"*That's the hope!*" Danna asserts.

"*Another goal for us is ensuring researchers are launched from academia into industry with the right skills to continue the progress of developing these technologies. It's about helping* ***to launch a whole new generation of developers****. Every person that participates makes a difference.*"

One surprising aspect of this work has been how much **private information** is included in the reviewed data, from credit cards to passports and even pregnancy tests. People are deciding **the trade-off to share and learn about private information is more important than not sharing it**. This discovery has created a new branch of research for Danna, figuring

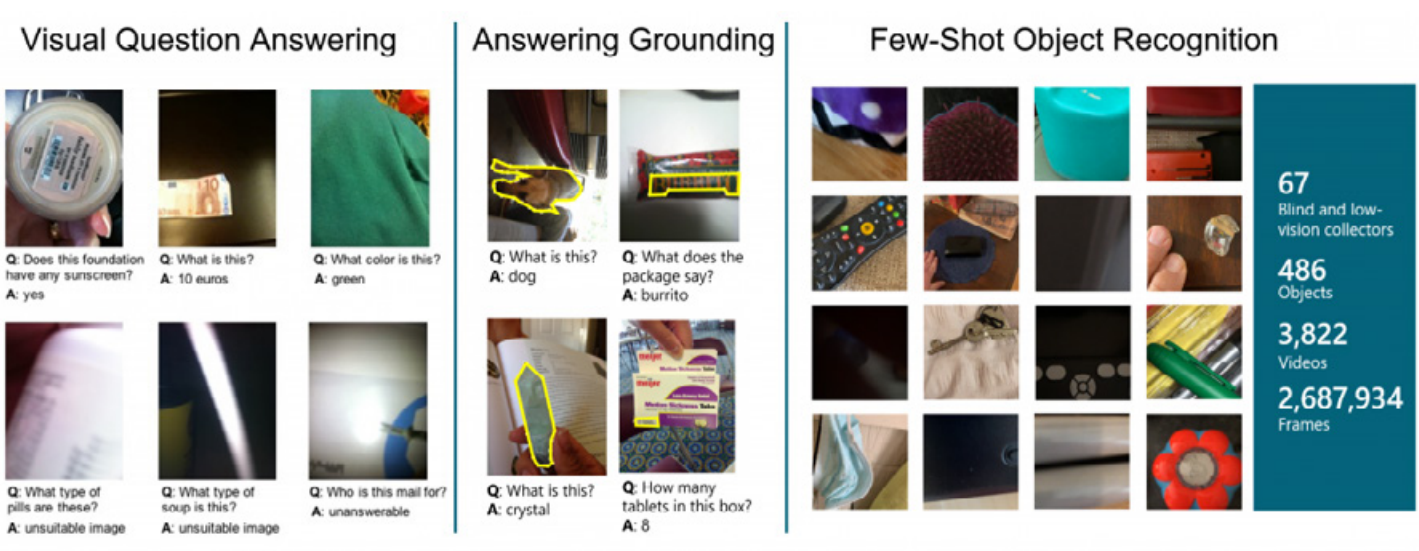
"People are deciding the trade-off to share and learn about private information is more important than not sharing it!"

out how to design algorithms that analyze private information **so that people don't have to share it with a remote human**.

"*You need to understand each individual's privacy concerns,*" she tells us.

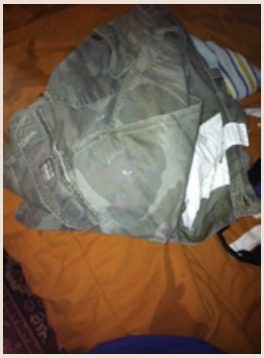
"*It's an interesting aspect of this problem we've discovered through user studies with people who are blind.* ***One privacy solution is not going to meet all needs.*** *It's a niche problem of a set of larger problems we deal with. I think that inherently it's often a few-shot learning problem. We have a new few-shot learning challenge this year, and we'll see more in future years. It's not that people have to focus specifically on privacy; it's just focusing on the challenges that are intrinsically intertwined with privacy.*"

Challenges for VizWiz Workshop



Grand Challenge Workshop at CVPR

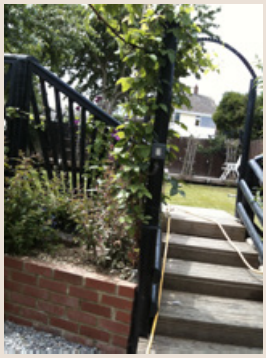
Visual questions asked by people who are blind



Do these clothes look dirty?

What flavor this is?

What types of pills are these?



Please describe what you can see.

The workshop will feature invited speakers and panelists from industry and computer vision research, as well as people who are blind.

To succeed in submitting abstracts, Danna encourages people to think about how their work could be applied to a new population they might not have considered before.

"Just because someone's not working on a dataset that's specifically from people who are blind, it doesn't mean it's not relevant," she points out.

"I would encourage people to think about why their work might be valuable to this population. Then they can show up to the workshop, test their ideas, and ***learn from industry, academia, and blind people to improve their understanding of why this work matters****. This event is a great place to hear different perspectives for anyone curious about what people do outside of the core computer vision community. I learn every year from this workshop, and I've been in this space for a long time!"*

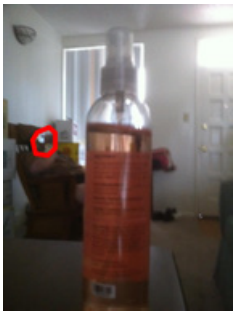
The workshop will be taking place as part of **CVPR 2022** next month, which this year is back with a physical presence in **New Orleans** after two years online. Does Danna have any advice for people attending their first in-person conference?

"These conferences are such an amazing and intense experience, where you can talk to people who generally you could never get to respond to an email!" she laughs.

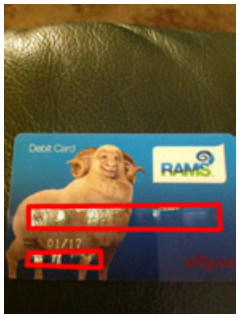
"I encourage people to spend as much time as possible visiting posters, going to talks, developing a network, and just recognizing there's a community they can belong to. The most effective way to do that is in person. There's no

"It's about helping to launch a whole new generation of developers!"

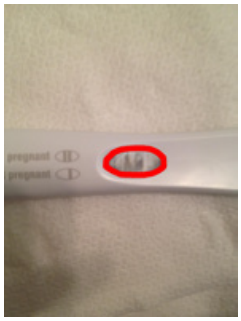
12% of 44,799 images show private content



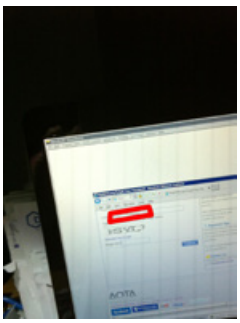
Face



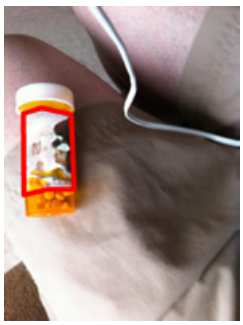
Credit Card



Pregnancy Test



Computer Screen



Prescription Pills

[Gurari et al. CVPR 2019]

replacement for showing up and making those connections. I would urge everyone to take advantage and engage."

Danna tells us this is not the only CVPR dataset challenge she is a part of this year. She is also working on the **MOT Challenge**, part of **the IEEE Workshop on Computer Vision for Microscopy Image Analysis**, which introduces a human-annotated live-cell imaging dataset for cell tracking.

When she is not busy organizing challenges and workshops, Danna is an Assistant Professor.

"*I mentor students, which I love!*" she grins.

"*I care about societal impact and trying to build a bridge between the work we're doing and how it translates into something that will change real people's lives. This workshop is a strong example of how you can have a real-world impact by translating research into industry technology. I also have students working in the medical space. That's another place where we can have real-world impact.*"

> ***"One privacy solution is not going to meet all needs."***

Two students in Danna's group have been chosen for an **oral presentation at CVPR**. Their work relates to designing unbiased visual question-answering algorithms. She invites everyone to hear her students speak about their work and to attend their poster session for a more detailed conversation during the conference.

Before we finish and completely unprompted, we might add, Danna has **a final thought**:

"*Ralph, you ask excellent questions! It's a privilege to be a part of your magazine. Thank you!*"

That is very kind, and it comes from a very old friend of **Computer Vision News**!

MEDIAPIPE (2/2)

By Marica Muffoletto (twitter)

Dear affectionate and new readers, welcome to a second chapter on **MediaPipe**. If you haven't read the first one and you want a general overview of this tool, please check the March edition of the magazine **here**.

In the previous article, we introduced this versatile machine learning framework, which is built with C++ and deployed to several platforms. We looked at the basic functionalities and the integration with Python. When embarking in the world of MediaPipe, the writer realised it was worth splitting the topic in two issues, to dedicate an entire part to the **integration of MediaPipe with Android**. It's important to mention that similar tools have been developed **for iOS and Javascript**. If interested in those, please check out the tool website or feel free to get in contact for more information!

Installing MediaPipe

Let's start with the installation of MediaPipe on your local machine. You can follow the steps below if you have a Linux machine with Ubuntu distribution (or visit **here** for Mac/Windows/Docker installation).

1. Install Bazelisk from this **link**.

Move into the desired workspace directory and clone MediaPipe repository:

```
cd $workspace_dir
git clone https://github.com/google/mediapipe.git
```

2. Automatically build OpenCV

```
cd mediapipe
sudo chmod u+x setup_opencv.sh
./setup_opencv.sh
```

3. Run Hello World! In C++ to check installation

```
export GLOG_logtostderr=1
bazel run --define MEDIAPIPE_DISABLE_GPU=1 mediapipe/examples/desktop/
hello_world:hello_world
```

If you have a GPU with EGL driver support, you can also install mesa GPU libraries, to compile your MediaPipe solutions using GPU:

```
sudo apt-get install mesa-common-dev libegl1-mesa-dev libgles2-mesa-dev
```

```
bazel run --copt -DMESA_EGL_NO_X11_HEADERS --copt -DEGL_NO_X11
mediapipe/examples/desktop/hello_world:hello_world
```

This last step should print Hello World! multiple times in your terminal window. If this happens, your setup is complete.

Building MediaPipe solutions on Android Studio

You will need Android Studio for this step. Download it from **here**, and run the following to open it on Ubuntu.

```
tar -xf android-studio-2021.1.1.23-linux.tar.gz
sudo mv android-studio /usr/local/
cd /usr/local/android-studio/bin
./studio.sh
```

Import mediapipe/examples/android/solutions directory from your previously defined `$workspace_dir` into Android Studio, by clicking on the Open button on the initial window.

Here, first sync the project by clicking on File -> Sync Project with Gradle Files (Fig.1).

If you are not very familiar with Android Studio, remember you need to setup an emulator or connect your own device before being able to run a script and see any result. You can do that through the **drop-down menu** (Fig.1) on the top right. In my window, you can see I setup a Pixel 2 XL emulator. I suggest you use the same, or, if you go for another one, you might need to define the api-level for the NDK in the workspace folder.

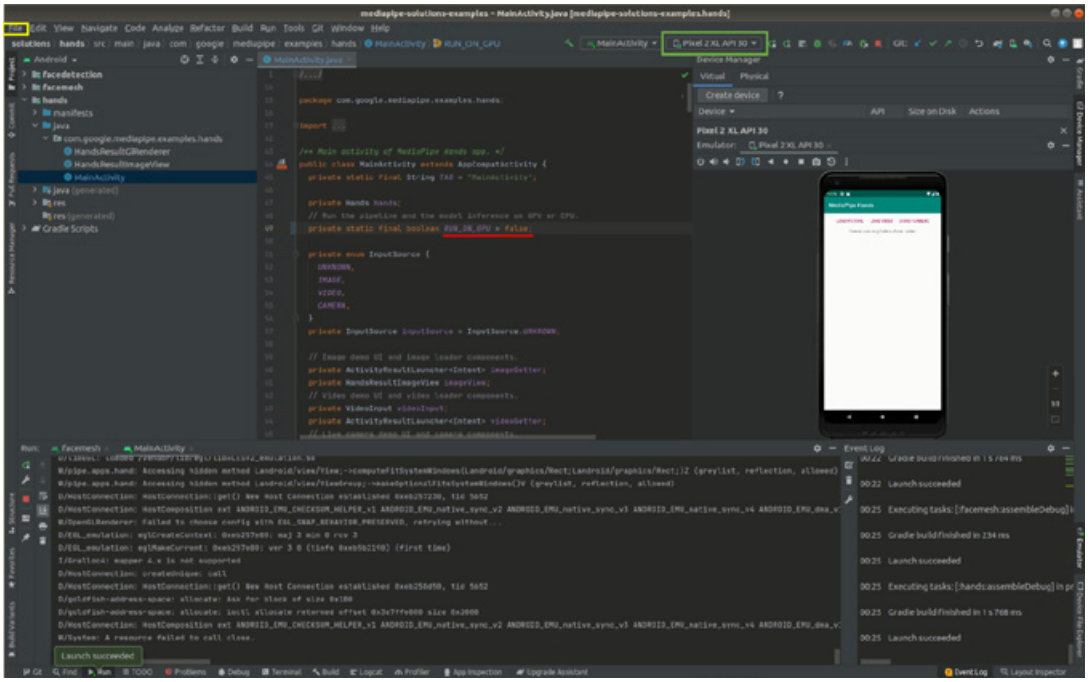


Figure 1: Screenshot of Android Studio while building the Hands solution. On the left, the Project folder shows the files found in the directory. On the right, you have the Emulator view. When the program is run, this is where you can test your application. Yellow, Green and Red boxes highlight buttons of interest for steps to follow.

As you might be expecting, since the last tutorial on MediaPipe showed a hand detection algorithm, building these solutions on Android requires a camera. Hence, if you are setting up an emulator, don't forget to link your camera to it. Press on the **pencil button** (Fig.2) linked to the chosen device and change the settings for your front and back camera.

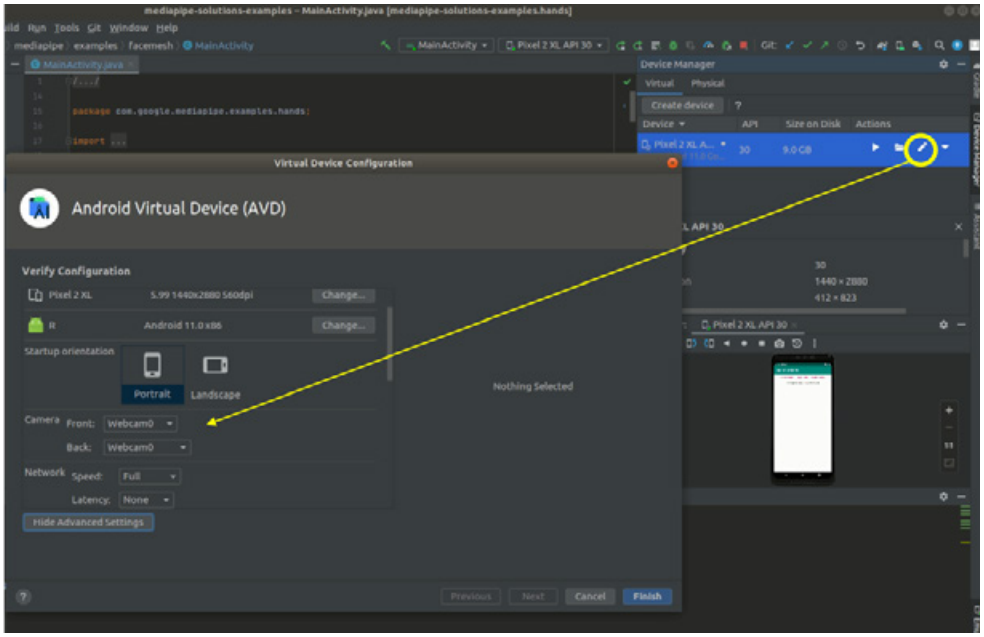


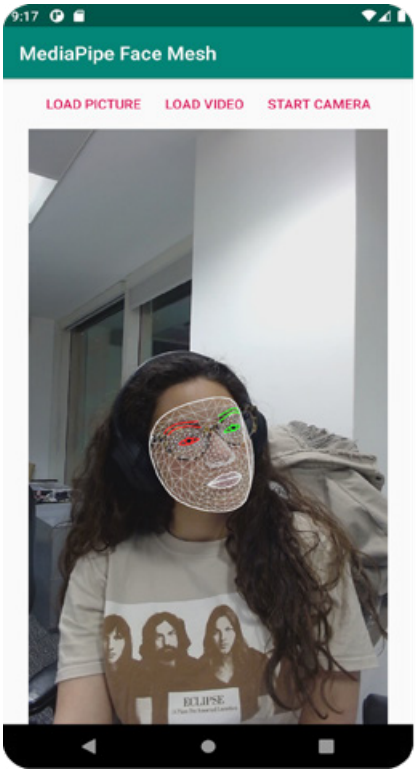
Figure 2: Screenshot of Advanced Device settings in Android Studio to setup camera for testing MediaPipe solutions.

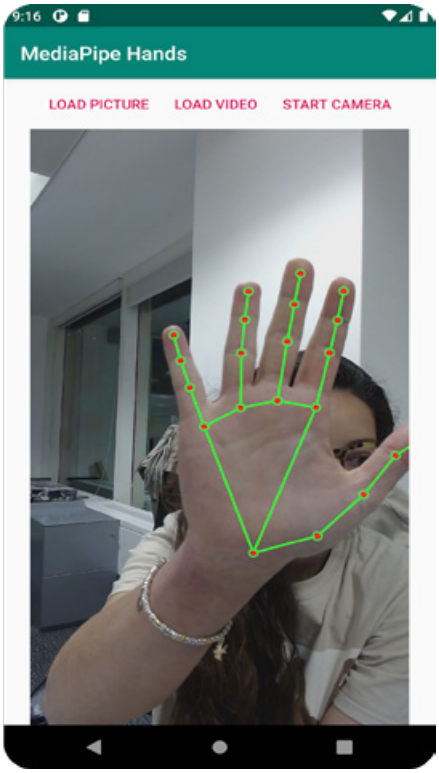
You should be ready to run any of the three available example solutions available through MediaPipe for Android. These are the Face Detection, Face Mesh and Hands examples.

It's fundamental that if you are running this without GPU support, you set the variable **RUN_ON_GPU** (Fig.1) to 'false' in the Main Activity file for each example.

We have completed all the necessary steps to build our examples. You should now be able to test the app on your chosen device.

We tried the Face mesh and Hands solution with the Pixel 2 XL emulator: results below!





A peak into the integration with Javascript

Let's finish this article with a brief overview of the MediaPipe solutions available on Javascript and how to access them.

You will find all ML pipelines available with MediaPipe staged in NPM. Hence, if you download this from **here**, you can just proceed by installing any package locally using ***npm install***.

```
npm install @mediapipe/face_mesh
npm install @mediapipe/drawing_utils
npm install @mediapipe/camera_utils
npm install @mediapipe/control_utils
```

The commands above install the face mesh solution, the utilities to draw landmarks and connectors, to operate the camera and to show sliders and FPS widgets.

Another great service offered by MediaPipe are the demos on Codepen. You can find them **here** and test any solution live. Through the button on the top right, you can alternate between a full-page view for testing the pipelines or an editor view which shows you the HTML, CSS and JS codes. On the left of the Codepen window, you can find sliders to change parameters of the MediaPipe app. We introduced these for the Hands demo in the former article. The *max_num_hands* can be used to set the maximum number of hands to detect, *model_complexity*, *min_detection_confidence* and *min_tracking_confidence* to vary the lower bound confidence for both the hand and landmark detection.

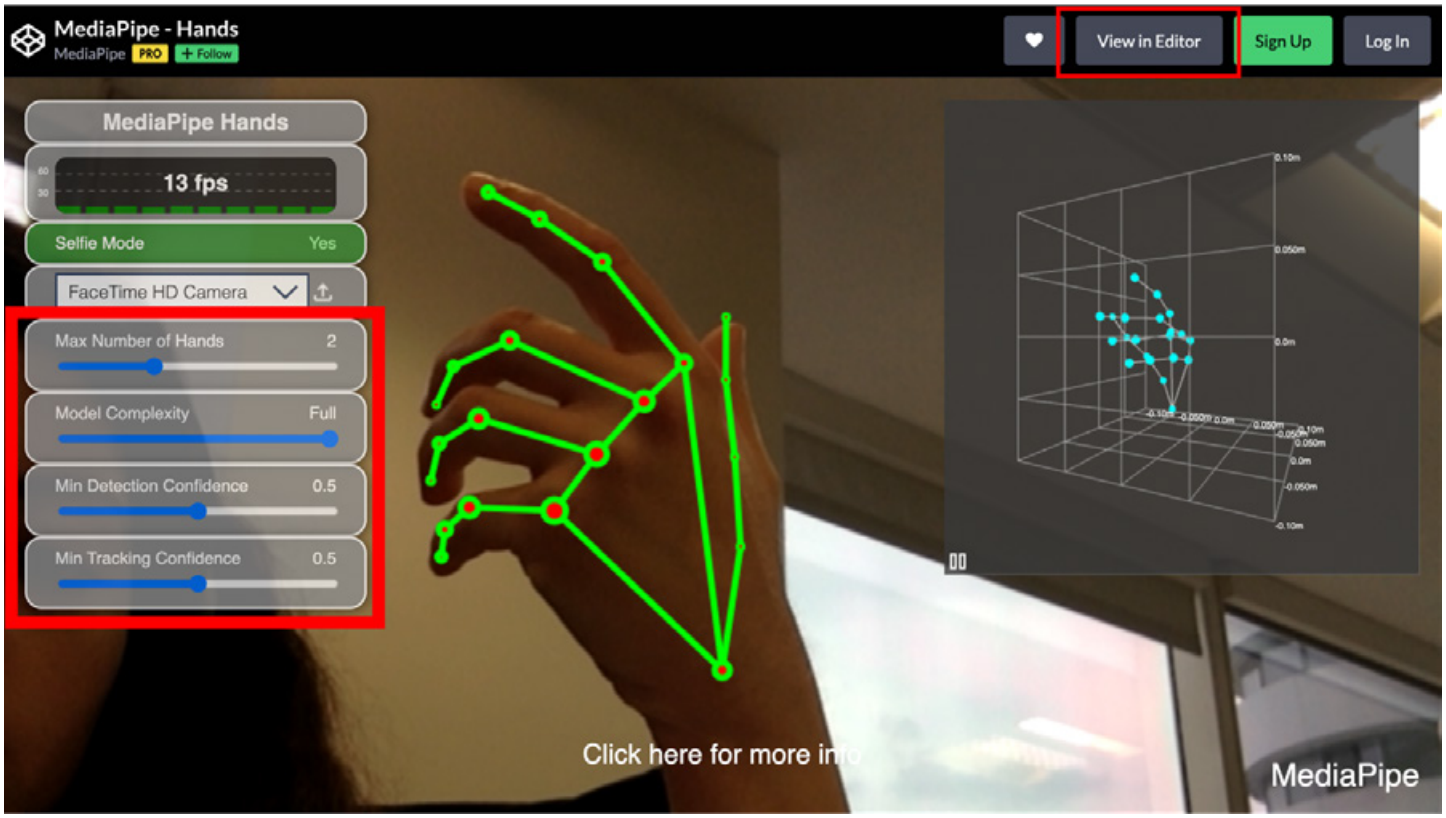


Figure 3: Full-page view of the Hands demo on Codepen.

Moving on to the Editor view, you will find the JavaScript code that interacts with the MediaPipe tools. Changing this will allow you to modify the demo.

Below, I am reporting the last part of the JS code where the user can change the range or the default values for the parameters.

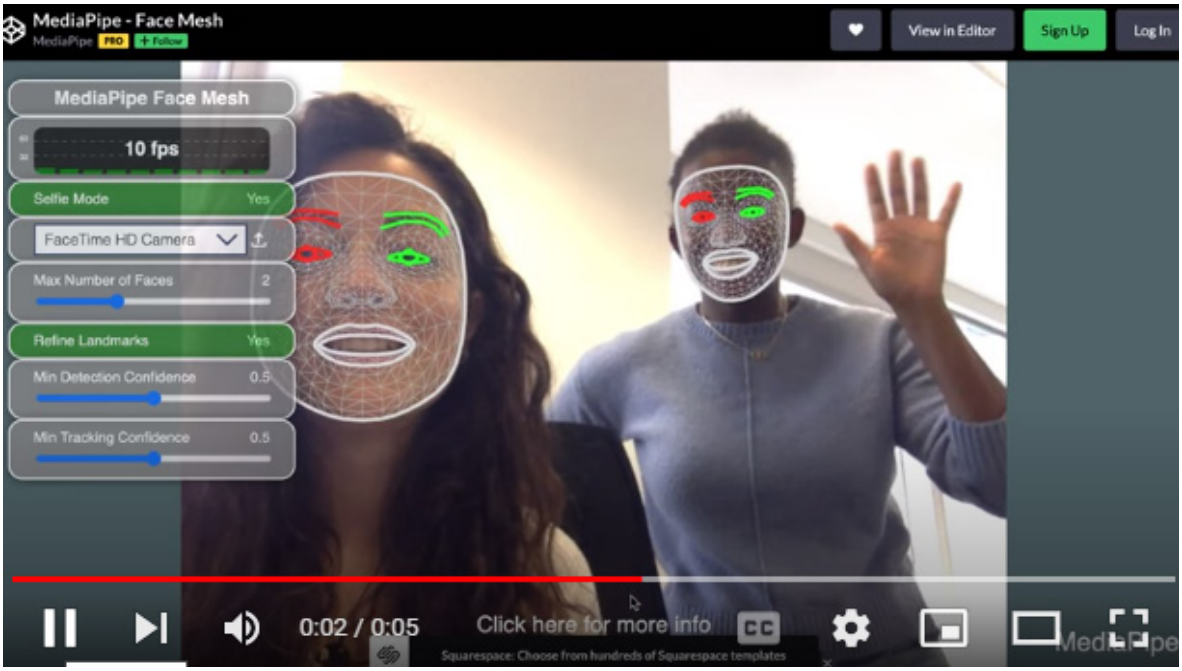
```
// Present a control panel through which the user can manipulate the
solution
// options.
new controls
    .ControlPanel(controlsElement, {
      selfieMode: true,
      maxNumHands: 3,
      modelComplexity: 1,
      minDetectionConfidence: 0.5,
      minTrackingConfidence: 0.5
    })
    .add([
      new controls.StaticText({title: 'MediaPipe Hands'}),
      fpsControl,
      new controls.Toggle({title: 'Selfie Mode', field: 'selfieMode'}),
      new controls.SourcePicker({
        onFrame:
            async (input: controls.InputImage, size: controls.Rectangle)
=> {
              const aspect = size.height / size.width;
              let width: number, height: number;
              if (window.innerWidth > window.innerHeight) {
                height = window.innerHeight;
                width = height / aspect;
              } else {
                width = window.innerWidth;
                height = width * aspect;
              }
              canvasElement.width = width;
              canvasElement.height = height;
              await hands.send({image: input});
            },
      }),
      new controls.Slider({
        title: 'Max Number of Hands',
        field: 'maxNumHands',
        range: [1, 4],
        step: 1
      }),
```

```
    new controls.Slider({
      title: 'Model Complexity',
      field: 'modelComplexity',
      discrete: ['Lite', 'Full'],
    }),
    new controls.Slider({
      title: 'Min Detection Confidence',
      field: 'minDetectionConfidence',
      range: [0, 1],
      step: 0.01
    }),
    new controls.Slider({
      title: 'Min Tracking Confidence',
      field: 'minTrackingConfidence',
      range: [0, 1],
      step: 0.01
    }),
  ])
  .on(x => {
    const options = x as mpHands.Options;
    videoElement.classList.toggle('selfie', options.selfieMode);
    hands.setOptions(options);
  });
```

We are at the end of this MediaPipe review.

We hope you liked it and that you will use this customizable platform for your own projects.

Special greetings below!



Congrats, Doctor!

(Jacob) Zhiyuan Fang recently completed his PhD at the Active Perception Group, Arizona State University. His research interest spans over the intersection of Computer Vision and Natural Language Processing, with the focus of building efficient and strong Vision-Language Models from weak supervisions. He developed Vision-Language learning algorithms from the perspective of knowledge distillation and implicit supervisions on efficient deep neural architectures. During his PhD, he also spent times in Microsoft Azure AI as research intern and he is joining Amazon Alexa AI-Lab126 as a research scientist in June 2022. Congrats, Doctor Jacob!

An important objective of AI is to understand real-world observations and build up interactive communication with people. The ability to interpret and react to the perception reveals the important necessity of developing such a system across both the modalities of Vision (V) and Language (L). Although there have been massive efforts on various VL tasks, e.g., Image/Video Captioning, Visual Question Answering and Textual Grounding, very few of them focus on building the VL models with increased efficiency under real-world scenarios. The focus of my research is to comprehensively investigate the very uncharted efficient VL learning, aiming to build lightweight, data-efficient, and real-world applicable VL models. The proposed studies in my research take three primary aspects into account when it comes to efficient VL:

1) **Data Efficiency**: collecting task-specific annotations is prohibitively expensive and so manual labor is not always attainable (see Figure 1). Techniques are developed to assist the VL learning from implicit supervisions, i.e., learn the associations between visual concepts and semantics in a weak/un-supervised fashion. For example, when given an image, learn which specific region best corresponds to a textual query “A man in red topping”, when we have only an image-level general description without knowing the exact region corresponds to specific visual concepts. Part of my works aim to build textual grounding system in images and moment localization in videos through language in weakly supervised fashion.

2) **Efficient representation learning with increased scalability**. It is challenging

Figure 1: VL tasks like referring expression and video moment localization require explicit annotations (i.e., bounding boxes of target objects or temporal boundaries for the moments) which requires human annotations.

how to effectively mine the hidden visual-textual associations at scale for representation learning. My work in two other papers studies the knowledge distillation (KD) technique for generic V and VL Representation Learning which proves to bring substantial performance gain than the regular representation learning schema. Empirical studies show that our KD assisted representation learning method is more data efficient and brings better performances.

3) **Architectural Efficiency**. Deploying the VL model on edge devices is notoriously challenging due to their cumbersome architectures. Much of the existing VL models focus on large models that suffer from high latency and large memory footprints at the time of inference, as well as the inconvenient two-stage design which requires the use of object detector as visual feature extractor. These all limits their deployment to resource-constrained edge devices for real-world applications. I study how to train small and efficient VL models from the perspective of model compression. To further extend these advancements to the real world, a novel one-stage VL architecture is designed recently in my work to tackle the inference bottleneck and the inconvenient two-stage training, which brings great training flexibility. Extensive discussions have been conducted on several critical aspects that prominently influence the performances of compact VL models.

Ana Serrano

Assistant Professor at Universidad de Zaragoza in Spain.

Spanish Minister of Science visit

Ana, can you tell us about your work?

My research spans several areas. In particular, I work on computational imaging, which is the capture part of image perception. It's about how we see or how we perceive different images or rendered content. Also, in the visualization part, digital reality at the moment. All these things are quite different, but I enjoy doing different things. The one thing in common, which fascinates me the most, is focusing on perception-motivated solutions. I really enjoy creating and applying fundamental knowledge about how the human visual system works and to use all this knowledge to improve different experiences, like virtual reality visualization, capturing devices, and so on. I really enjoy learning how things work and to apply all this to whatever work I'm doing.

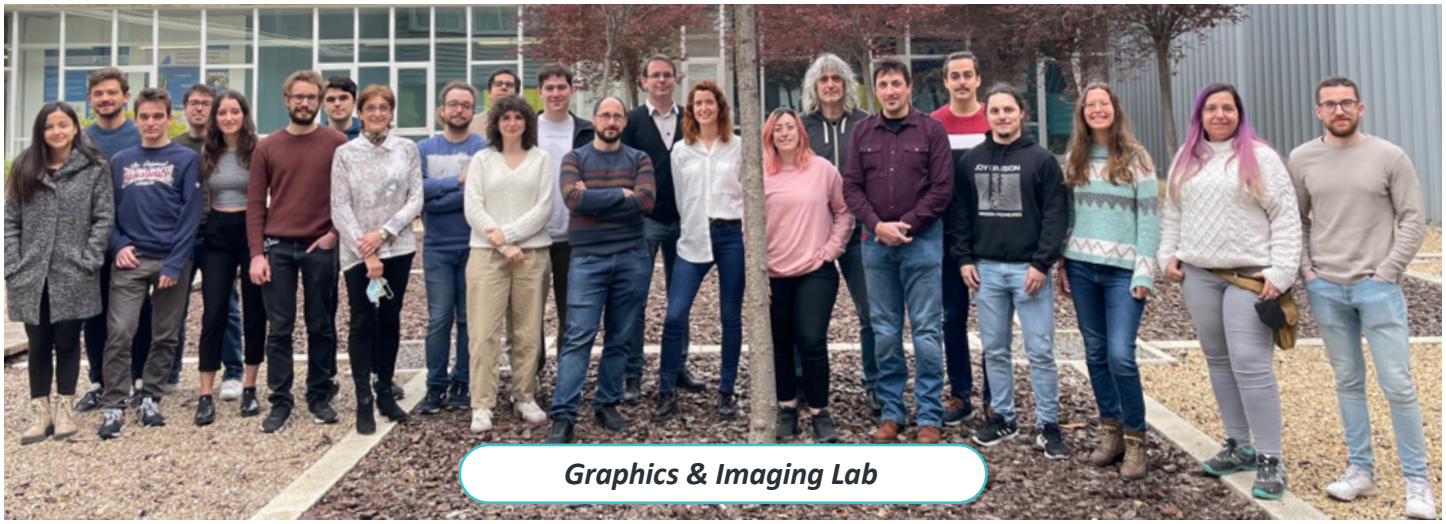
All these things have one more thing in common, which you did not mention: they are very complicated.

[*laughs*] Yes, indeed. They are complicated. Once you get into it, it becomes easier with time. There are people that work on one single thing, and they become experts. I don't want to say I'm not an expert, but that's a bit different for me because I like working on a broader set of things. Sometimes I don't have as much time to devote to a single one. That's good, I think. I like to apply things I learned in one field to other fields. In the end, that's also rewarding. It also gives me ideas about how to interact with all these fields. I'm really happy.

When did you discover that you had a gift for complicated things?

[*laughs*] Quite late actually. I was never a brilliant student at school, in high school or the beginning of university either. I did well, I guess, but I wouldn't say I considered myself brilliant. I persevere. I keep going in one direction if I think this is the correct direction. This paid off in the end. During the last two or three years that I was in college, I started getting to know research. I was interested in research in general. I wasn't pursuing a particular field, not in computer graphics or computer vision. I was just a regular engineer, let's say. Then I got in touch with computer graphics and computer vision. I started to realize that I was doing good! I was doing better at research than regular problem-solving. I was very happy that I finally discovered what I wanted to do. It was not clear to me... [*laughs*]

I am happy you did! You mentioned perseverance as one of your main strengths. This seems very common among scientists. What talent do you have that most scientists don't have?



Graphics & Imaging Lab

I would say it's a combination of talents. I don't think of myself as a brilliant person, but I have a combination that makes me a good researcher. I am very curious. This is also common if you do research. I like to know how things work. I like to dig deeper. Also, perseverance as we mentioned. One thing that helped me a lot is to be very organized. It doesn't sound very impressive, but it helps me a lot when there are tons of papers to read and tons of tasks. I am very organized. I am meticulous with everything I do. I don't think any of these talents are exclusive. Most people that do research have these talents.

Yoshua Bengio told me: I'm not smarter than other researchers, but I'm able to focus a lot!

I agree. Of course, you need to have ideas and know how to carry out these ideas. Personally, I don't think that brilliance is necessary. There are many other things that are important.

Can you tell us more about your work at your lab?

I would like to talk about the virtual reality part of it. I enjoy everything, but I am more excited about this work. It's newer and trending now. I'm interested in understanding how people explore virtual environments and how people interact with virtual reality in different ways. These days, virtual reality has become a huge thing. The hardware is getting better and better. It's already pretty good, but I'm confident that it will just keep getting better. One thing that is going a bit slower is content creation: we were just trying in virtual reality whatever we knew about regular video games, regular cinematography. That may work, but it may not be the best way of exploiting the capabilities of virtual reality. When movies started, one of the first movies, *Arrival of a Train* by the Lumière Brothers, was very, very different from the films we have today. Cinematography evolved, and content creation also has to evolve. What I'm doing today is coming up with questions that are interesting. For example, one work we did not long ago was about cinematic virtual reality. What happens when you have a cut? When you work on movies, you have one scene, cut, then move to another. What happens in virtual reality? You're suddenly getting teleported somewhere else. We come up with these questions. We try to see what happens and how people perceive this. Most of my work is devoted to trying to

come up with interesting questions and trying to come up with interesting studies. That is the most important thing for me.

In the end, the users are going to consume this content. It's a key part of my work to bring the person in the loop. I usually don't do super engineering research. I'm interested in applications later down the road. Mostly, I'm interested in understanding how people perceive these things, if an experiment is not well designed. When I started working on perception, many years ago, I wondered, "*How hard can this be?*" You're just asking people questions. Turns out, it's pretty hard! [*laughs*] I remember one of my first experiments. It was about material appearance perceptions. You only have to ask people, "*How glossy is this material?*" I had to try the experiment like ten times. I was not asking the right questions. I was not getting in the mindset of a person who is not a researcher, who may not be very familiar with the technology. It's really a hard task, and it's a task that I enjoy a lot.

What breakthrough finding would make you the Lumière Brothers of virtual reality?

That's a good question! I don't think there's a single point that you say, "*Oh! This is it!*" It's hard to be the person that comes up with this one thing. It's a constant effort, not just for myself but all the people I work with, also all the people I don't work with. It's a community effort, in this case, to bring virtual reality forward.

This is your moment to speak to the community about what you want to create!

That is a big question! Everything is good as it is. Recently, I was on a committee for a small workshop about virtual cinematic language. The community is doing good. We have a healthy community. Everyone wants to read others' work and just keep pushing. One thing that would be good is to try to communicate more with practitioners, with people who may not be in research but doing the work every day, for example, cinematic companies that focus on virtual reality. There is already communication. Sometimes you see research works that answer the questions that we believe are interesting. It would be much more effective to communicate more with the

Graphics & Imaging Lab

practitioners that actually do this. That's what I hope for, to have great collaboration between practitioners and researchers. That would really take us forward.

While we are talking, I can see behind you lots of badges from conferences.

Yes, when I started collecting, I didn't think that I would have so many! I started my PhD in 2015. I thought that going to one of these huge conferences would be once in a lifetime or I was just lucky that I got my paper in. It turns out that wasn't true. Now with COVID, there were virtual conferences. I don't have the badges. This year, I started adding badges again to the collection.

Will we see you next month at CVPR?

No, unfortunately, I cannot go to CVPR this year. One thing that happened this year is that I changed jobs a couple of times very quickly. Then I started teaching so I had to prepare for all the classes. This year is a bit complicated for me. Hopefully, next year everything will be more settled, and I will be able to go to more conferences again.

Years ago I interviewed another brilliant researcher from Zaragoza, Clara Fernàndez Labrador. She was just finishing her PhD. Do you know her?

Yes, I do. Zaragoza is not a huge city. The university is not huge. Everyone that works in computer science, computer vision, computer graphics, artificial intelligence, and computer architecture, we are in the same building. Sometimes we are in the same corridor. So I do know her. I haven't worked with her, but I had the pleasure to read some of her works. We all know each other.

Is there anything you wish you could work on, that you don't have time for now?

Augmented reality! I haven't been working on this. As a person who enjoys video games and virtual reality, I think it has a lot of potential, even more than virtual reality,

because augmented reality has a social component. Think at Pokémon Go, a game that had a boom like two or three years ago. It was great! You would go in the street and see an old lady buying groceries. She would come out with groceries, on her phone, chasing Pokémon! This is great! Everyone enjoys it. It has a social component. You gather with people. I am interested in augmented reality as a consumer, but I would also like to work on this!

One thing that strikes me is how many internships you have had.

I would definitely recommend it. I enjoyed it, and I benefited a lot from many perspectives. One thing I learned is how to work with many different people in many different environments with other supervisors. It's great to adapt to all those situations. I learned from all of this. I also got to know all these great people from Stanford, Max Planck, and Adobe Research. It also helped me decide what I wanted to do, whether I wanted to go to academia or industry. That helped me a lot. I also had a lot of fun. I have always wanted to travel since I was very young. You go there to work, but you also go there to meet people, to have fun, to know the places. They were great experiences.

IEEE VR Japan

Switzerland United Nations

What was the most fun of all your adventures?

This is actually on the last year before I started my PhD, at CERN - the European Organization for Nuclear Research. I was there doing my first internship ever, before I started research. I didn't know what I was doing. I had great fun! We weren't supposed to, but we did it. You could go under the research facility. There were a lot of tunnels. We used to gather there to discuss things and have fun. I come from Zaragoza, which is not a huge city, and I've managed to come to CERN. I am now discussing with all these great people and having fun underneath the research facility at, I don't know, 10 pm! That was amazing! This is the internship I probably enjoyed most, not in terms of research, but it was the first time I was like, "*Whoa! I'm here!*" I'm meeting great people! I'm at this amazing place!

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

Andrew NG to start a new Machine Learning Specialization

Let's start with one of the pillars of our community. This tweet was already seen by many people, but maybe not by everybody. At the time of **Coursera's 10 years anniversary** (good job, guys!), **Andrew NG** announces that he is going to start a new **Machine Learning Specialization**, created by **@StanfordOnline** and **@DeepLearningAI_**. It will be a new and separate program but is an expanded and updated version of the original standalone course. Many people started their career on Andrew's first ML course and now he asks to help him spread the word. Here you go, Andrew! Read More

Expressive and Emotional Speech Generated from a Neutral Prompt, with Laughter, Yawns, Hum, Etc.

We mentioned Andrew, let's talk about Yann. Yann Lecun shared **Meta**'s recent announcement about a few more milestones toward more expressive **NLP models**, which add up to last year's breakthrough model called **Generative Spoken Language Model (GSLM)**. Apparently, Meta claim they are able now to model **expressive vocalizations, like laughter, yawning, and cries**; as well as spontaneous, real-time chit-chat, occasional overlaps or pauses and other nuanced social cues and signals, like interruptions, as well as positive or negative feedback, helping researchers to build more natural AI systems

Read More

Researchers Warn: Risks of Using AI To Grow Our Food Are Substantial and Must Not Be Ignored

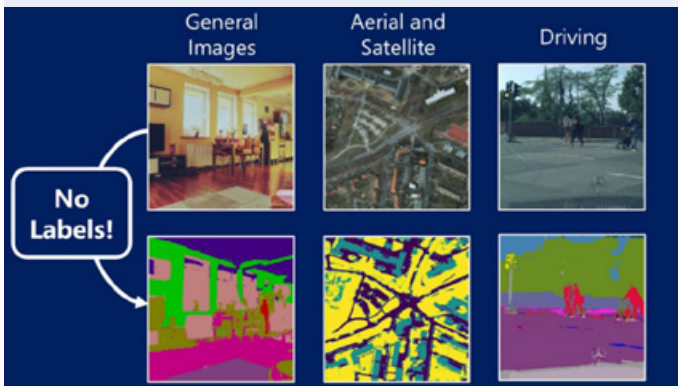
This article informs us that **Artificial Intelligence** is on the cusp of driving an agricultural revolution, and helping confront the challenge of feeding our growing global population in a sustainable way. Which is very good, and we sort of knew that already. But researchers warn that **using new AI technologies at scale holds huge risks**: hackers messing things up, security failing, environment mismanagement, exploitation of disadvantaged communities and so many other bad consequences. Moreover, the benefits will likely be more visible in the more developed and internet-savvy regions of the world. Read More

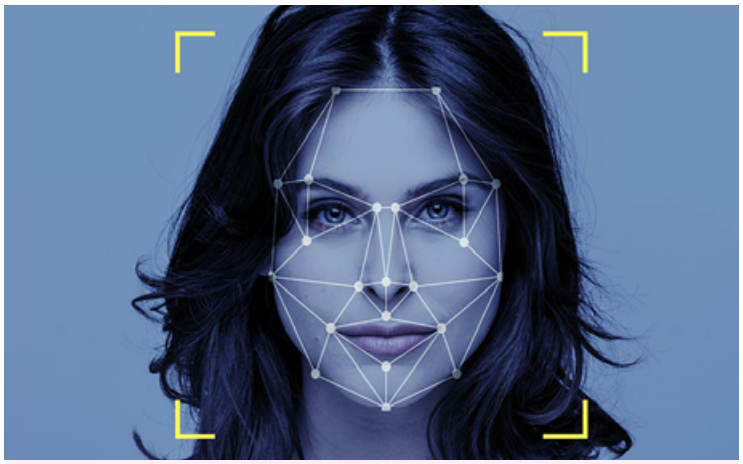
Israeli-Based Tech Startup Brings Your Old Family Photos to Life with Amazing Artificial Intelligence

This Forbes article relates of something amazingly beautiful and touching. Leveraging **Artificial Intelligence** and sophisticated technology, the company has created a unique, animated, live portrait, which animates the photos of long-lost relatives or whoever you'd like to see, as if they are in the room with you. Its tech makes people come alive and look realistic and natural. The feature, **Deep Nostalgia**, lets users upload a photo of a person or group of people to see individual faces animated by AI, breathing life into old black-and-white photos of grandma and grandpa stored in boxes up in the attic. **Read More**

A New State of the Art for Unsupervised Computer Vision

That would sound too ambitious, if it didn't come from the guys at **MIT CSAIL**: they created an algorithm to solve one of the hardest tasks in computer vision: assigning a label to every pixel in the world, without human supervision. **"STEGO"** can jointly discover and segment objects without any human labels at all, down to the pixel. To discover these objects without a human's guidance, STEGO looks for similar objects that appear throughout a dataset. It then associates these similar objects together to construct a consistent view of the world across all of the images it learns from. **Watch the Video**





An Optimized Solution for Face Recognition

More News from MIT. They're good! When AI is tasked with visually identifying objects and faces, it assigns specific components of its network to face recognition - just like the human brain. Which means that it's a specific area of our brain that enables each of us to recognize hundreds or thousands of people's faces; and that AI trained to identify faces and other objects discovers a surprisingly brain-like strategy to sort them all out. What stroke MIT's McGovern Institute for Brain Research scientists is that two completely different systems have independently devised a very similar solution to the same problem. **Read More**

VISION UNDERSTANDING AND MACHINE INTELLIGENCE (VISUM) SUMMER SCHOOL 2022

Ana Rebelo is a Senior Researcher at INESC TEC. Sara Oliveira, Wilson Silva, and Tiago Gonçalves are PhD students at the Faculty of Engineering of the University of Porto (FEUP) and Research Assistants at INESC TEC. Wilson and Tiago are also Invited Assistants at FEUP. Together, they are four of the organizers of a fantastic computer vision and machine learning summer school in Porto later this year and are here to tell us all about it.

\The VISion Understanding and Machine Intelligence (VISUM) Summer School launched a decade ago when Ana completed her PhD and recognized there was a gap in Portugal on computer vision. She spoke to her colleagues at the time, and they came up with the idea of launching a summer school on the topic.

"*We wanted to create a different summer school from the others,*" Ana tells us.

"*We decided to focus on computer vision and machine learning. We had a theoretical license, but we also had a practical license, so students could put into practice what they were learning.*"

The summer school has evolved each year since then. Now, it includes a **computer vision project challenge** that runs across the week and an **industry day**, with company representatives showcasing real-world applications and proposing mini challenges to students.

The team changes a little bit each year, with new people bringing new ideas for improving things, which Ana says keeps the event fresh.

"*When I started this, I chose to do it with young people because most seniors are in academia, they are professors, and I wanted* ***an environment where students were networking with other students****,*" she explains.

"*This year, we have Sara and Wilson on our main organizing committee, and I think it's the best one we've done. We have great speakers, a very nice challenge, and a completely different industry day. There's a real spirit of teamwork throughout the entire event, and I think that's why we all enjoy it so much.*"

Sara adds:

"*It's so rewarding to know that everyone has enjoyed it at the end of the week. That's the most important thing for us. Also, even though it's a lot of work, it's important that we enjoy it too! As Ana said, we have* ***a great team and spirit!***"

Wilson agrees:

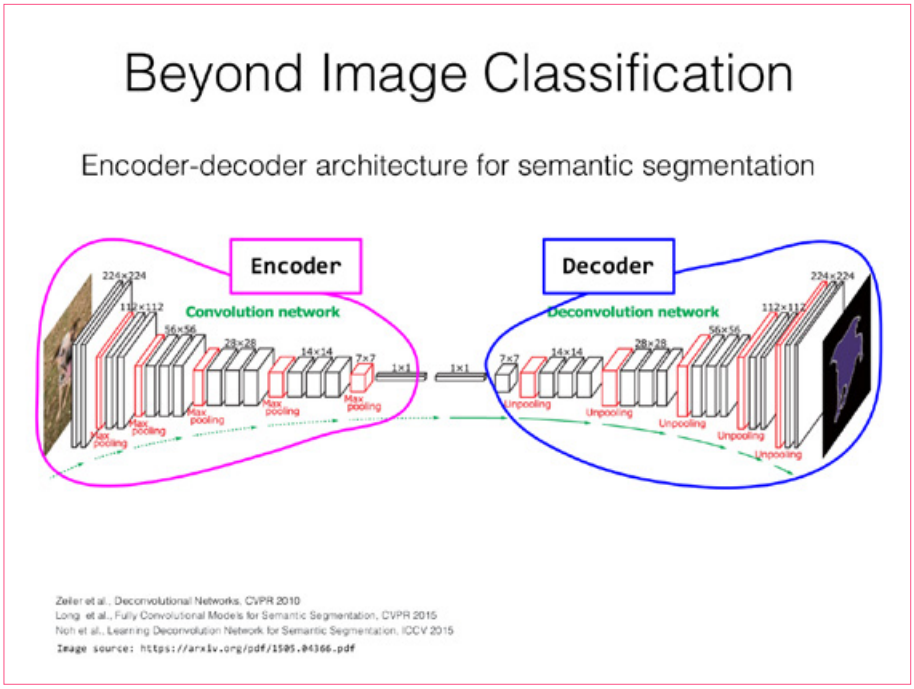
"It's been such fun to work with colleagues on a different project from our everyday work. I like the ***multicultural environment*** *– we have participants from many different countries. We learn a lot culturally from everyone."*

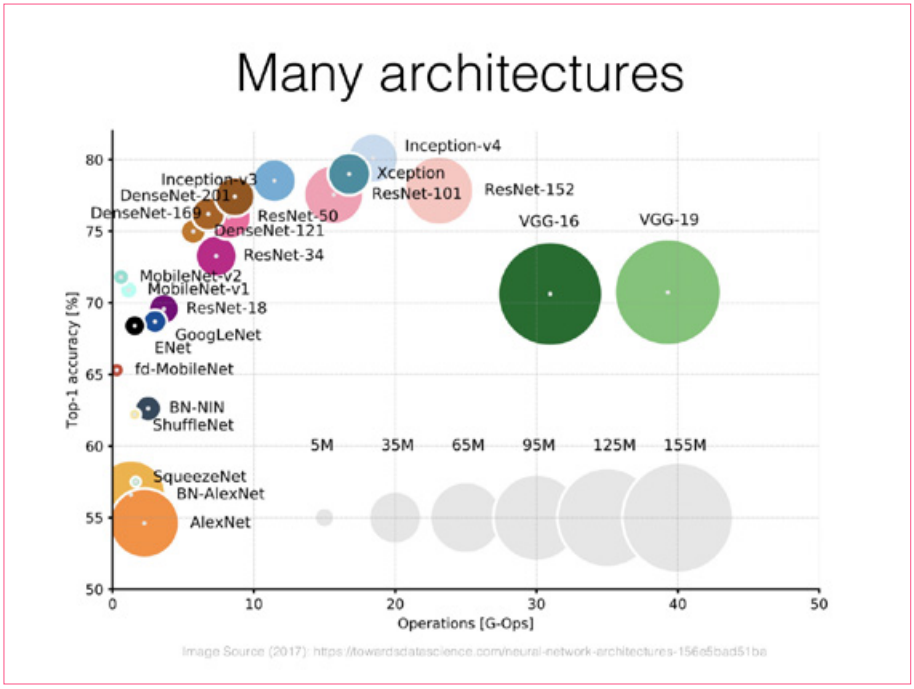
Up to 80 people are expected at this year's **in-person English-speaking event in Portugal**, which follows two digital editions in 2020 and 2021. Attendees are generally in the last year of their masters or early in their PhD, as well as researchers and people from industry who want to learn more about computer vision. Applications closed last month, but if yours has been accepted, do not forget to register as soon as you can to secure the best value ticket.

Tiago joined VISUM a couple of years ago, and this year he is leading **the project competition**.

"VISUM is one of the best ways to travel without leaving home!" he declares.

"We get to meet new people, new cultures and broaden our views. The best part is





working with real-world problems, not just theoretical ones from the lab. I can get out of my box, join a team, prepare data, and discover new ways of solving things so that participants get a controlled version of a problem to work on during the week. In one week at VISUM, ***you will experience as much as you would across one or two years otherwise****!"*

In this 10th edition, running parallel to the challenge, the event has **a mentorship program**. Previously, the team identified that some people did not feel comfortable participating in the competition, so this year, attendees will be able to choose to work on small research problems with a mentor. **Mentors have been selected from different backgrounds, including research and industry**.

"We welcome people who don't have a computer vision or machine learning background but want to learn and challenge themselves," Sara points out.

"Even if you're a newbie, come along, and if you're not comfortable enough to join the challenge, you can get started with the mentorship program instead. We also have basic sessions where we introduce all the fundamentals of computer vision. Everyone is welcome at VISUM, even if it's not yet your area of expertise!"

Finally, we ask: what else should encourage people to come to Porto?

"The food is the best part!" Wilson laughs.

"Not only in Porto but in Portugal in general. And the good weather, especially in July!"

VISUM will take place in Porto, Portugal from 10-16 July.

COMPUTER VISION EVENTS

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

Embedded Vision Summit
Santa Clara, CA
17-19 May

Image Analysis and Processing ICIAP
Lecce, Italy
23-27 May

CARS 2022
Tokyo, Japan
7–11 June

CAOS 2022
Brest, France
8–11 June

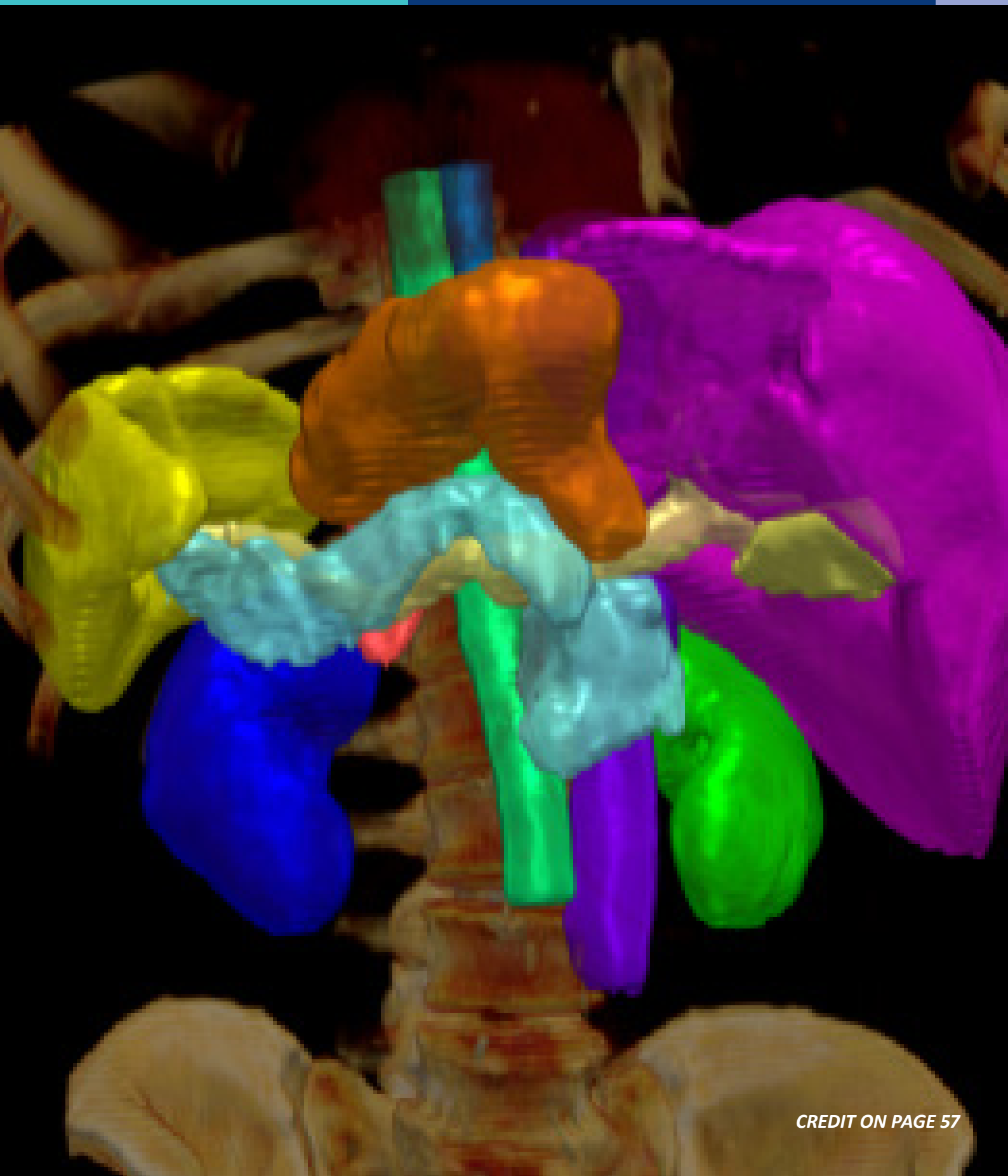
CVPR 2022
MEET US THERE
New Orleans, LA
19–24 June

MIDL 2022
Zürich, Switzerland
July 6-8

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

MAY 2022



CREDIT ON PAGE 57

SEMI-SUPERVISED DEEP EXPECTATION-MAXIMIZATION FOR LOW-DOSE PET-CT

Vatsala Sharma is a third-year PhD student at the Indian Institute of Technology in Bombay. Vatsala and her colleagues have just won the Best Paper Award at ISBI 2022 for their work "Semi-Supervised Deep Expectation-Maximization for Low-Dose PET-CT", proposing a new system for low-dose PET-CT image enhancement in the case of limited data availability. Vatsala speaks to us about their acclaimed research.

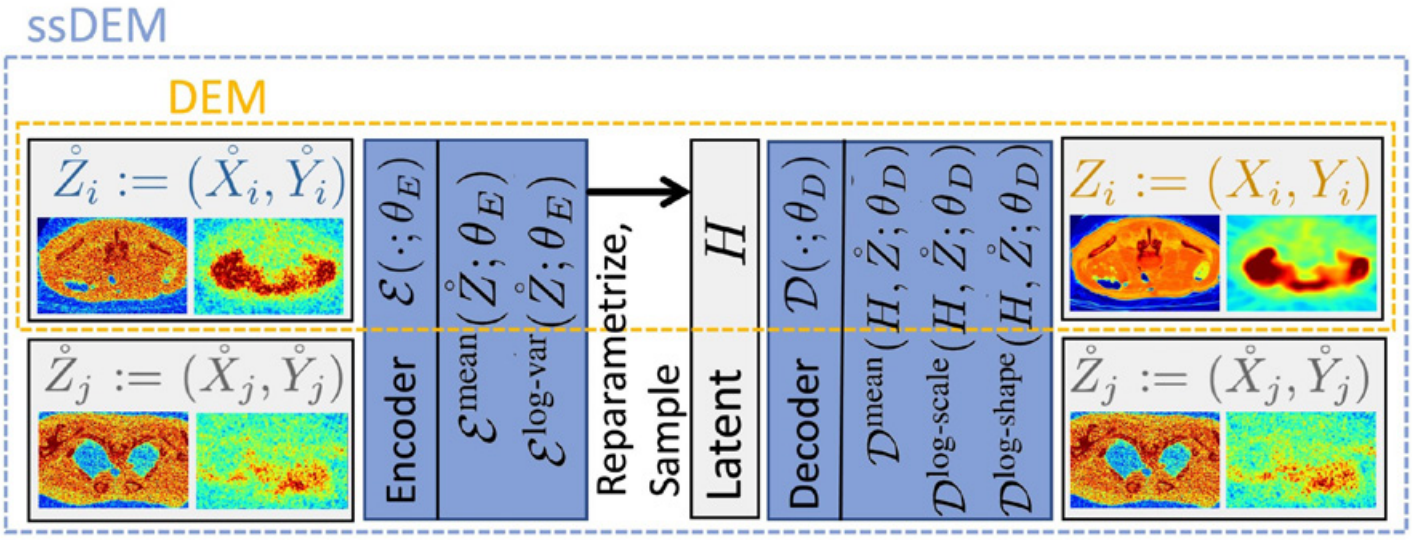
Patients are subject to radiation during the scanning process for both PET and CT. If someone needs multiple scans over their life, this can be detrimental to their health. Using **low-dose PET-CT imaging** reduces the health risks associated with radiation. However, **noise and other undesirable artifacts** can reduce the quality of images when scanning at low doses.

This work proposes using semi-supervised learning to overcome this, with a solution based on **deep learning**, developed using Python and the PyTorch library.

Vatsala explains that the work had many different parts, and to make it more robust, the team each worked individually on separate sections before trying to make the whole thing fit together. Did they succeed the first time?

"*Definitely not on the first try!*" she laughs.

"*When you're focusing on the bigger picture, sometimes you miss simple things like small errors in your script, and then after speaking to your co-authors, you realize you've made a silly mistake. For example, I got an order number error in my code, and I couldn't understand why, but it was simply because of the way I was calculating a probability value. There was*



Overview of the framework proposed in the paper. DEM stands for Deep Expectation Maximization which uses paired low-dose and full-dose PET-CT images. The ssDEM framework is its semi-supervised version which additionally utilizes data where no full-dose images are available.

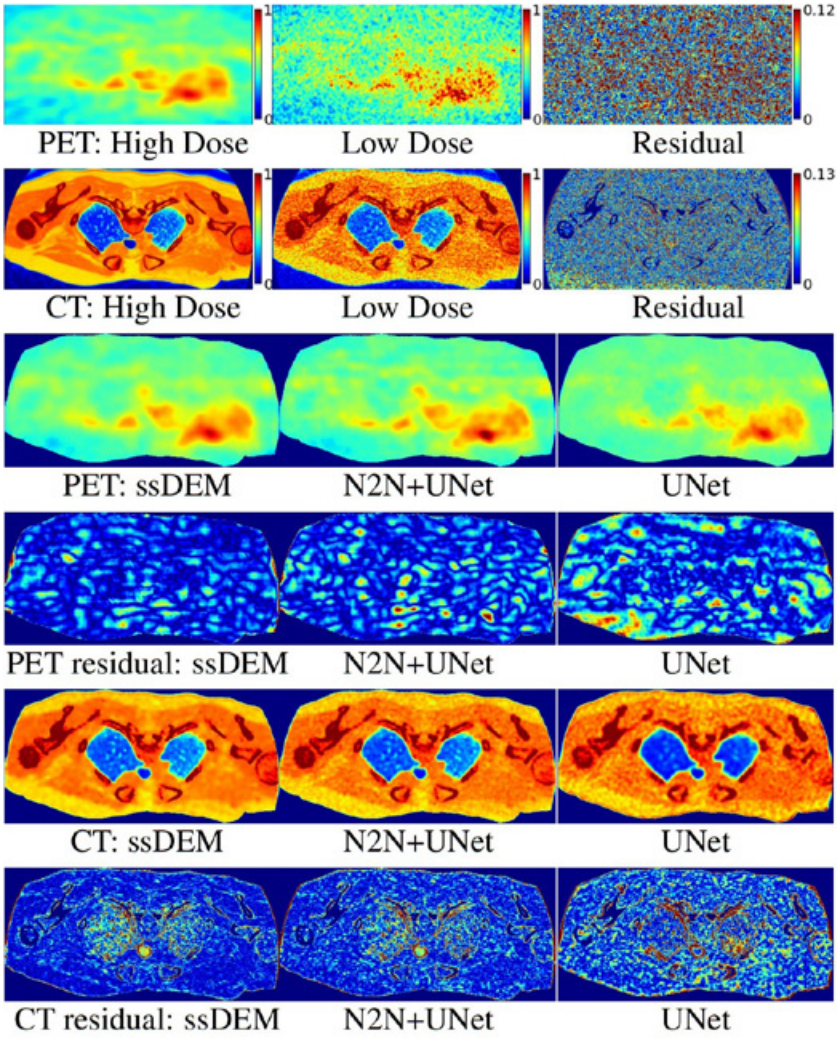
a lot of back and forth between us!"

Winning a Best Paper Award is an outstanding achievement, especially as this was Vatsala's first oral presentation at a big conference. Congratulations! What does she think convinced the jury?

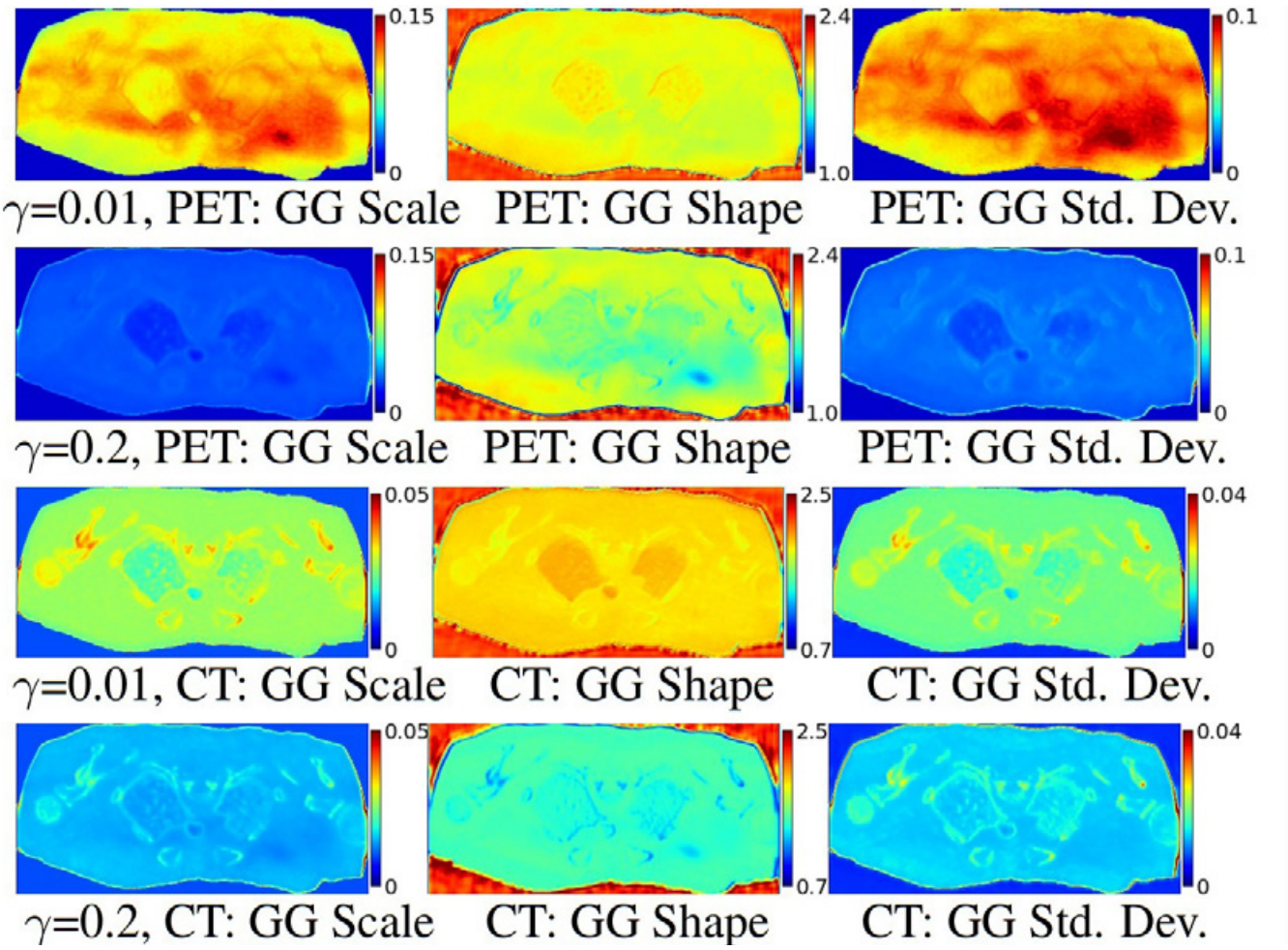
"*I think it's the novelty of combining traditional statistical approaches with more modern techniques,*" she tells us.

"***Expectation-maximization*** *is a very old technique used in different fields over many years. We're combining that with the far more recent field of deep learning.*"

Despite living in India, Vatsala presented her work virtually as she could not travel to Kolkata due to issues related to the pandemic.



Some qualitative results produced by the ssDEM framework, compared against UNet (a fully supervised framework) and N2N+UNet (a semi-supervised framework we proposed as a baseline).



The proposed framework is uncertainty aware, i.e. along with predicted outputs, it also provides the uncertainty maps associated with the outputs. These are obtained as the standard deviation of the Generalized Gaussian distribution that the network produces as output. These are shown in this image, where higher supervision during training leads to lower uncertainty.

"With a hybrid conference, you have to prepare your talk in advance, so I had plenty of time to go over my notes and all the work we had done," she says.

"Obviously, there were some nerves before the presentation started, but I was well prepared!"

Thinking about the next steps for this work, whilst this paper is focused on PET and CT, she would like to explore whether the framework could work for other problems, both within and outside of medical imaging, and **if it could be extended to non-vision tasks**.

In terms of her own future, Vatsala says she is still pondering what to do once she has completed her studies.

"As a PhD student here, my work is divided into two," she points out.

"Research – working on these new concepts and figuring out problems – and being a teaching assistant."

What if she had to pick one?

"I would choose research," she reveals.

"Even in academia, many researchers also teach, but I think research is the most enjoyable part!"

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

MRI findings associated with poor thrombectomy outcomes after stroke

Recent research published in **Radiology** (based on the MRI brain scans of 366 patients after **endovascular thrombectomy - EVT**) found that as the presence of **white matter hyperintensity (WMH)** volumes increased, so too did patients' risk of poor three-month poor post-procedure clinical outcomes. The scientists from **Montpellier** in France believe that there is a presumed relationship between vascular origins and cognitive dysfunction, depression, increased risk for stroke, higher rate of disability and death after stroke, which suggests that patients with increased WMH burden might respond less favorably to EVT. Read More

O-RADS Externally Validated for Differentiating Between Benign and Malignant Ovarian Lesions

Another article from the good folks at HealthImaging.com. Researchers from the Department of Medical Imaging at the **University of Toronto** recently completed an external validation for the **Ovarian-Adnexal Reporting and Data System (O-RADS)** for classifying lesions as malignant or benign, for which validation had been previously lacking. To the authors' knowledge, there are no current studies that have examined the system's efficiency at distinguishing between malignant and benign lesions using surgical and nonoperative treatment as the reference standard. Ultrasound were conducted by sonographers in North America. Read More

Artificial Intelligence to Assess Dementia Risk and Enhance the Effectiveness of Depression Treatments

Even though **EEG** is the most used brain imaging technique in the world, **magnetoencephalography (or MEG)** - which measures the magnetic fields generated by the brain's electrical activity - provides signals that are easier to interpret than EEG, because the skull and other tissues don't distort magnetic fields as much. Researchers from Aalto University in Helsinki tell us that **MEG coupled with AI** shows that communication between neurons begins to deteriorate well before the initial clinical symptoms of dementia. Inversely, magnetic stimulation is used to treat severe depression and neuropathic pain. **Read More**

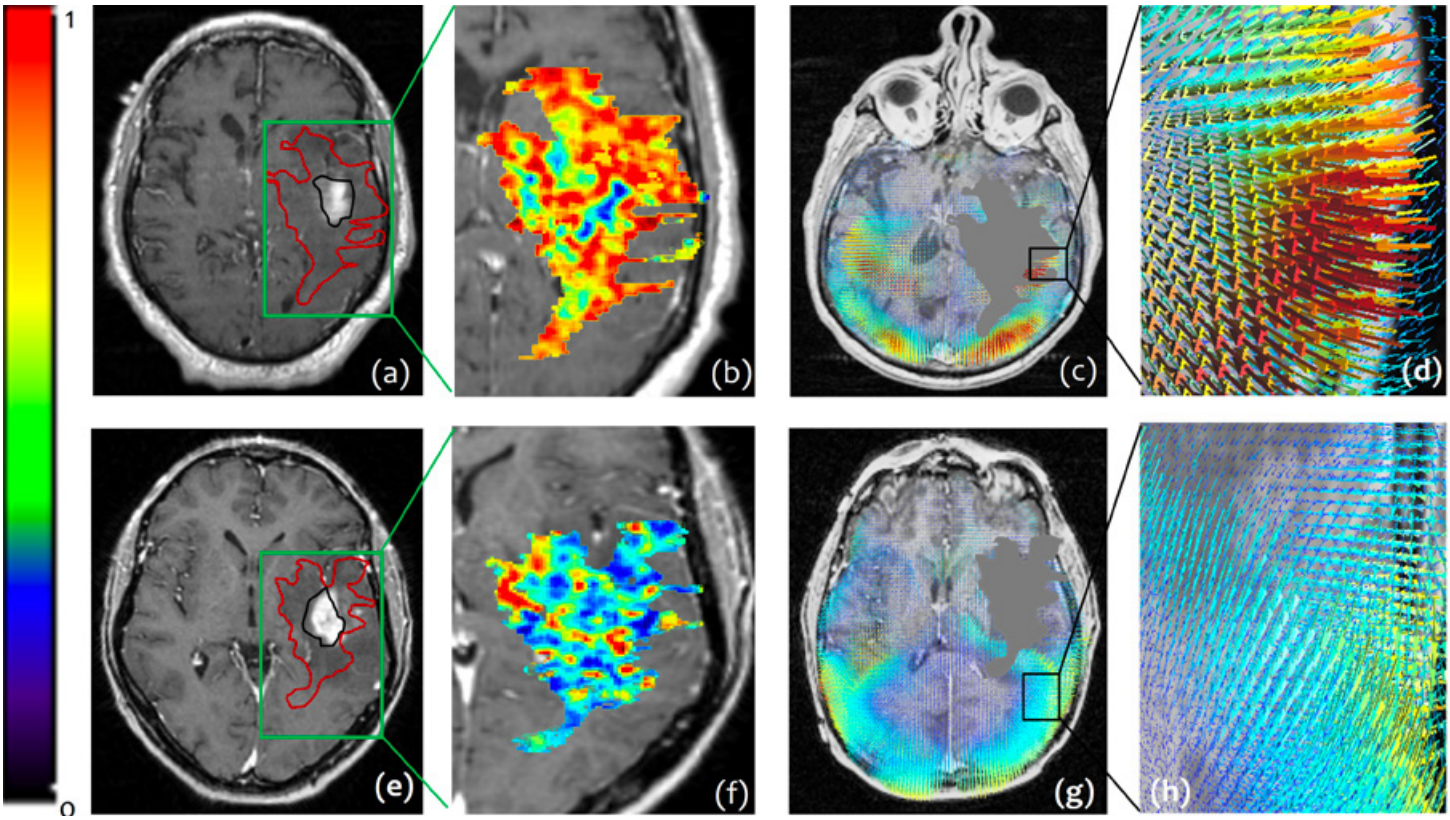
Pallavi Tiwari is an Assistant Professor of Biomedical Engineering and Director of the Brain Image Computing (BrIC) laboratory at Case Western Reserve University. She speaks to us about their innovative work using AI and machine learning in the treatment of brain tumors.

Can you tell us about your work?

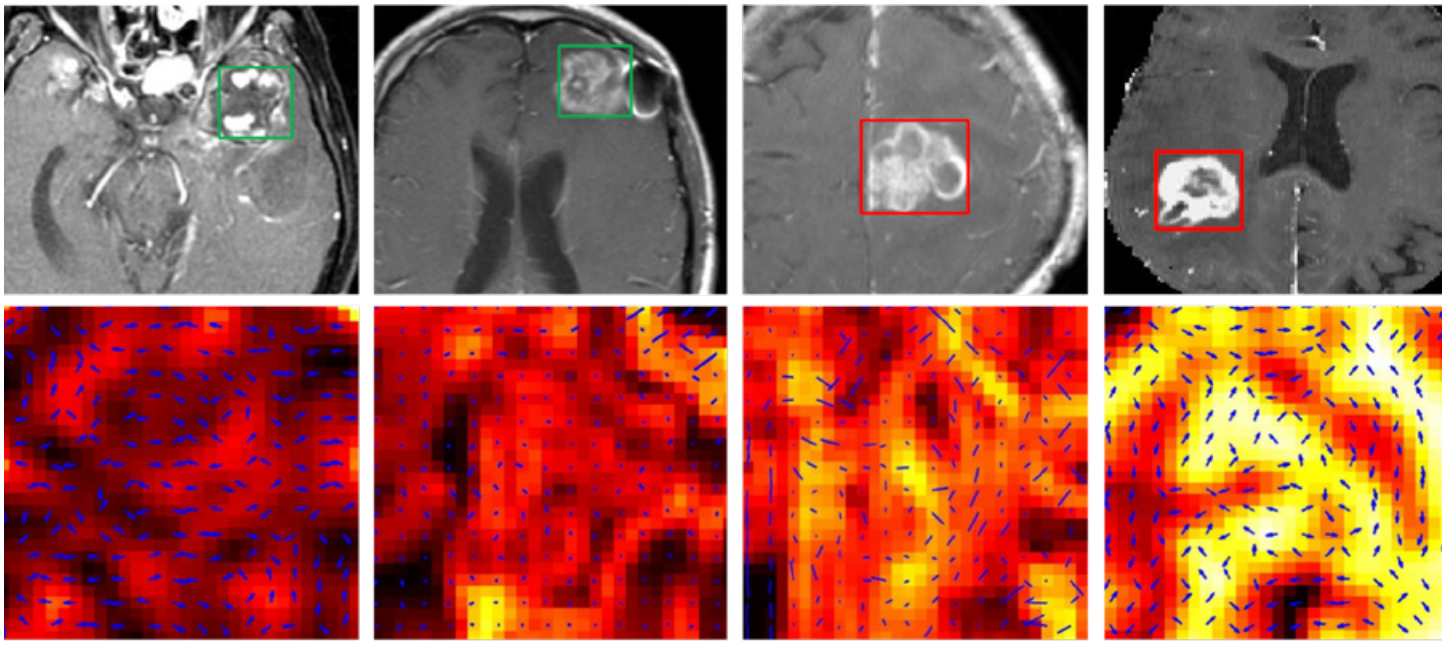
My lab is developing AI and machine learning-driven approaches for the diagnosis, prediction, and treatment evaluation of neurological disorders, particularly brain tumors. Recently we've been looking at post-treatment response assessment. Glioblastoma, one of the most aggressive and highly malignant forms of brain tumor, has a median survival rate of 15 months. Patients know there will be recurrence at some point after treatment. They're asked to come back for follow-up MRI scans, and if a suspicious lesion is seen, the question is, is it tumor recurrence, or is it just a side effect of the aggressive radiation they've received? The only definitive diagnosis comes from surgical resection. Some of these patients, who have already undergone chemoradiation treatment, must undergo another surgery only to realize it was a benign condition and could have been avoided. That affects their quality of life, overall outcomes, and survival.

How do you solve it?

We use AI and machine learning on routinely acquired MRI scans. We don't look at any fancy imaging. That has allowed us to collaborate with institutions globally to obtain MRI scans. We're close to 90%



Work from Tiwari's group published in IEEE Transactions of Medical Imaging (Ismail et al, 2022). Top row shows a GBM patient with an overall survival of 30 days, while the bottom row shows a GBM patient with an overall survival of 691 days. (b), (f) demonstrate the COLLAGE heatmaps generated for each of the two subjects, with higher values (red) being more prevalent in the patient with poor survival, compared to the patient with prolonged survival. (c), (g) illustrate the extracted deformation field magnitudes respectively for each of the two patients. For the patient with poor survival (d), higher magnitude values (represented in red) were observed in close proximity of the tumor, whereas lower values were observed (blue) for the patient with prolonged survival (h).



accuracy for distinguishing between a benign condition or tumor recurrence. By comparison, for neuroradiologists, the accuracy is close to a coin toss.

How did you end up working with the brain?

My master's and PhD work at Rutgers was focused on prostate cancer, but I've always been fascinated by the brain. My cousin has been suffering from multiple sclerosis for several years. That was another reason for me to delve more into neurological disorders. Towards the end of my PhD, I started talking to a neurosurgeon, who told me how limited the options for brain tumor patients were. That resonated with me, so right after my PhD, I started working in this field.

How good is the community today at tackling brain tumors?

The clinical community still has many questions that need to be addressed. There have been many clinical trials but not much success so far. Multiple trials are coming up to determine if any alternative treatments could work. Compared to other cancers which have moved towards personalized treatments, brain tumor patients get standard chemoradiation. That's not changed in the last 25 years. Our group has been developing tools and AI and machine learning algorithms to begin to move the needle towards personalized medicine. The aim is to identify patients who might be more suitable for a specific treatment instead of giving them the same aggressive dose of chemoradiation that may be unnecessary.

How is the machine learning community getting involved?

The technical community is recognizing these challenges more and more. A prominent example of that is the BraTS challenge through MICCAI, which has been running for ten years now. Through BraTS and other ongoing brain tumor challenges, more data is being made available, which hugely benefits our community and helps us answer these more challenging questions.

How fast do advancements translate into practical applications in real life?

A big challenge is that groups are working on these problems in a silo. Multiple people must come together both in terms of datasets and algorithms for us to start translating this work into clinical practice. We need a shared resource where we can test each other's algorithms. To some extent, that's started to happen now with federated learning and Spyros Bakas's work at the University of Pennsylvania.

Are brain tumors different for you as a medical imaging expert from other lesions that the human body may have?

One thing that's challenging about brain tumors is that they're highly heterogeneous. We presented a paper at MICCAI in 2014 and then published a journal paper in 2016 about a CoLlAGe feature descriptor that captures this intra-tumoral heterogeneity locally. The techniques we're developing

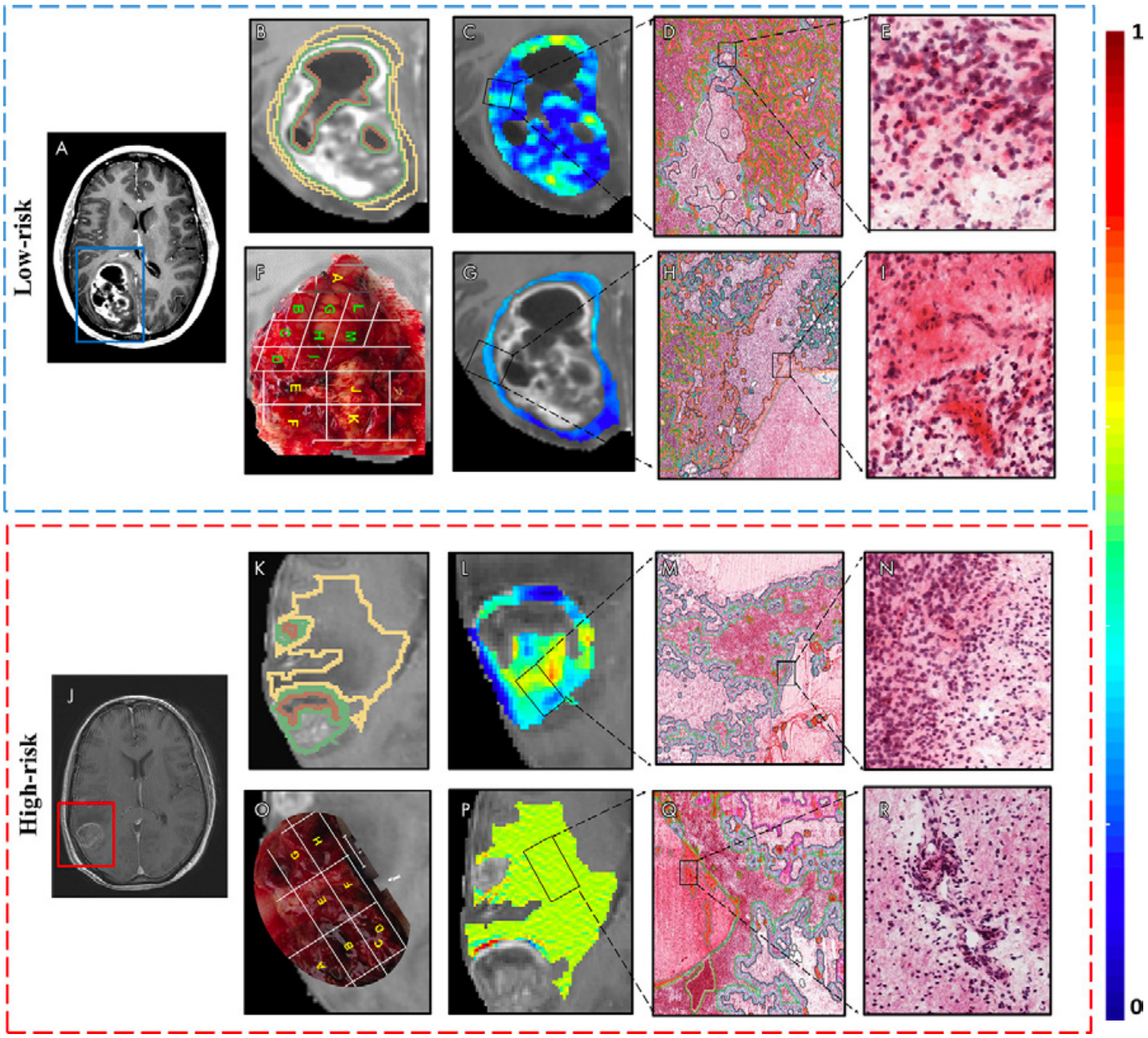
"We need a shared resource where we can test each other's algorithms."

are trying to capture this heterogeneity to predict outcomes in treatment-naive patients and patients after treatment.

Are you able to predict malignancy and the growth patterns of tumors?

When we think about a brain tumor, it's like a ball sitting in the patient's head which is expanding. This ball impacts the rest of the brain in what is known as mass effect. If the tumor grows quickly, this mass effect will impact the rest of the brain. We've developed a feature looking at the periphery of the tumor and the region outside it. It captures this mass effect on imaging and uses it to predict how long the patient will live. We compute it by measuring biophysical deformations on

Brain Imaging Computing lab (directed by Dr. Tiwari) in 2019 before the COVID-19 pandemic.



routine MR imaging. A paper just came out from our group in IEEE Transactions on Medical Imaging, where we've been exploring the impact that the tumor has on the entire brain. Not just thinking about heterogeneity or growth but bringing all these pieces together into a machine learning model and using that model to deduce patient outcomes and their response to specific treatments.

What are the most significant problems brain cancer still has for you and researchers like you?

That's a great question! There are so many problems I wish we could solve. We've already been building solutions for the issue of distinguishing between benign radiation effects and tumor recurrence so that patients can avoid unnecessary aggressive treatment. Another problem is picking the right treatment for the patient. Experimental treatments such as Immunotherapy often don't work for most brain tumor patients but do for some and have been shown to improve their quality of life and extended their survival. The

> **"The aim is to identify patients who might be more suitable for a specific treatment instead of giving them the same aggressive dose of chemoradiation that may be unnecessary."**

challenge is to determine which patients will be suitable for immunotherapy or other combination therapies.

Will we see you at MICCAI in Singapore this year?

Most likely!

As one of its early members, can you tell us more about your work with Women in MICCAI?

I care deeply about increasing the visibility of women in MICCAI and medical image analysis. If any women are reading this and are interested in reaching out, I'd be happy to help them or connect them to the right people within the community who would be great resources for them to use.

Have you seen our Women in Science series?

Yes, it's great you put women on the front page and give them that visibility. That's also what we've tried to do with Women in MICCAI.

Do you have any final thoughts?

I want to give a shout-out to all my students, postdocs, and research associates because they are the driving force behind all of our work. Also, my clinical collaborators and everyone I work with. I'm glad brain tumor is a relatively rare disease given it's challenges, but this means we don't have as many patients per institution to work with. Machine learning models need big data. We've been fortunate to work with collaborators worldwide who send us scans to evaluate, and we continue to grow this collaboration. They deserve a big shout-out as well for all their help!

Vision Transformers in Medical Computer Vision

IOANNIS VALASAKIS, KING'S COLLEGE LONDON

@WIZOFE

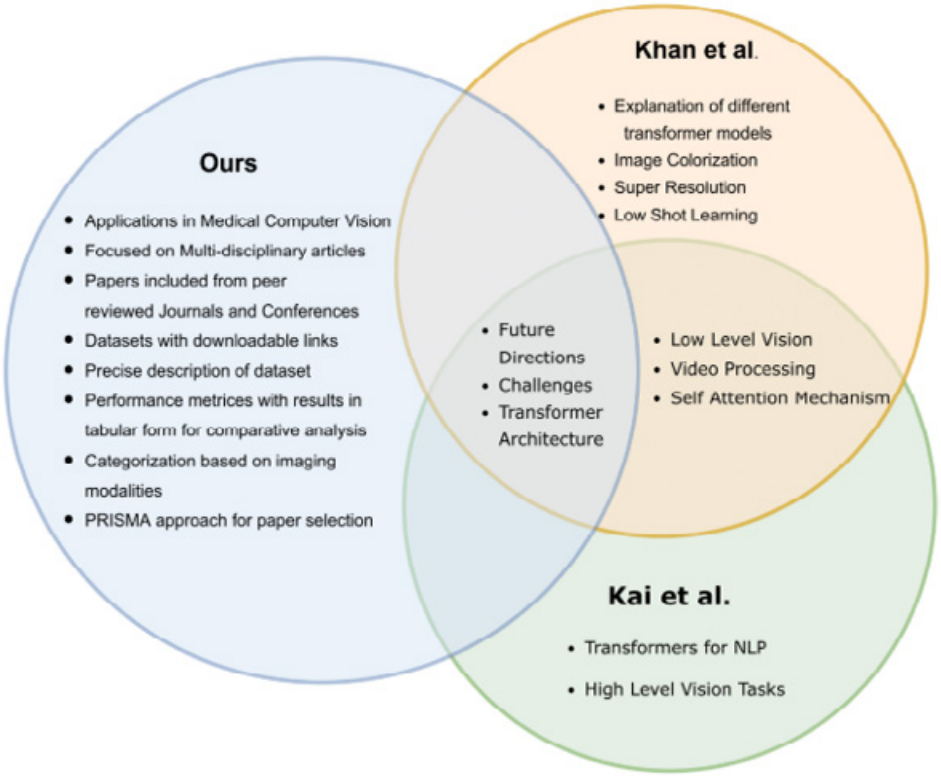
Hi everyone! This month we will review a paper writing about vision transformers! I will try to keep this article as simple as possible and to the point. I will link to the paper and other supplementary reading material at the end of the article and as always: be open, explore the original paper and create your own stories.

The title of the article is **Vision Transformers in Medical Computer Vision**, written by **Arshi Parvaiz** et al from **The School of Electrical Engineering and Computer Science** in Islamabad, Pakistan and it introduces the readers to what a vision transformer is how's used as a state-of-the-art approach in medical computer vision and pros and cons which are discussed in depth.

One of the reasons for choosing the article this month is my own curiosity on the subject and whether it's suitable for my own research on explainable classification neuroscience models. Are you curious? Me too!

Here we will give an overview before we start reviewing the article. I hope you enjoy!

Vision Transformers



Vision Transformers (ViTs) are evolved as one of the most contemporary and dominant architectures that are being used in the field of computer vision. These are immensely utilized by many researchers to perform new as well as former experiments. This article investigates the intersection of **Vision Transformers and Medical images** as an overview of various ViTs based frameworks that are being used by different researchers in order to decipher the obstacles in

Medical Computer Vision.

A survey is performed on the application of Vision transformers in different areas of medical computer vision such as image-based disease classification, anatomical structure segmentation, registration, region-based lesion detection, reconstruction using multiple medical imaging modalities that greatly assist in medical diagnosis and hence treatment process, and in the following figure from the paper you can see how different reviews have approached the subject.

Medical images contain an ample information that is the key for medical diagnosis and hence treatment. The healthcare data comprises 90% of imaging data, so considered as the primary source for medical intervention and analysis. Multiple medical imaging modalities such as Computed Tomography (CT), ultrasound, X-ray radiography, MR Imaging (MRI), and pathology are commonly used for medical imaging diagnostics. The analysis of these images by analysts is limited by human subjectivity, time constraints, and variation of interpretation.

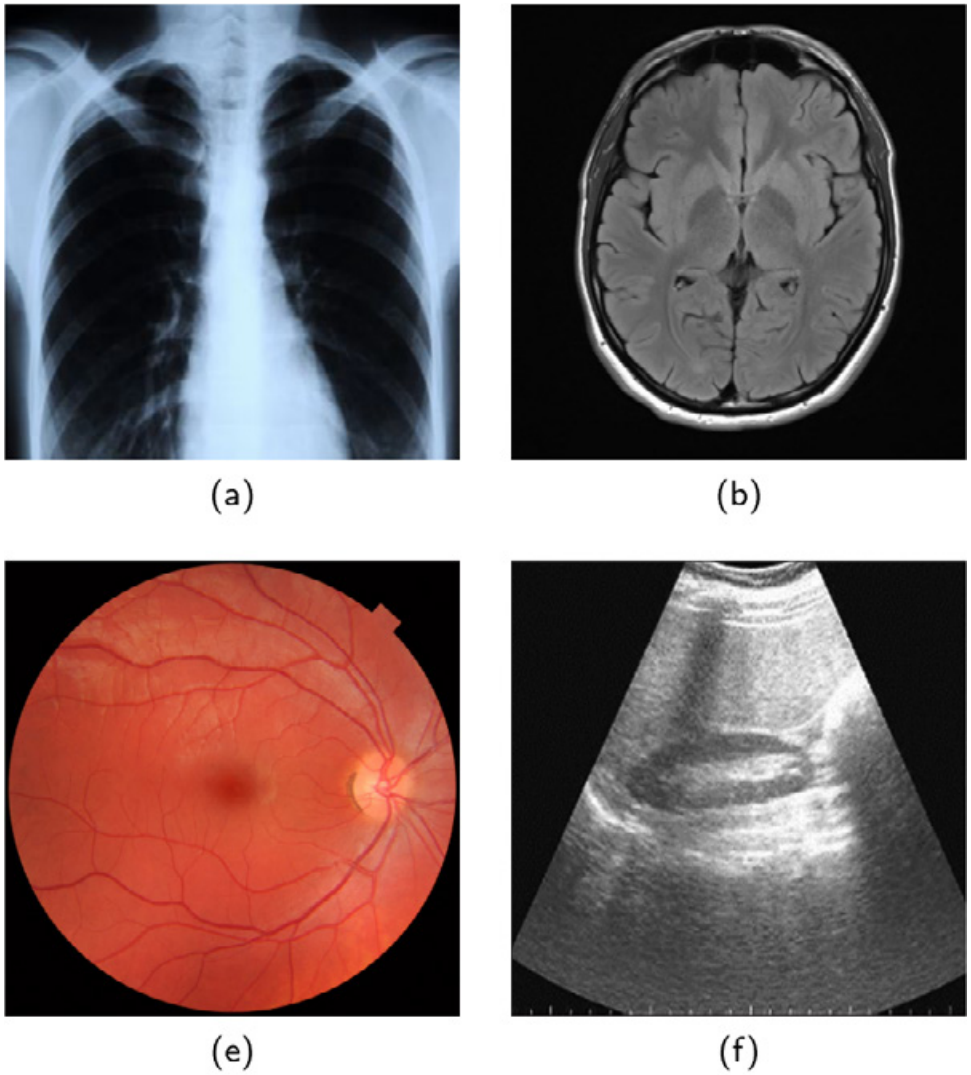
Several challenging factors associated with medical imaging modalities such as expensive data acquisition, dense pixel resolution, lack of standard image acquisition. Those techniques in terms of tool and scanning settings, modality-specific artefacts and hugely imbalanced data in negative and positive classes are major hindrance in translating AI based diagnosis into clinical practice.

Convolutional Neural Networks (CNN) are a type of deep learning architecture. CNNs are potentially the most popular deep learning architecture for its distinguished capabilities to exploit the spatial and temporal relationship between the features of images. CNNs have achieved notable accomplishment in medical imaging applications, such as, determining the presence and then identifying the type of malignancy (Classification), locating the patient's lesion (Detection), extracting the desired object (organ) from a medical image (Segmentation), placing separate images in a common frame of reference for comparing or integrating the information they contain (Registration), synthesizing images for balancing dataset (Generative Modeling).

CNNs are very good at feature extraction tasks.

CNNs lose the global context of the features. **Increasing the number of filters improves the representation capacity but at the cost of computation**. Various architectural changes are suggested by researchers for an efficient solution over time and leading to attention mechanisms. Using attention mechanism, regions of the image are captured, to which a CNN should pay attention, and forwarded to deeper layers. Researchers have demonstrated that replacing convolutional layer with attention has improved performance. It gets the best out of the attention mechanism to incorporate global context in the image features **without compromising computational efficiency**.

The potential of the vision transformers is further explored by many researchers for solving various problems. In this work it is highlighted the contribution of vision transformers to circumvent the challenges in automatic diagnostic of diseases using medical imaging modalities and their applications in medical computer vision tasks.



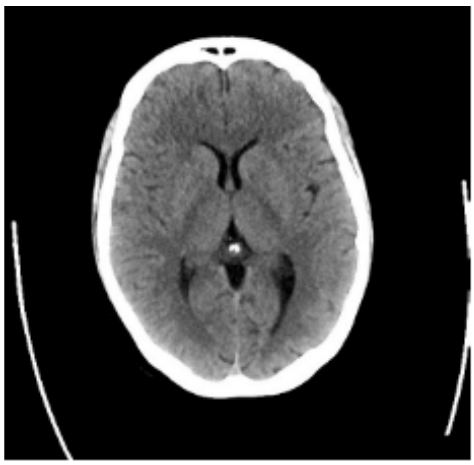
Medical images are used for medical diagnosis and treatment. Medical images can be acquired by various modalities, such as computed tomography (CT), ultrasound, X-ray radiography, MR imaging (MRI), Positron emission tomography-computed tomography (PET-CT), pathology fundus images and Optical coherence tomography (OCT). The images acquired from these modalities are shown in the figure below with more details about these image modalities in the following section.

In the previous figure clockwise: chest X-Rays, brain MRIs, fundus image on the rear eye for detection and grading of hypertensive retinopathy and liver ultrasound imaging techniques are shown.

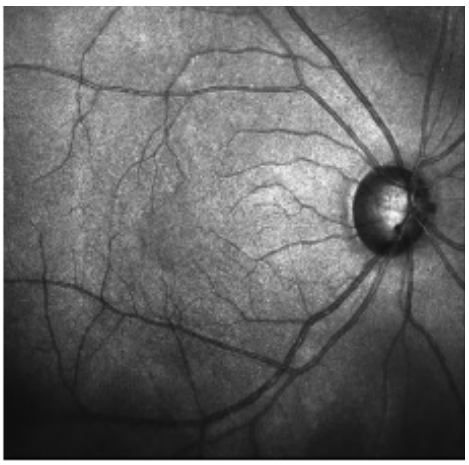
In the next figure brain CTs, such as those used to detect lesions tumors or clots, OCT images to diagnose retinal disease , whole slide images (WSI) used in computational pathology and PET-CT scans which are mainly used for cancer patients.

I will follow with a short description of each of those imaging modalities.

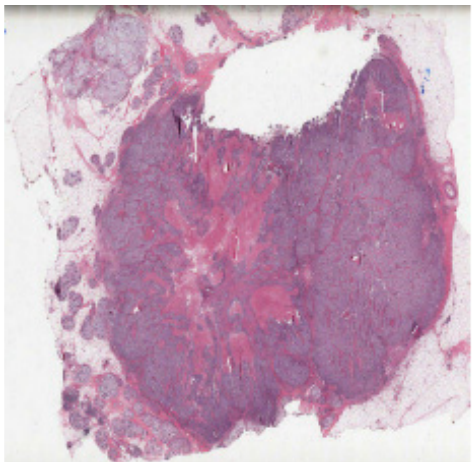
X-rays are non-invasive imaging techniques. X-rays are mostly used for diagnosing bone fracture. X-rays are also used by mammograms for breast cancer detection. Other most



(c)



(d)



(g)

(h)

familiar uses of X-rays are for breast tumors, enlarged heart, blocked blood vessels, conditions affecting your lungs, infections, osteoporosis, arthritis and tooth decay.

CT Scans are used for identifying the various organs/slices of the body. CT Scans are used for detecting cancer, tumors excess fluid, pulmonary embolisms (blood clots), lung infections and emphysema or pneumonia.

Optical coherence tomography (OCT) captures invasive cross-section images of the retina using light waves. OCT can be used to examine the retina's distinctive layers which help in mapping and measuring their thickness. OCT can be helpful in diagnosis of glaucoma, macular pucker, macular oedema, central serous retinopathy, diabetic retinopathy, macular hole.

Magnetic resonance imaging (MRI) can capture 3d anatomical images noninvasively. MRI scanning does not use any radiation which make it an ultimate choice of capturing when frequent imaging is required in the treatment process especially in the brain. MRI is particularly suitable for capturing the soft tissues of the body, but it is more costly as compared to x-rays and CT scanning. MRI can be used to capture different parts of the body for example MRI are used for diagnosis of aneurysms and tumors, as well for

differentiating between white matter and grey, in brain. There is a specialized MRI called functional Magnetic Resonance Imaging (fMRI), which is used for observing brain structure and locating the areas of the brain which are activated during cognitive tasks.

Ultrasound can be used to internal organs within the body, noninvasively. Ultrasound images are captured in 2D, 3D, but it can also capture 4D images which is 3D in motion such as a heart beating or blood flowing through blood vessels.

The **Whole-Slide Imaging (WSI)** refers to capturing the microscopic tissue specimens from a glass slide of biopsy or surgical specimen which results in high-resolution digitized images. Specimens on glass slides transformed into high-resolution digital files can be efficiently stored, accessed, analyzed, and shared with scientists from across the web using slide management technologies. WSI is changing the workflows of many laboratories.

PET scans can be used for cancer detection and diagnosis. PET scans can be used to determine spread of the cancer, determining the recurrence of cancer, metastasis, evaluating brain abnormalities like tumor and memory disorder. PET scans map normal human brain and heart function.

AI and healthcare analysts can use deep learning concepts, techniques, and architectures to bridge the gap between them.

Deep neural networks in computer vision have contributed to various fields of study. For instance, while assessing medical images, practitioners can recognize if there is an anomaly. CNNs generally consist of three kinds of layers: convolution layers, pooling layers, and full-connected layers. Convolution layers are responsible for learning features and capturing the Spatial and Temporal dependencies between the features by application of relevant filters. The pooling layer is responsible for reducing the size of feature maps to capture more semantic information than spatial information. Before the fully connected layer, the output of the convolutional and pooling layer is flattened to make a fully connected layer. A loss function is used to calculate the error and the is back propagated to update the values of learnable parameters. In recent years several CNN architectures are developed with various such arrangements: AlexNet, VGGNet, GoogleNet, ResNet, ResNeXt, Squeeze and Excitation Net, DenseNet, and EfficientNet.

Convolutional neural networks are used in various applications in the categories of image classification, detection, and segmentation, etc. They are known to be a black box, as the training is according to the task and domain. One major limitation is the unclarity of results i.e., the reason for a particular outcome.

One way to tackle this problem head-on is to have such a model that focuses on relevant parts of the image and can be visualized by the doctors. Attention models were proposed. Transformers consist of multiple blocks of identical encoders and decoders, which were composed of self-attention block and feed-forward networks. Transformers consist of an extra attention block, which focuses on the relevant part of the sequence. The performance of the transfer models was state-of-the-art in tasks related to natural language processing.

The model was called Vision Transformers (ViT). They divided the entire image into small

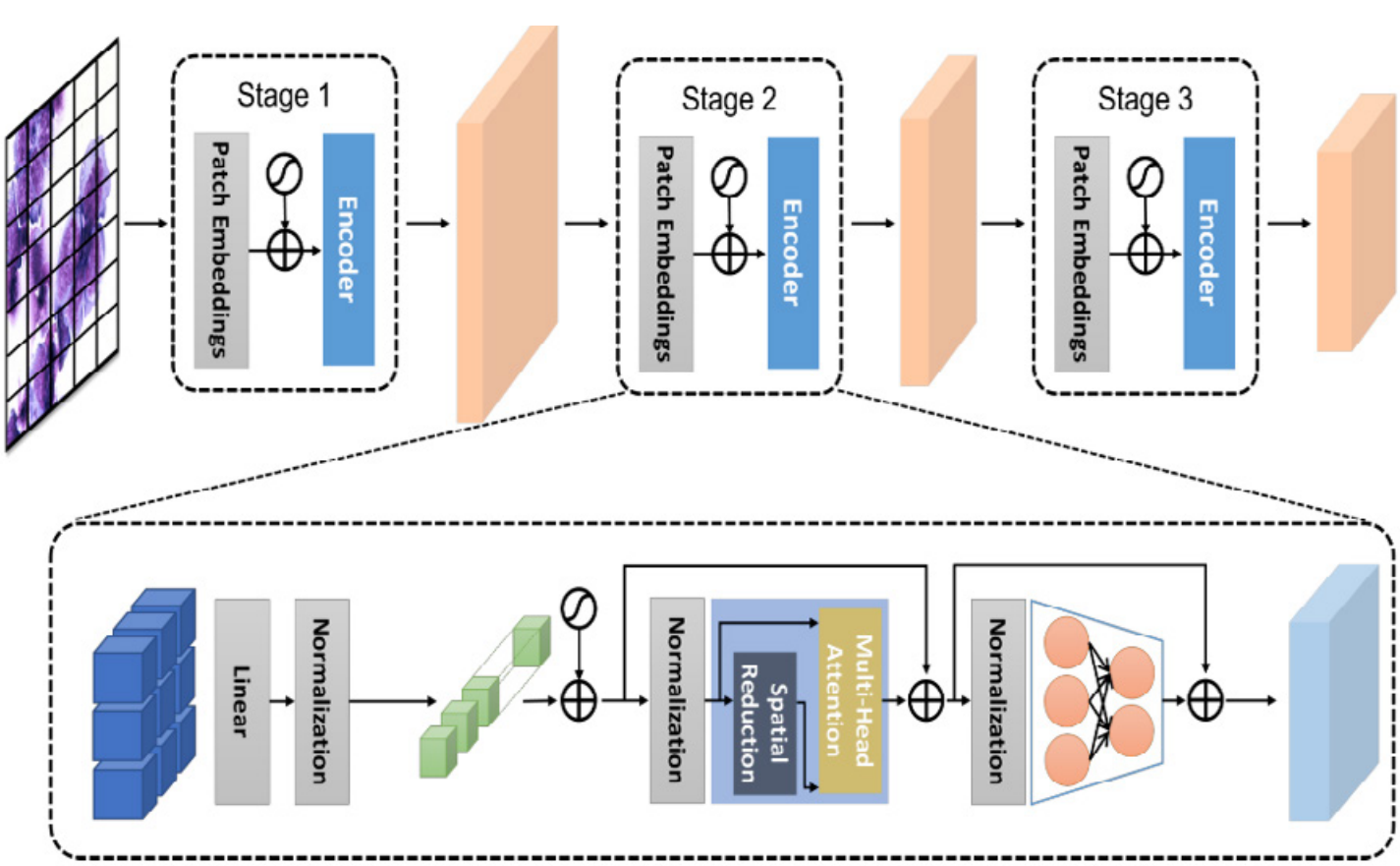
image patches of 16x16. They introduced simple numbers 1, 2, up to n as positional embeddings for specifying the positions of the patches.

Vision transformers have the capability of modelling global context which assists in more accurate results. Medical images are considered as the input for vision transformers.

For medical imaging classification the practitioners give their diagnosis by analyzing the medical images. Image classification has various applications in the medical domain. At present, image classification has various applications in the medical domain.

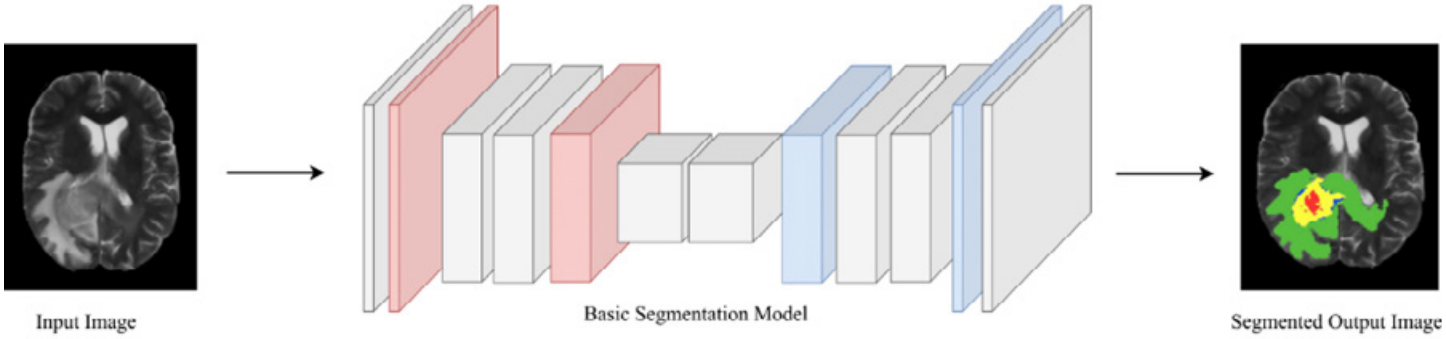
Image classification using CNNs can be used for various applications. The applications were achieved through various CNN architectures such as AlexNet, VGGNet GoogleNet , ResNet. More resource-efficient architectures were proposed i.e. MobileNet, Squeeze and Excitation Net, and EfficientNet.

Convolutional Neural Networks (CNNs) have been the most dominant deep neural networks for autonomous medical image analysis applications such as image classification during the last decade. These models, however, have shown poor performance in learning the long-range information, due to their localized receptive field. Transformer architecture, proposed by Vaswani et al., is currently the most popular model in the field of natural language processing (NLP). Self-attention modules are used in these models to learn the relationship between the embedded patches. In these models, the overall training process is predicated on dividing the input image into patches and considering each embedded patch as a word in NLP. They propose three different approaches to summarization. The performance matrices summarize the performance of each approach. They include a table



that summarizes the performance matrix. In the following figure there's a representation of visual transformers.

In the case of segmentation ViTs have proven very successful. Clear cut and detailed segmentation are a decisive step in image guided treatment and computer-aided diagnosis. A great deal of image segmentation models has been proposed. In the last 40 years, from traditional models to deep neural networks, they have outperformed all the state of art segmentation models. Transformers functions prominently in error free segmentation of medical images because of their capability to model the global context. As the organs lay out over a wide receptive field, hence, transformers can easily encode these organs by modelling the association of pixels that are distant spatially. The background is dispersed in medical scans; for that reason, gaining the understanding of the global context between those pixels that relate to the background will be beneficial for the model to do the unerring classification, as sheen below.



Xiong et al. proposed a novel hierarchical neural network architecture using reinforcement learning to generate a long coherent medical report. They incorporated the self-critical reinforcement learning method into the detector, encoder, and captioning decoder. They used DenseNet-121, pre-trained on chest X-ray 14 dataset, to detect the region of interest (ROI) proposals using a bottom-up attention mechanism. The region detector outputted a set of ROI proposals along with classified classes and some associated attributes. They used top-down transformer visual encoder to extract further pixel-wise visual information from proposed ROI using pooling operations. Their proposed architecture outperformed the state-of-the-art methods for the CIDEr evaluation metric on the IU-Xray dataset, but for the BELU 1 metric, their model could not perform state of the art. Their model over-fitted as they used only the findings portion of the generated medical report. They need to use a larger labelled dataset to solve this problem.

There are different ways to approach the transformations:

- Generative vision transformer based unsupervised MRI reconstruction architecture to increase the receptive field.
- Generative non-linear mapping over latent and noisy space to improve invertibility of the model.
- Cross attention to improve context of image features. Extensive experiments on accelerated multi-contrast brain MRI dataset.

They proposed an ASFT network to reconstruct the high-resolution MRI scans from low resolution scans. They introduced a multi-branch features transformation and extraction (MFTE) block. They filtered out the useless information using MFTE block. Their model achieved the state-of-the-art performance for super-resolution task.

Medical Image Synthesis: Tissue morphology information acquired from multimodal medical images play an important role in the clinical practice. GAN is a CNN based architecture that shows locality bias and spatial invariance across all the positions. Double-scale GAN showed efficient performance on benchmark IXI MRI dataset.

The authors propose a dual transformer network (DTN) model for the diffeomorphic registration of MR images. DTN uses self-attention mechanisms to facilitate contextual correspondence between anatomies. DTN has two branches to learn the relevance based on the embeddings of separate one-channel images and concatenated two-channel images. DTN has two branches to learn the relevance based on the embeddings of separate one-channel images and concatenated two-channel images. DTN uses feature enhancement, based on global correspondence, to infer the velocity field and registration filed. DTN is used to optimize metric space. DTN is unsupervised.

Vision Transformers (ViT) are now one of the hottest topics in the discipline of computer vision. Although CNNs are matured enough for the development of applications that can ensure an efficient and accurate diagnosis, in the medical field - where an inaccurate output might endanger lives - the concept of attention in vision transformers has paved its way for more precise outcomes. A variety of approaches have been proposed in recent years to explore and utilize the competency of vision transformers. These approaches showed excellent performance on a wide range of visual recognition tasks, including classification, lesions detection, anatomical structure segmentation, and clinical report generation. Nevertheless, the real potential of transformers for computer vision has yet to be fully explored.

Next month

Let’s meet next month with the continuation of the previous month’s coding article. I hope that you tried some of the ideas presented, as now we’ll see a more hands on approach. Until next month, have a great time and always be curious! ☺

PI-RADS ASSISTANT: NEW AI TOOL FOR PROSTATE MRI ANALYSIS

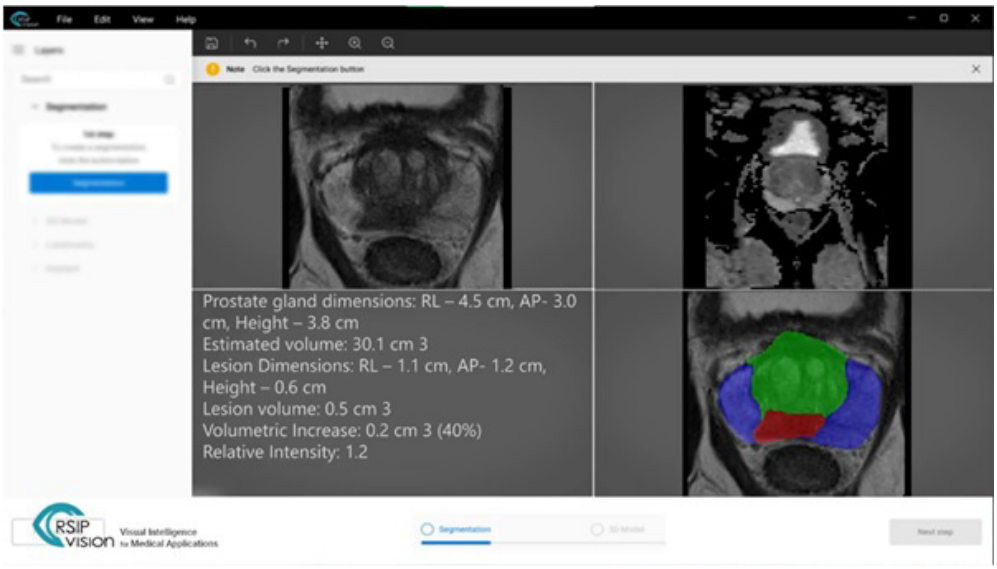
In a recent article published on our website, we described the challenges in Prostate Imaging–Reporting and Data System (PI-RADS) scoring, and that Artificial Intelligence (AI) plays a critical role in facing these challenges. We are now excited to share RSIP Vision's solution: the PI-RADS Assistant.

by Oren Wintner

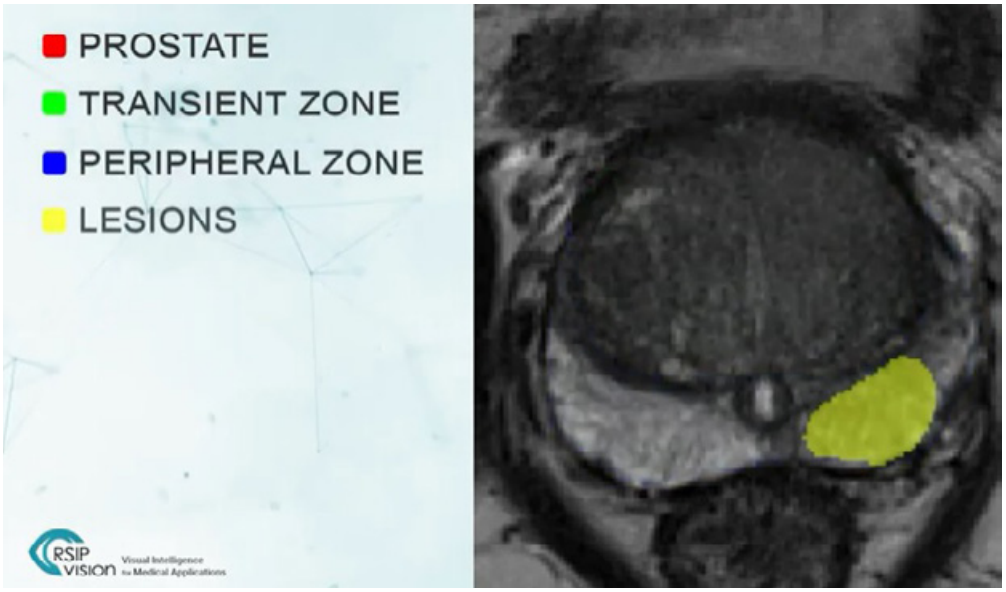
Prostate cancer is the most common male-cancer, affecting nearly %10 of the population. Similar to all cancers, early diagnosis and treatment significantly improve survival rates and reduce complications. **Diagnosis of prostatic cancer requires several clinical and imaging tests**. Firstly, blood PSA level is tested. Imaging tests usually involve ultrasound (US) and MRI scans. To construct an objective measure for risk of prostate cancer, **the Prostate Imaging–Reporting and Data System (PI-RADS) scoring system was developed**. It utilizes several MRI sequences and specific lesion characteristics to provide a numerical score for risk assessment.

However, **calculating the PI-RADS score is time-consuming and subjective**. Image quality, which depends on technician experience and hardware, highly affects the lesion characteristics. Also, each radiologist views the MRI sequences differently, and determines the lesion based on personal experience, leading to inter-user variability. This process takes time, and as this condition is prevalent, there are many prostate MRI scans to view, taking up a large portion of the radiologists' workload.

RSIP Vision has recently developed a new tool to assist in PI-RADS scoring. This tool performs segmentation of the prostate, its sub-sections, and lesions. It analyzes the lesions' intensity, restriction, size, and shape, and provides a baseline for Prostate Imaging–Reporting and Data System score. This information can be used



by the radiologist **to assess cancer risk and calculate the PI-RADS score faster and in a robust manner**.



There are several steps for the development of the PI-RADS assistant. Initially, image preprocessing was conducted to provide adequate input to the system. T2 and ADC image protocols undergo resolution equalization and normalization. Also, gentle registration of the protocols is conducted to compensate for patient movement throughout the MRI scan.

Using deep learning (DL) algorithms, the processed T2 sequence image data is passed through a **U-Net architecture neural network**, whose output is the segmented regions: whole prostate, transition zone (TZ), and peripheral zone (PZ). A separate network with similar architecture is used for lesion segmentation, utilizing both T2 and ADC sequences.

Standard methods are used for calculation of the parameters relevant for the PI-RADS score. Size and volume are retrieved directly from the segmentation mask. Edge smoothness is calculated by the gradient along the mask's edge, and it also is fitted to an oval shape and results in an ovalness score. Mean intensity and standard deviation is calculated within the lesion in all sequences to support the scoring protocol. Lesion position was detected by comparing the segmentation labels' position to determine lesion restriction.

As our current dataset includes T2 and ADC sequences only (**The ProstateX dataset**), with limited resolution, **we expect to improve performance by introducing other sequences with higher resolution**. Increasing the dataset size is also expected to have an effect on performance.

As described previously, this tool can be expanded to other risk-scoring methods: **BI-RADS (for breast cancer), TI-RADS (for thyroid cancer)**, etc. Adjustments to the different imaging modalities need to be made, but the concept remains similar - automating region detection and calculation, to save scan reviewing time. Implementing AI methodology in these fields will significantly reduce radiologist workload and improve patient care. As AI is slowly penetrating various use-cases in healthcare, cancer screening should benefit as well. **Read more about AI in urology.**

Congrats, Doctor!

Elisabeth Preuhs (née Hoppe) recently completed her PhD at the Pattern Recognition Lab at Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) in close collaboration with the Magnetic Resonance (MR) predevelopment of Siemens Healthineers. Her research focused on novel methods for multidimensional MR data to enable an extensive and quantitative analysis of, e.g., brain tissues or the beating heart. For her pioneer work in this field, she was awarded as an Artificial Intelligence Newcomer of 2019 by Deutsche Gesellschaft für Informatik, as well as with the Innovator Award 2020 from the FAU. Currently, she is working in the innovation department of Siemens Healthineers for minimally invasive interventions in the scope of new robotic-assisted workflows. Congrats, Doctor Elisabeth!

The application of deep learning (DL) increased tremendously in recent years. In my PhD work I made use of these technique for the improvement of medical MR Imaging (MRI), especially to enable a quantitative reconstruction (i.e., maps with physical T1 and T2 relaxation times) of MR data, as well as to facilitate the workflow of dynamic cardiac MRI. Using such quantitative and dynamic MRI instead of the conventional qualitative MRI is beneficial for a thorough analysis of tissues (e.g., differentiation of normal and pathological tissues), as well as for the simultaneous evaluation of the function of moving tissues like the heart.

In a first project, DL for the reconstruction of so-called MR Fingerprinting (MRF) data was applied. MRF acquires multiple image contrasts of the same tissue with strong undersampling by modifying the parameters for each acquisition. The aim is to generate characteristic signals (so-called fingerprints) for different tissues. The State-of-the-Art (SOTA) reconstruction compares the measured fingerprints with a presimulated dictionary to derive the underlying quantitative maps of T1 and T2 values. This is a highly non-efficient method due to the exhaustive search. To overcome these limitations, recurrent neural networks (RNNs) for the MRF reconstruction [**1**] were introduced, which simultaneously handle spatial and temporal correlations

Figure 1

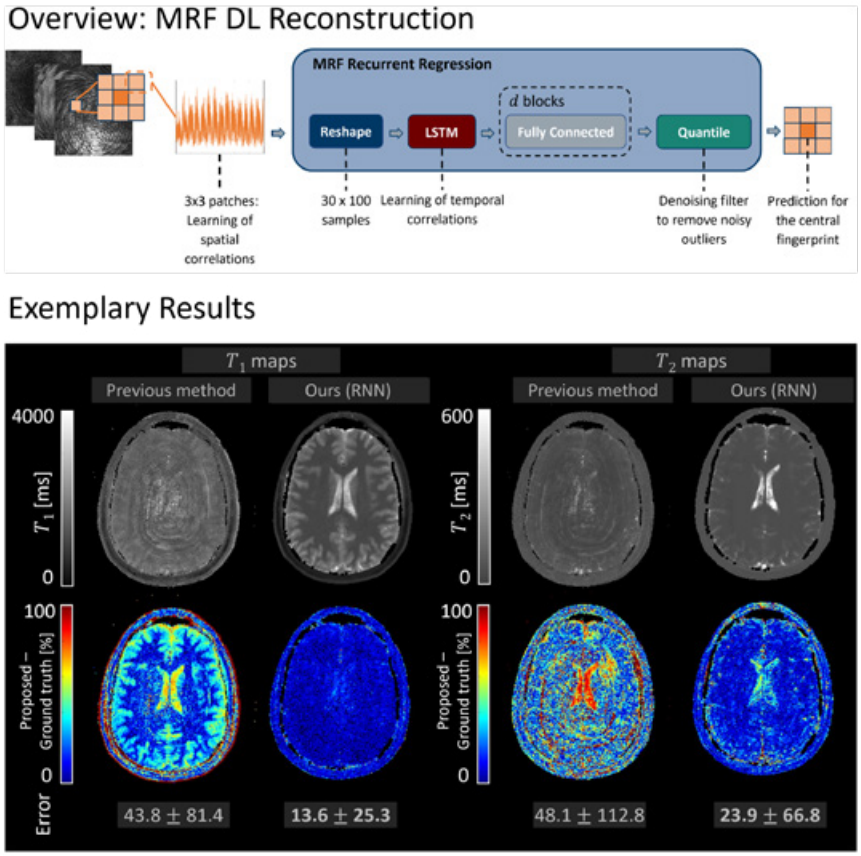
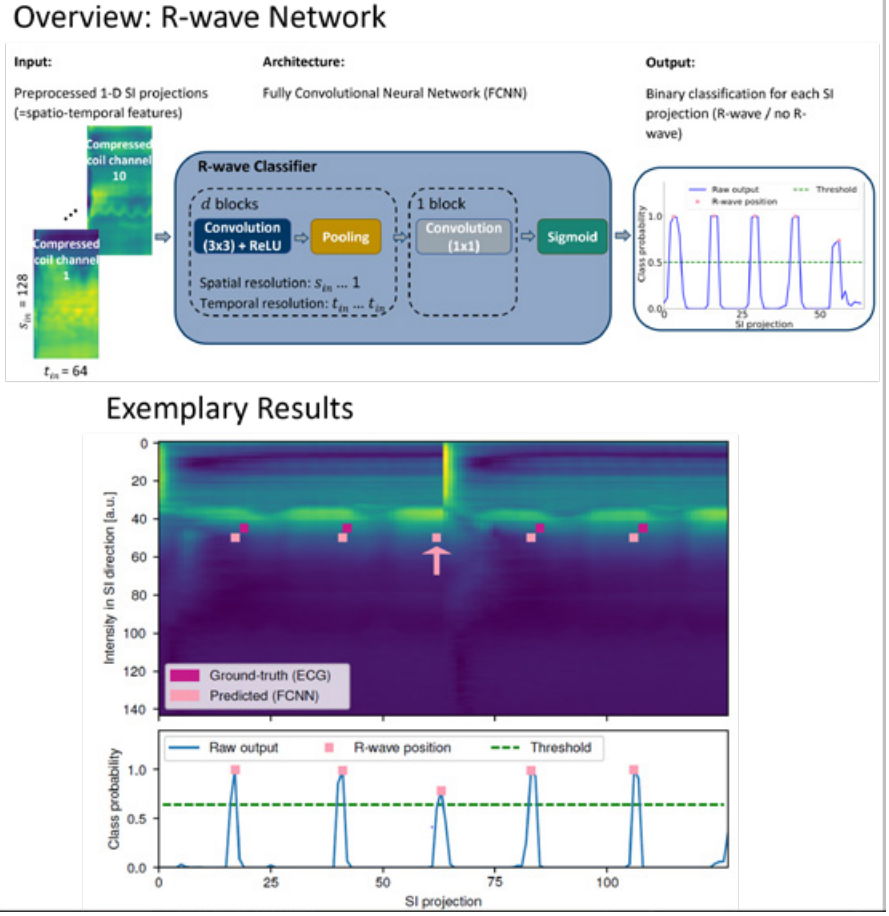


Figure 2



in the fingerprints (Fig. 1). The results show that RNNs reconstruct high detail maps with clearly reduced artifacts compared to previous DL methods (Fig. 1). In addition, the reconstruction time can be reduced up to a factor 30 with RNNs compared to the SOTA method.

In a second project, DL-based cardiac navigation for MRI of a beating heart was developed. A motion-resolved reconstruction (i.e., a movie of a beating heart) requires the knowledge of the cardiac motion in a scan. The ECG-sensor is the clinical SOTA to derive this motion, which is not always reliable within a MR scanner. My framework [**2**] directly uses parts of the acquired MRI data of a scan as input to a fully convolutional neural network (FCNN) which estimates the R-wave timepoints, i.e., the cardiac phases (Fig. 2). This simplifies the workflow, as no external ECG-sensor needs to be attached to the subject. The results also showed, that by using the features from the MRI data instead of the electrical ECG-signal, limitations of the ECG-signal (like missed R-waves because of its interference with the magnetic fields) could have been overcome (Fig. 2).

PREOPERATIVE TO INTRAOPERATIVE LAPAROSCOPY FUSION (P2ILF) CHALLENGE

Sharib Ali is a postdoctoral researcher at the University of Oxford and co-organizer of the MICCAI Preoperative to Intraoperative Laparoscopy Fusion (P2ILF) challenge, which aims to improve the complex task of fusing preoperative 3D CT/MRI scans with intraoperative 2D images in laparoscopic liver surgery. He speaks to us about the first edition of this exciting event.

Augmented reality-assisted laparoscopic liver surgery uses key landmark detection from intraoperative 2D video frames registered to a preoperative 3D liver model from CT/MRI data for tumor localization.

"*Sometimes tumors are not visible in laparoscopy images because they're embedded inside the liver,*" Sharib tells us.

"*However, these tumors can be visible in preoperative CT/MRI scans, so the idea is to* ***fuse the two in real-time to give the precise location of the tumor during surgery****. This allows the surgeon to resect it completely, reducing the risk of recurrence and ultimately saving lives.*"

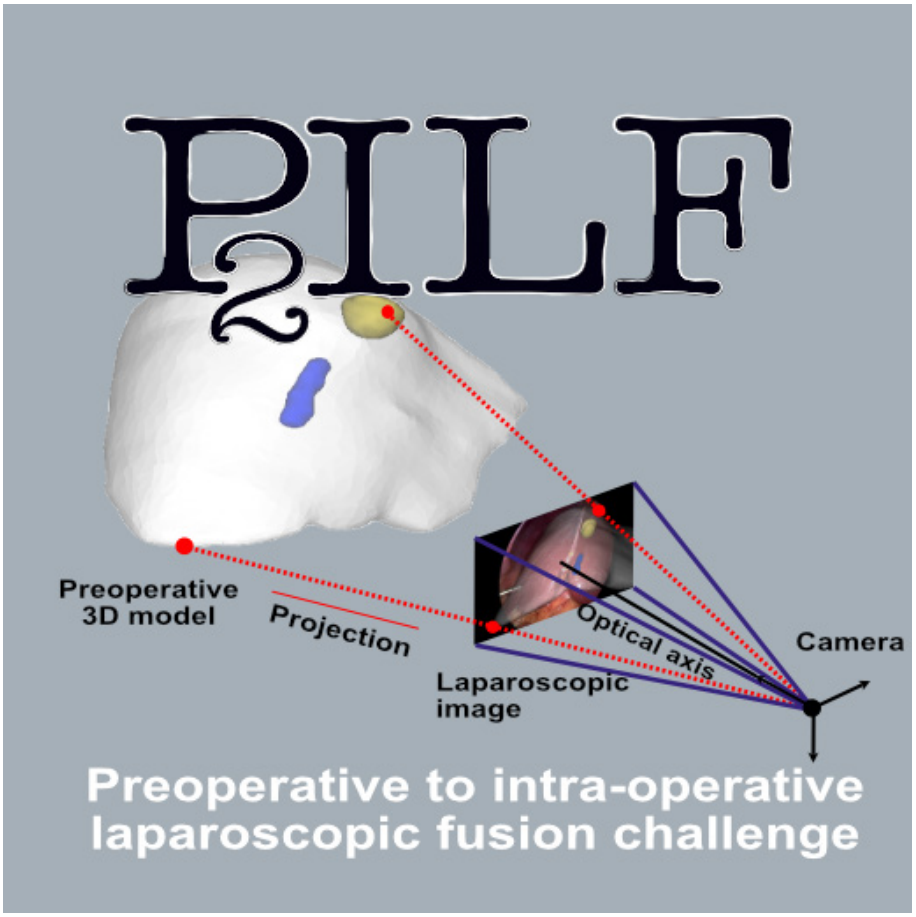
The P2ILF challenge asks participants to use machine learning methods for two tasks. They must first **segment five liver anatomical curves**, including silhouette, falciform ligament, left and right ridges, and liver boundary, from the 2D intraoperative images and preoperative 3D liver model. The second task is to **register those segmented curves in 2D laparoscopy** to the corresponding landmarks in the 3D model.

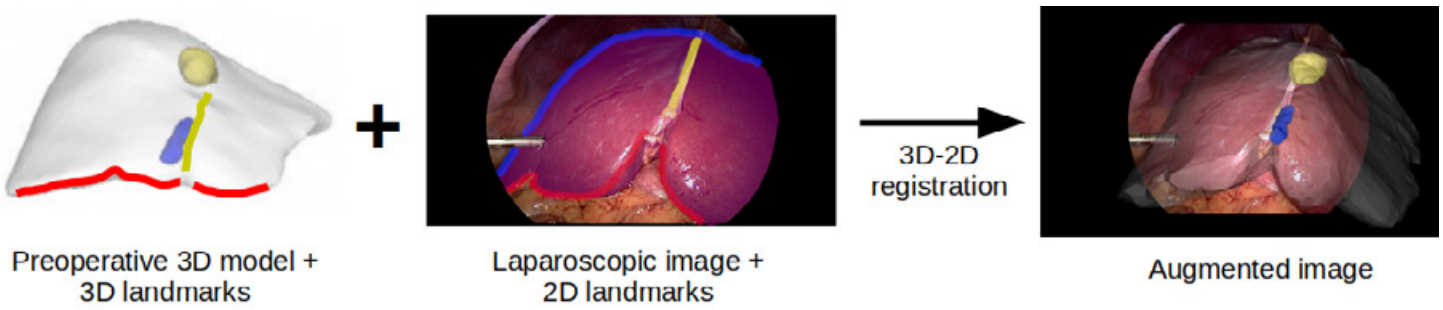
Most current methods use traditional computer vision methodologies to do this. Sharib approached **Adrien Bartoli**, Professor of Computer Science at Clermont Auvergne University, and discussed designing tasks using deep learning rather than a mathematical model.

The challenge uses single image frames, but he does not rule out using video in future editions. The plan is first to assess where the community is at in solving this kind of complex problem in the era of deep learning.

"*I don't want to scare people by saying it's complex,*" Sharib interjects.

"*The only way to approach it is to truly understand the problem and what you're trying to achieve. There are two tasks, but the registration task is particularly important and can be challenging with liver views at different positions and angles. People must take great care when* ***finding the landmarks and understand which***





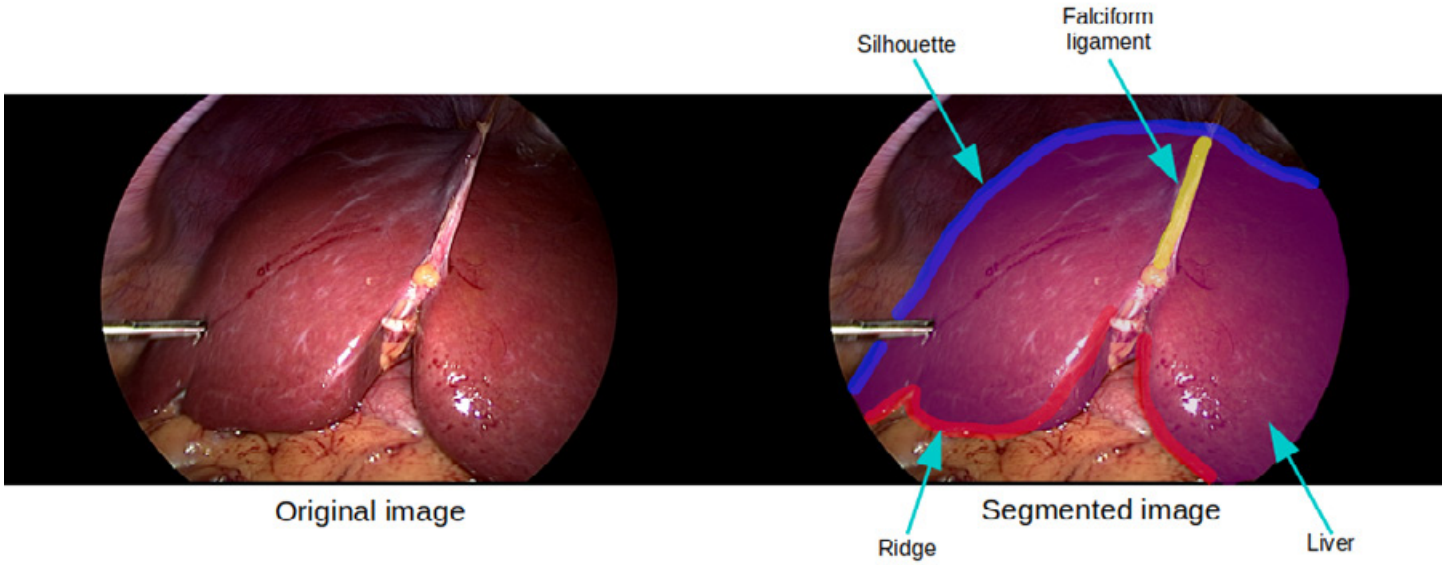
***landmarks are critical for registration and which are not**. If they always keep that in mind, they'll be successful!"*

Compared to other organs, the liver does not move much, which is one of the reasons the team picked it for this first edition of the challenge. However, preoperative to intraoperative laparoscopy fusion can be performed with other organs, and the plan is to explore this in future editions.

Sharib and his fellow organizers – **Adrien, Yueming Jin, Yamid Espinel López, and Lena Maier-Hein** – have put a great deal of thought into designing this challenge and curating the data. They are still working on collecting additional data from their clinical collaborators, which is all manually annotated to get the ground truth and provide a clear picture of the metric that will be used to evaluate the challenge.

The team recognizes the potential for metric failure and has introduced a 2% tolerance for the predicted anatomical curves with respect to the ground truth.

"We're doing our best to pose a great challenge with a metric that will be clinically



valuable, but at the same time, we don't want participants to be too disheartened if it's not working," Sharib points out.

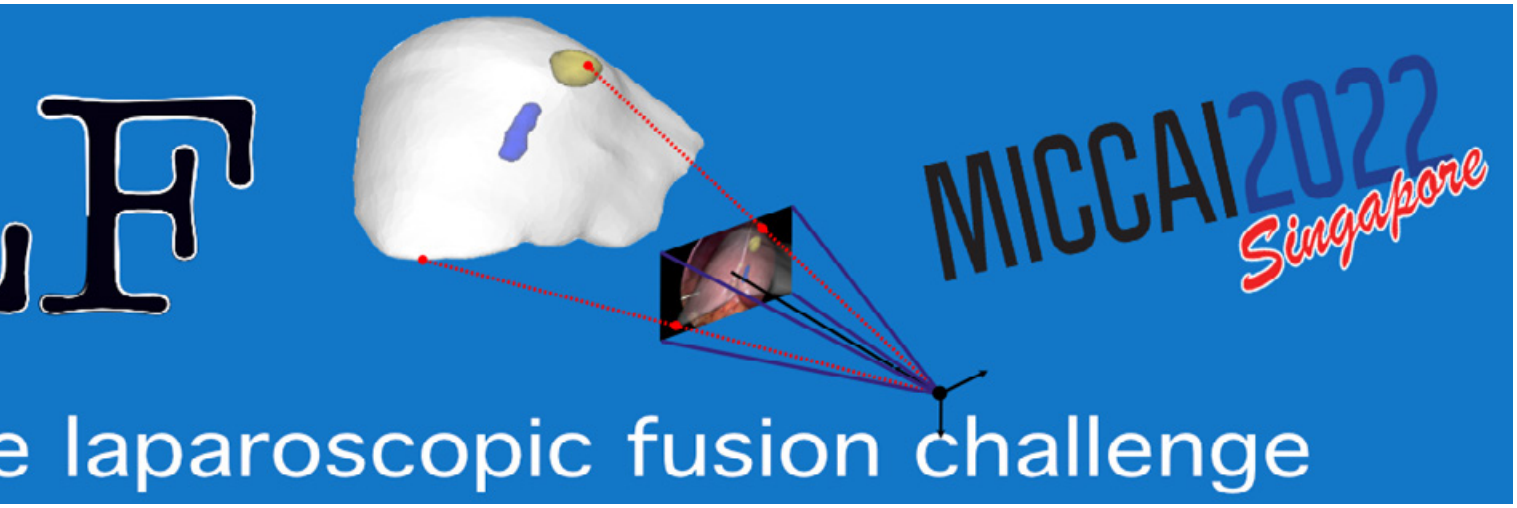
"*You just have to be close enough, and we'll compensate for the rest!*"

The final results of the challenge will be announced at **MICCAI in Singapore in September**, which will be its first in-person event since Shenzhen, China, in 2019 and the first MICCAI conference hosted in Southeast Asia.

Sharib adds that we have a responsibility to **tackle these complex problems together as a community**.

"*I'd like to invite the whole community working in this domain to contribute,*" he declares.

"*As a community, we can try to solve it, or at least understand where we're at. Independently, we can do our own research going forward too, but collectively, I think we'll have a much bigger impact on society. That's the best way to use our time as researchers.*"



BVM WORKSHOP 2022

Jasmin Metzger and Peter Neher are research scientists working in Klaus Maier-Hein's Medical Image Computing Division at the German Cancer Research Center (DKFZ). They are co-organizing this year's BVM Workshop 2022 to be held on June 26-28 in Heidelberg, Germany. Together with Klaus, they speak to us about what we can expect to see at the event next month.

Jasmin Metzge

Peter Neher

Klaus Maier-Hein

Established in 1993, the **BVM workshop** is an annual event where the mostly local computer vision community can come together. It is typically an informal and intimate affair, with around 100-150 participants, attracting young scientists and students who are looking to grow into the community, amongst more experienced members.

"*It's like a family meeting of the German medical image processing community,*" Peter tells us.

"*International events, which we all obviously want to go to, can be a bit overwhelming. This is one of the few situations where the whole German community can come together. That's the most appealing part.*"

Being a German event, many of the talks and much of the discussion is in German, but everyone is also free to present and speak in English.

Last year, BVM 2021 went virtual, after organizers planned for several scenarios in the face of an ever-changing pandemic. This year, the team have focused on organizing a physical event and, despite having to reschedule once already, are determined the show will go on and everyone will finally be able to meet in person.

Klaus tells us he is honored to be hosting the event in Heidelberg again:

"*Year after year, BVM continues to engage AI and imaging scientists from several countries and all career stages. This is an ideal conference to exchange the newest scientific ideas as well as personal updates in an extremely familiar atmosphere.*"

The scientific value of the event has been growing in recent years. Peter, who has been

attending BVM for a decade now and co-organized the workshop in 2013, has witnessed this change first-hand.

"*Organization of the event rotates every year, and we have new university partners and new professors now who have advertised it a lot more amongst their students,*" he points out.

"*In the beginning, BVM was more like a get-together and the scientific content was nice to have, but people didn't usually publish their most important results or findings here. Now, the spirit is changing, and I see particularly in the deep learning area that it's completely valid to publish your new results at BVM. That's reflected in this year's program.*"

Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211 (2021).
https://doi.org/10.1038/s41592-020-01008-z

Jasmin adds that **switching to different submission formats** has been another important change:

"*We have plenty of fresh work from young and older scientists, but we also have one-pager abstracts of papers which have already been accepted or published at other conferences or in journals.*"

There will be three keynotes with topics from local and international speakers, including **Mihaela van der Schaar from the University of Cambridge**, **Stefanie Speidel from NCT Dresden**, and **Ullrich Köthe from the Visual Learning Lab at the University of Heidelberg**.

Awards and prizes will be given out at the event for the best talks and posters, as voted for by attendees, as well as a selection of star submissions, chosen and ranked by the BVM Committee, including **Klaus Maier-Hein**, **Christoph Palm**, and **Andreas Maier**.

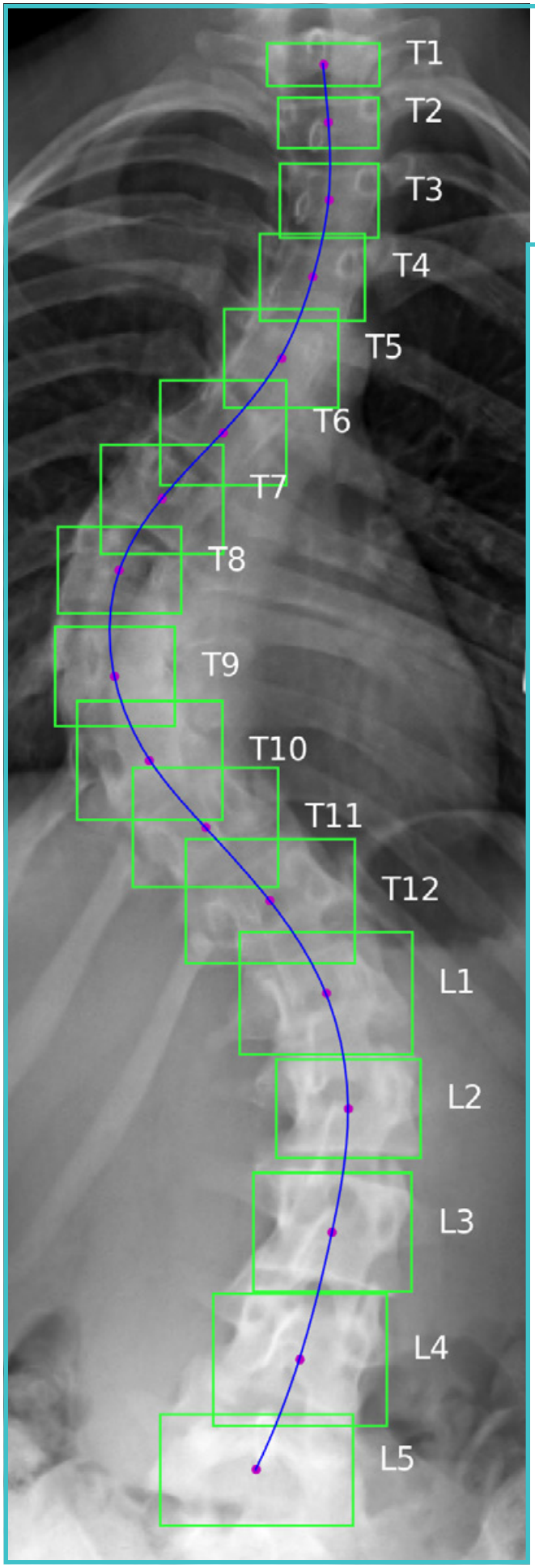
People have also been busy submitting applications for a special BVM Award for the best thesis in the field of medical image processing. This is separate from the event itself, but the award and a cash prize will be given out during the proceedings.

The main workshop runs on Monday and Tuesday, and on the Monday night, there is an exciting community event planned: a catered cruise along the Neckar River.

"*We very much hope we are allowed to do it,*" Peter tells us.

"*You're on a boat. No one can escape. You have to mingle!*" Jasmin laughs. But with such good German food and drink on offer, we do not think people will need much persuading.

"*Good food, but heavy food!*" Peter adds,



Sukesh, Richin; Fieselmann, Andreas; Jaganathan, Srikrishna; Shetty, Karthik; Kärgel, Rainer; Kordon, Florian ; Kappler, Steffen; Maier, Andreas K - Training Deep Learning Models for 2D Spine X-rays Using Synthetic Images and Annotations Created from 3D CT Volumes - In Bildverarbeitung für die Medizin 2022, edited by Christoph Palm, Thomas M. Deserno, Heinz Handels, Andreas Maier, Klaus Maier-Hein, and Thomas Tolxdorff. Informatik aktuell. Wiesbaden: Springer Fachmedien, 2022.*

laughing.

The **German Cancer Research Center** is part of the **Helmholtz Association**, one of the biggest research foundations in Germany, with strong ties to the **University of Heidelberg** and the **University Hospital Heidelberg**.

Peter's work at the **Division of Medical Image Computing** has been varied since completing his PhD, and he now coordinates one of its subgroups.

"*My area of expertise is **diffusion-weighted MR image processing**; reconstructing nerve fibres in the brain and analyzing them in case of pathologies,*" he tells us.

Jasmin went down a different route after getting her diploma. She did not do a PhD but has been working in research for 10 years now.

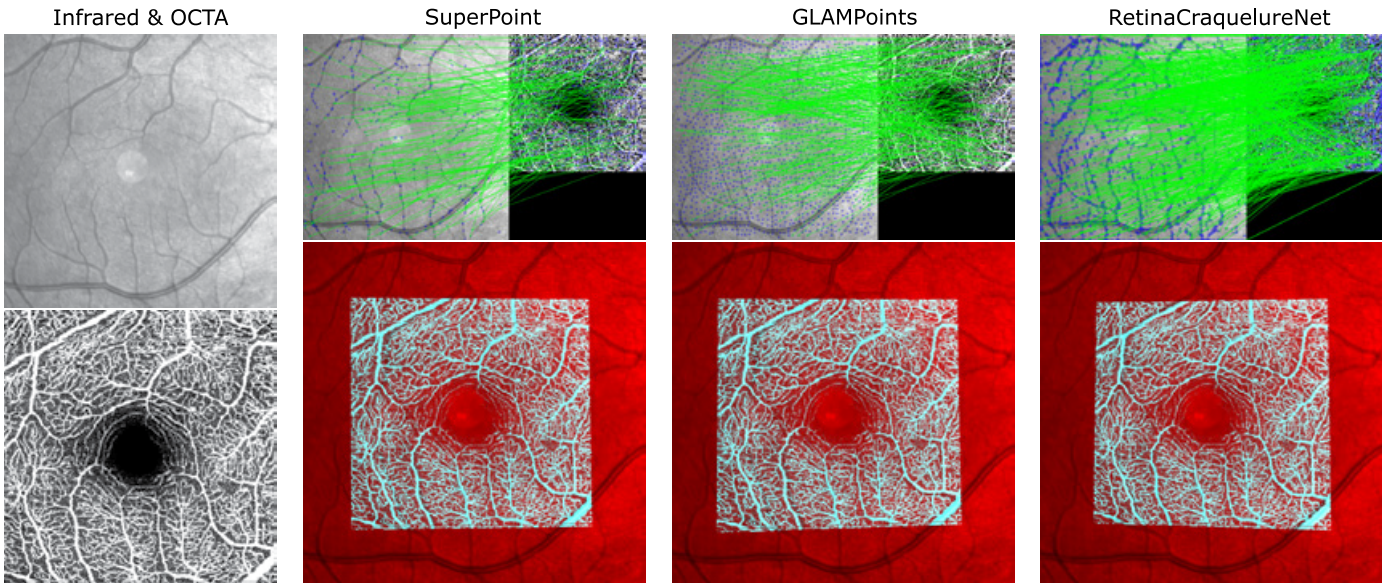
"*I'm more in the **project coordination** part,*" she says.

"*Since my parental leave last year, I've been working at the Kaapana platform, which is a platform for deploying AI methods to different sites. Currently we have this platform installed at a number of university clinics.*"

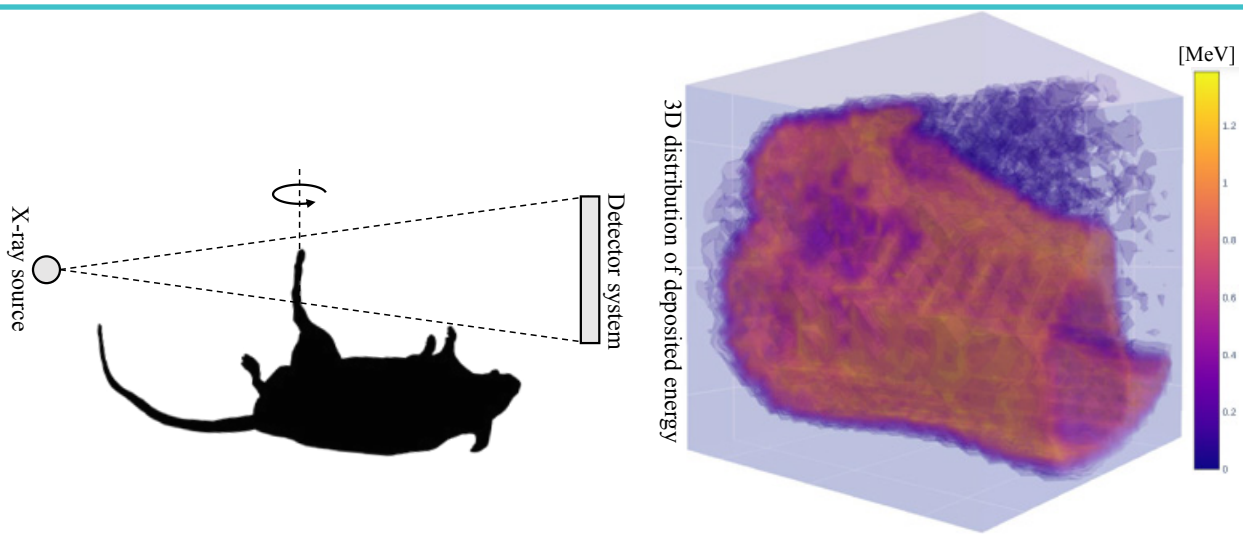
With so many young scientists and students in attendance at BVM, what advice do the team have for those young people who are attending their first event this year and want to make the best of it?

"*Just pick the parts that are really interesting for you and use the occasion to network with your peers,*" Peter advises.

"*To make extra space due to Covid, we've*



Sindel, Aline; Hohberger, Bettina; Fassihi Dehcordi, Sebastian; Mardin, Christian; Lämmer, Robert; Maier, Andreas K; Christlein, Vincent - A Keypoint Detection and Description Network Based on the Vessel Structure for Multi-Modal Retinal Image Registration - In Bildverarbeitung für die Medizin 2022, edited by Christoph Palm, Thomas M. Deserno, Heinz Handels, Andreas Maier, Klaus Maier-Hein, and Thomas Tolxdorff. Informatik aktuell. Wiesbaden: Springer Fachmedien, 2022.*



Wagner, Fabian; Thies, Mareike; Karolczak, Marek; Pechmann, Sabrina; Huang, Yixing; Gu, Mingxuan; Kling, Lasse; Weidner, Daniela; Aust, Oliver; Schett, Georg; Christiansen, Silke; Maier, Andreas K - Monte Carlo Dose Simulation for In-Vivo X-Ray Nanoscopy - In Bildverarbeitung für die Medizin 2022, edited by Christoph Palm, Thomas M. Deserno, Heinz Handels, Andreas Maier, Klaus Maier-Hein, and Thomas Tolxdorff. Informatik aktuell. Wiesbaden: Springer Fachmedien, 2022.*

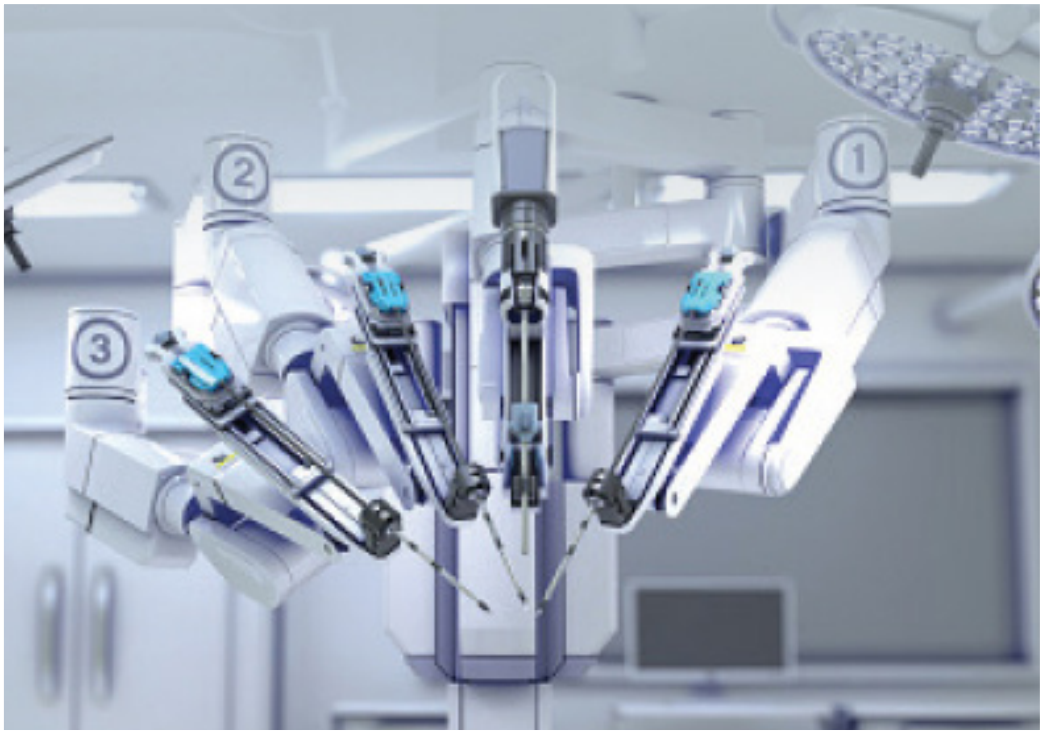
extended the poster sessions this year. There are three sessions with fewer posters each. This gives you the opportunity to really talk about the content, which is so much more important than hearing every talk and trying to understand every paper."

Jasmin adds:

"Don't forget to check out the tutorials on Sunday. We have three this year and they're all really cool. I highly recommend them for students. You'll also have an extra day in Heidelberg, and that is pretty cool as well!"

Klaus offers a final thought:

"After two virtual events, this year's BVM is the perfect opportunity for young PhD students to mingle with the German medical image processing community and expand their research network."



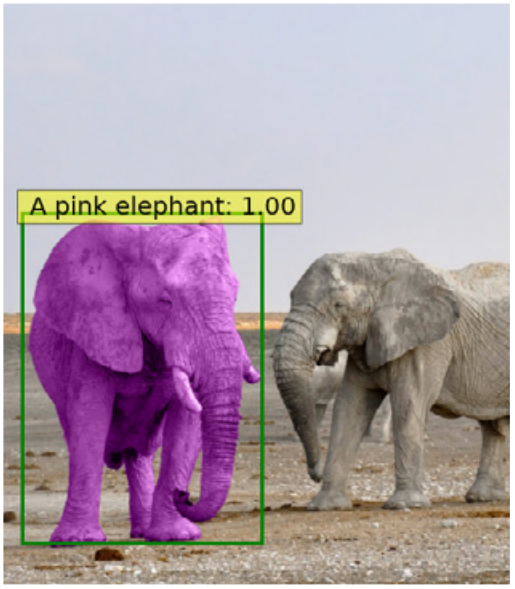
IMPROVE YOUR VISION WITH

Computer Vision News

SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!



RSIP VISION