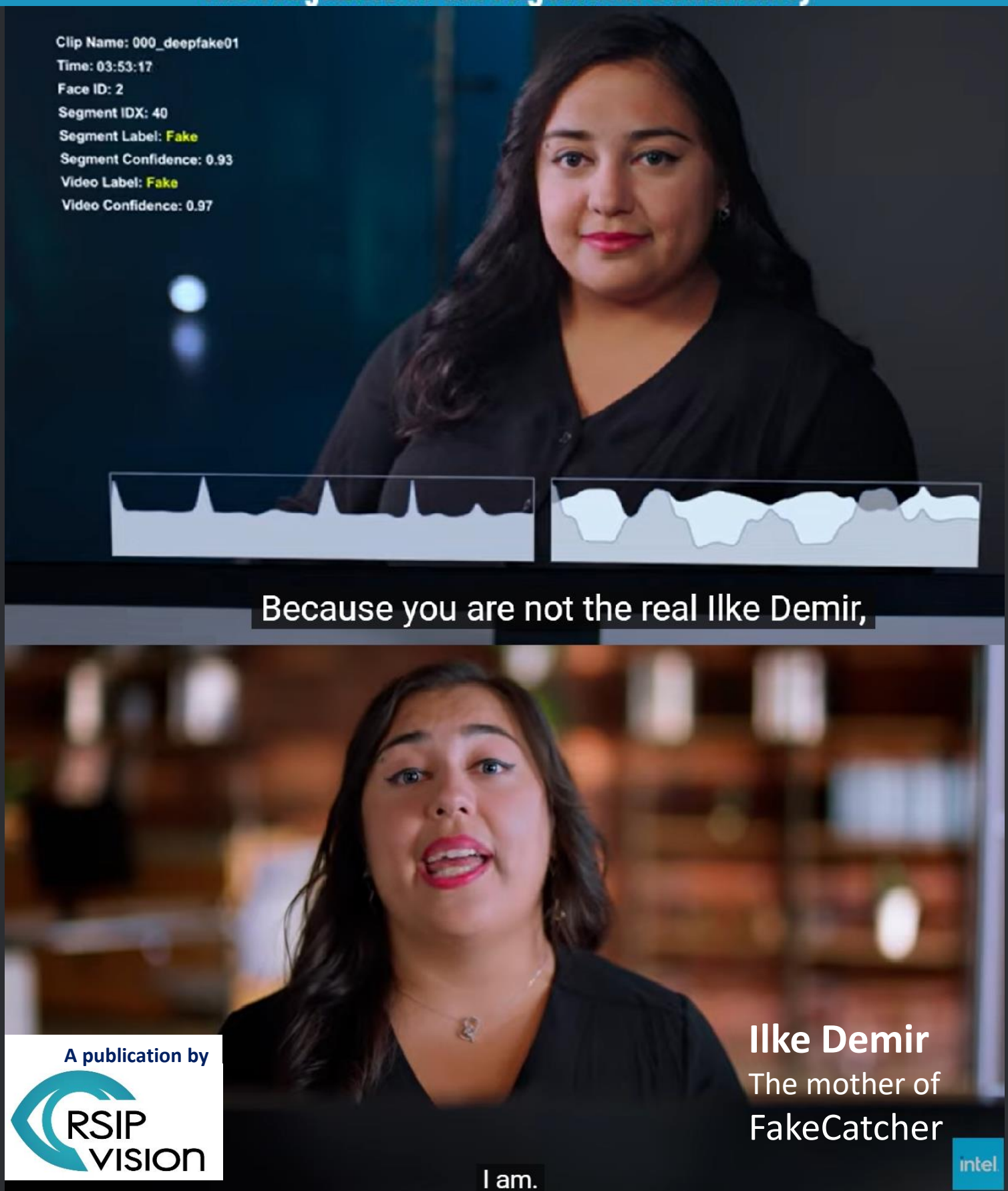
August 2023

Computer Vision News

& Medical Imaging News

The Magazine of the Algorithm Community



A publication by

RSIP VISION

Real-time Deepfake Detection Platform

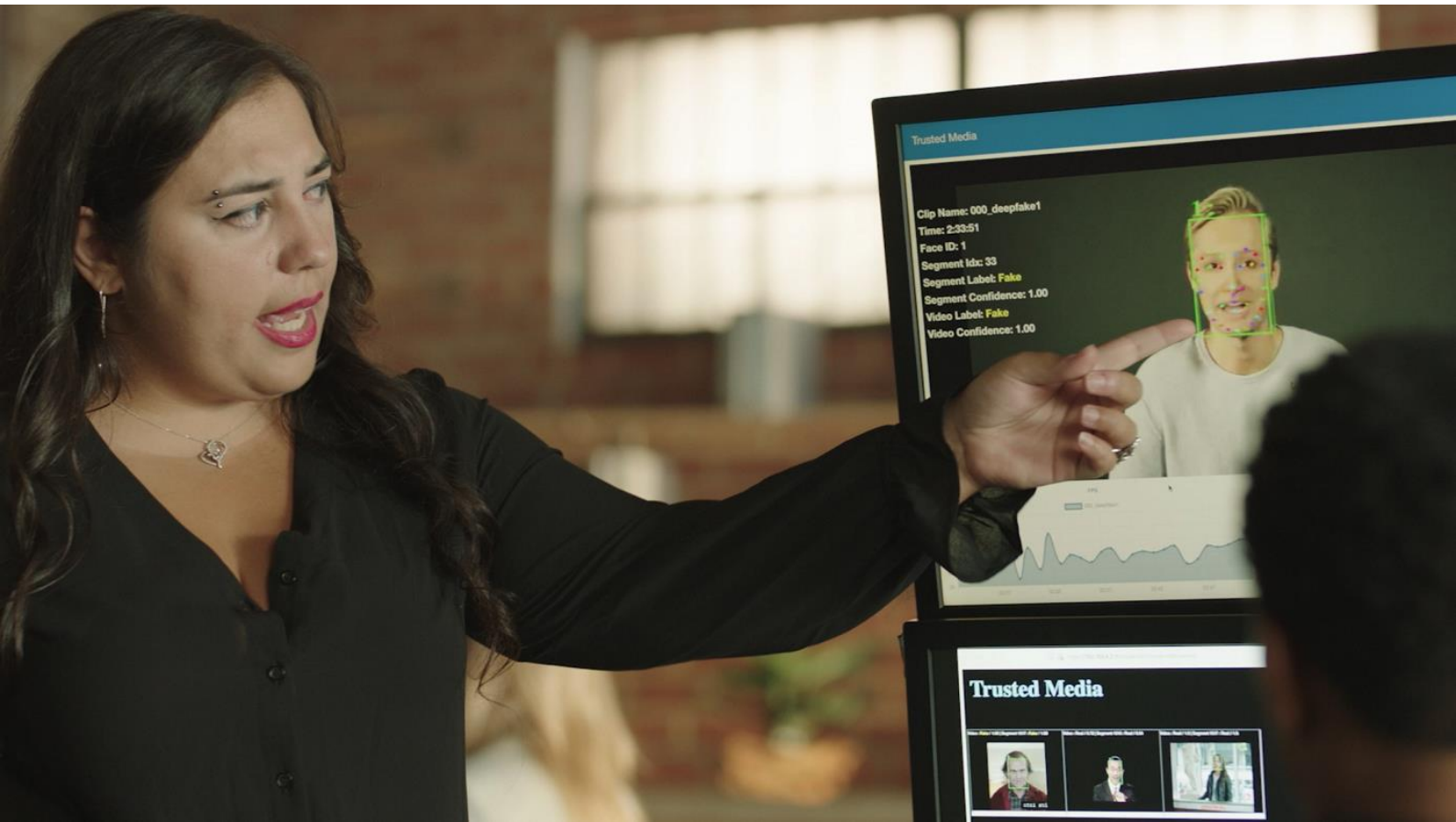
Ilke Demir is a Senior Staff Research Scientist at Intel Labs, leading the Trusted Media team, working on manipulated content detection, responsible generative AI, and media provenance.

One of the consequences of **AI democratization** and the growing emergence of generative models is the rise of **deepfakes**. Through its work to determine the authenticity of media content, **Trusted Media** aims not simply to identify artifacts of fakery, including broken hands and symmetry issues, but to answer a deeper question: **Is there an inherent watermark in being human?**

"*The first thing we look at is **blood**,*" Ilke tells us. "*When our heart pumps blood, it goes to our veins, and they change color. That color change is called photoplethysmography (PPG). We collect those PPG signals from the face, look at their temporal, spectral, and spatial correlations, and create PPG maps. On top of those, **we train a deep neural network to classify them into fake and real videos!***"

This technology is called **FakeCatcher**, one of several deepfake detectors developed by the team. Others examine whether eye gaze remains consistent over time and whether the motion in a video aligns with natural human movements. Then there is multimodal detection, such as exploring correlations between head movements and voice changes.

Deepfakes extend into scene and



object manipulation and even satellite imagery. To address this, the team has built a detector that breaks down large satellite images into smaller patches, enhancing their resolution using **super-resolution techniques** and utilizing a multi-head attention transformer network to extract textures and features to determine their authenticity.

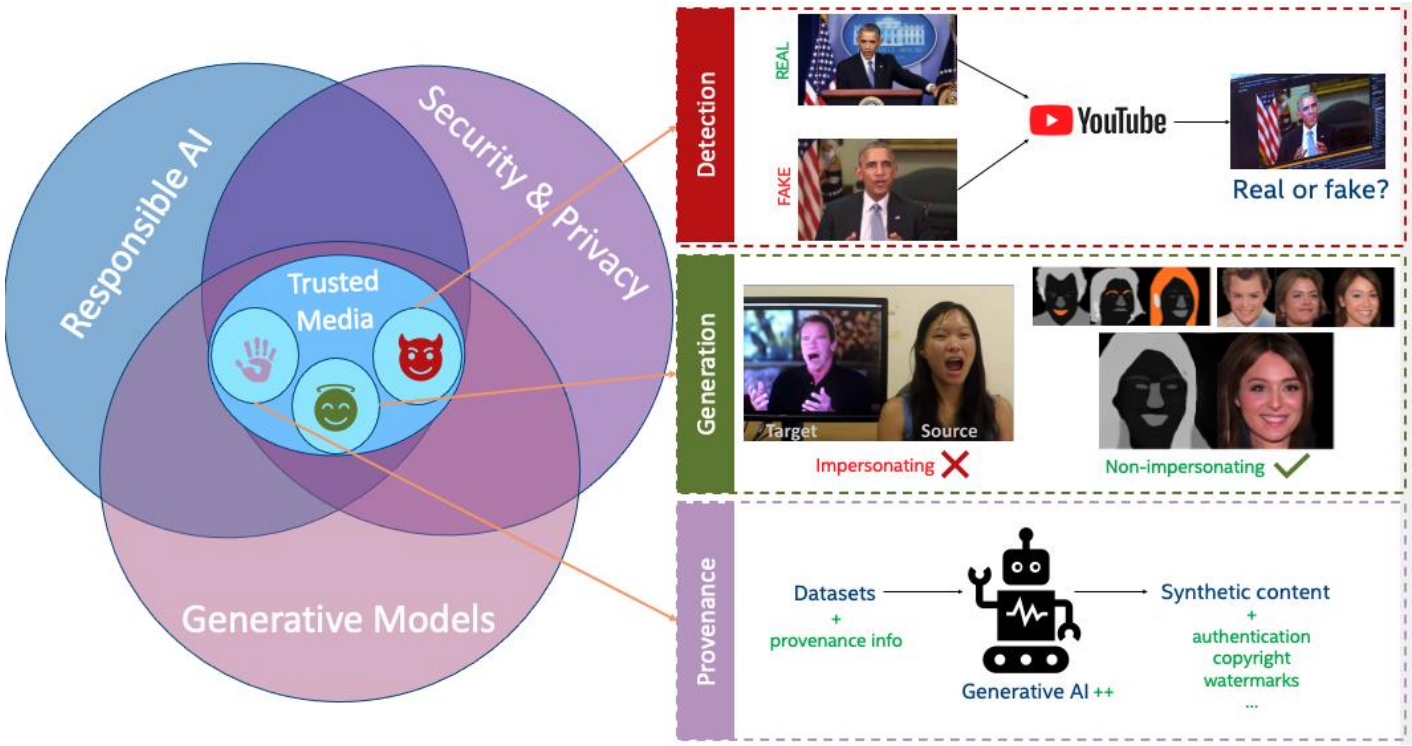
"*We have very high accuracy rates, especially for multimodal and multi-domain approaches,*" Ilke reveals. "*FakeCatcher has **96% accuracy**, our multimodal detector has over **98%**, and our satellite imagery detector has **99.62%**. It's getting better and better!*"

Ilke compares this work to an arms race – **as the generators improve, so do the detectors**, but then the generators improve again, and so on. It is an endless game of cat and mouse. However, she assures us the detectors are always one step ahead of any new generator, using authenticity signatures and priors in the data rather than conforming to fakery.

We have to ask: What could go wrong?

"***I would say people can go wrong!***" Ilke laughs. "*Instead of using our technology to enhance decision-making, if people try to use it as the absolute decision-maker for validating fakes, then that's bad. We're not saying any of the detectors are 100% perfect. That's why we provide so many different detectors. We don't want to decide for you. The biggest risk is always humans because **our systems are mostly deterministic, but humans are not**.*"

Trusted Media has also been building trust metrics around technical systems to enhance **interpersonal trust, societal trust,**



people's trust in systems, and systems' trust toward people. To this end, it conducts user studies, collaborates with social scientists, engages with customer groups, and utilizes its systems to investigate various trust-related factors. One of its main objectives is to minimize the risk in the current climate of marginalizing humans in this equation.

Regular readers may recognize Ilke from her [Women in Computer Vision feature in 2017]. Since then, her career has gone from strength to strength. As the self-declared **"mother of FakeCatcher",** her passion for the technology shines through as she talks about it, and she takes pride in the fact that it belongs to her rather than any corporation.

"*My background is in proceduralization, which is **computer graphics, computer vision, and machine learning to find interpretable representations from 3D data**,*" she explains. "*I've been looking at priors, distributions, and generative models my entire life – 20 years of research. When deepfakes were rising, I thought, 'Generative models also generate deepfakes, and generative models have those priors. Can we find some human prior in data to depend on in that generated content?' I was also working on human understanding in virtual reality. At that point, I was like, there are some human priors, and machine learning can predict humans, so we can build something...*"

Ilke then saw an MIT paper about PPG signals, looking at blood flow from videos, and realized its potential for analyzing deepfakes. Alongside her colleague **Umur Aybars Ciftci**, they began running experiments on the data and proving why PPG works. FakeCatcher soon

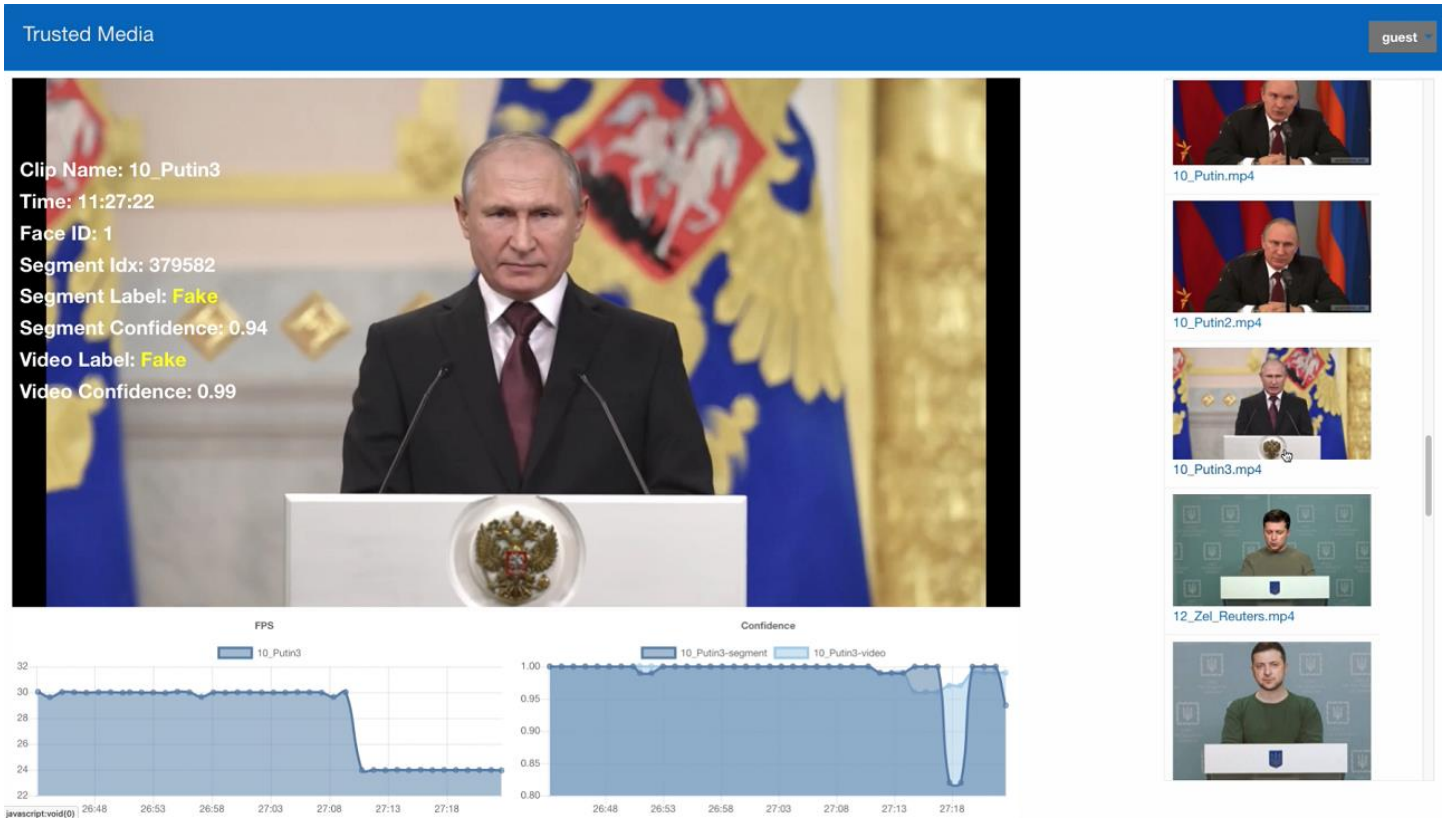
became accomplished at finding every deepfake. The team at **Intel** super-optimized it using **OpenVINO**, **VNNI**, **AVX**, and all the Intel AI accelerators to make it super real-time on **Intel Xeon processors**. Up to 72 concurrent FakeCatcher streams can be run on one machine.

Responsible generative AI is another research area for the Trusted Media team. Can deepfakes be responsible and used for good? Can inherently ethical and responsible generative AI be built with design priors for the architecture rather than trying to patch it afterward?

"*If we start with a responsible generative model, then everything else will follow through because we aren't leaving room for disinformation, misinformation, false data, or impersonation,*" Ilke points out. "*We're trying* ***to counter all the harmful aspects of generative AI by design, network architecture, loss choices, training data, and watermarking***."

Trusted Media's other research domain is **media provenance**, which involves establishing the authorship of media content. It addresses the issue of false ownership claims by providing crucial information about the origin, creator, creation process, purpose, and consent behind the media. The goal is to embed this information directly into the content, like a fingerprint. Even if a generative model is used to create the person in the video, it is acceptable if it is the consented and genuine version. Various information-theoretic methods, such as **authentication** and **watermarking**, can embed provenance

"Our Real-time Deepfake Detection Platform is the flagship product supporting our corporate responsibility towards the world!"



information into **synthetic data**.

We have to ask Ilke the million-dollar question: Why is a company that exists to make money investing so much time and energy into this endeavor? Is it doing humankind a favor, or is there another opportunity here?

"***That's a wonderful question,***" she responds. "*There are several value captures coming from Trusted Media. These algorithms are super-optimized to work on* ***Intel hardware****, including Intel Xeon and VPU, which supports more platform and hardware sales.* ***The OpenVINO team is doing wonders****. They were running* ***Stable Diffusion with one frame per second****. Wow! My team's detectors and responsible generative AI approaches follow that path. But all big corporations have some corporate social responsibility. At Intel, we call it* ***RISE initiatives*** *and our* ***Real-time Deepfake Detection Platform*** *is the flagship product supporting our corporate responsibility towards the world.* ***We support trust and human-centric and responsible AI through our products. These are key values for Intel!***"

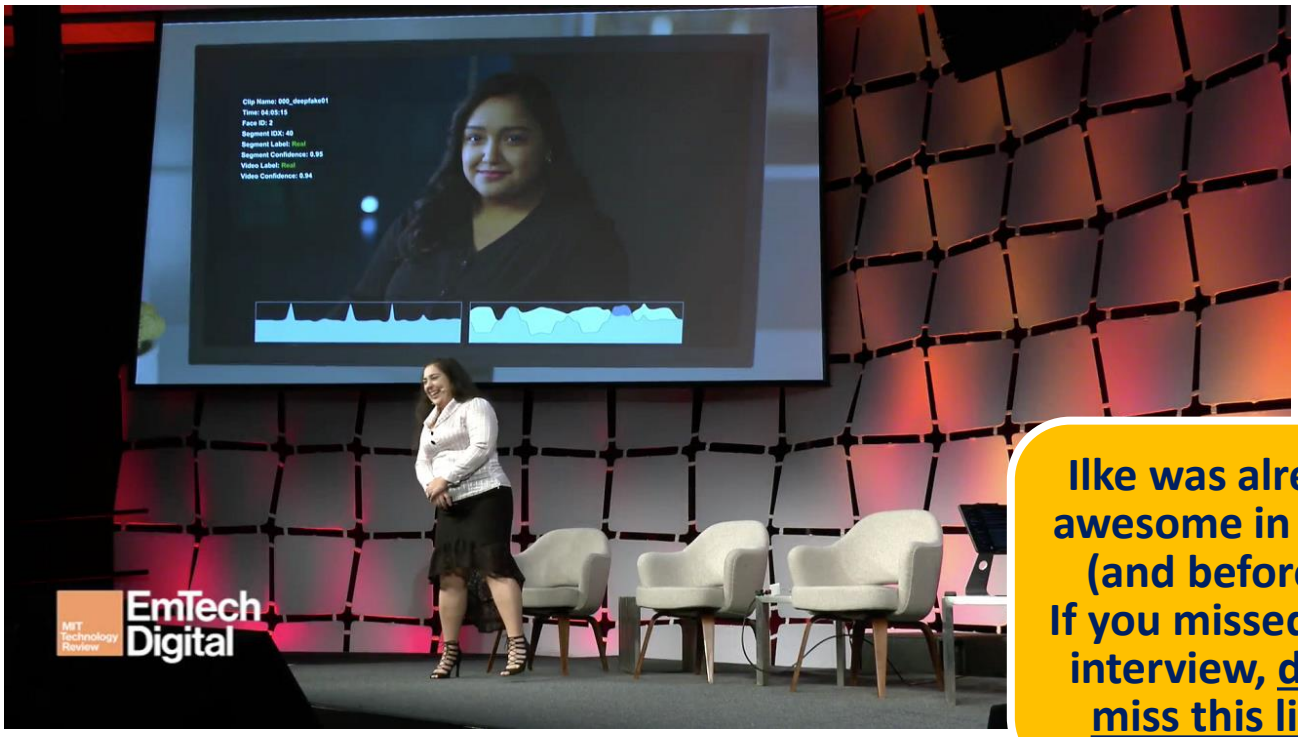
There is also a critical humanitarian angle to Trusted Media's work. It collaborates with human rights organizations, nonprofits, and civil organizations to combat misinformation and disinformation in cases where high-risk and high-impact deepfakes emerge. To stop the spread, it strives to **provide immediate results and accuracy to these organizations, ideally within the hour**. A crucial aspect is providing

information to people in emergency situations in war zones.

*"If you remember, there was the **Zelensky deepfake** giving misinformation about the **Russian invasion**,"* she recalls. *"We may think it's fake, but people on the ground might not know that. They see their president saying something and believe it. **We can help everyone decide if what they see is the truth!**"*

And that is AI for good.

"Absolutely, yes!"



Ilke was already awesome in 2017 (and before) – If you missed that interview, don't miss this link!

Computer Vision News

Editor:
Ralph Anzarouth

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

Publisher:
RSIP Vision

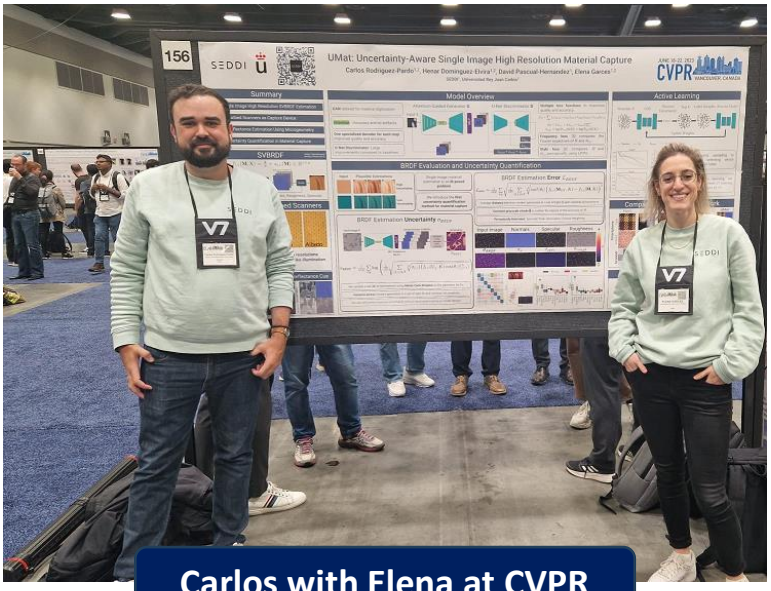
Contact us
Give us feedback
Free subscription

Read previous magazines

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

RSIP VISION Visual Intelligence for MedTech

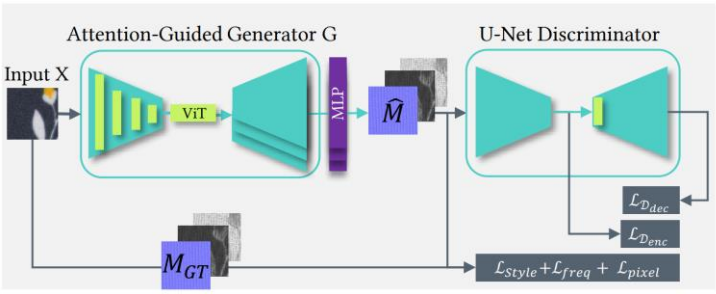
Neural Networks for Digital Materials and Radiance Encoding

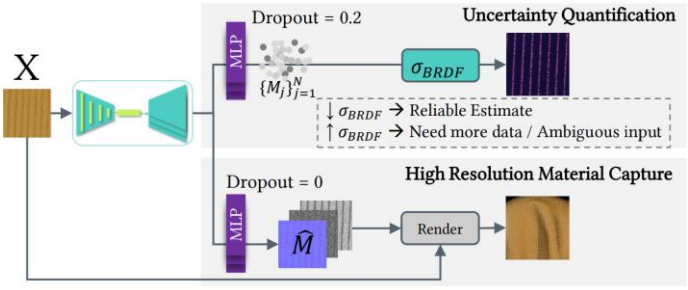


Carlos with Elena at CVPR

Carlos Rodríguez - Pardo did an industrial PhD at Universidad Rey Juan Carlos, which was fully funded by SEDDI, a startup based in Madrid focused on digitizing the fashion industry. His PhD was supervised by awesome Elena Garcés.

The focus of his thesis is to develop deep learning based methods for digitizing materials, inverse graphics, and encoding radiance for virtual scenes. Congrats, Doctor Carlos!





Realistic virtual scenes are becoming increasingly prevalent in our society, with a wide range of applications in areas such as manufacturing, architecture, fashion design, and entertainment, including movies, video games, and augmented and virtual reality. Generating realistic images of such scenes requires highly accurate illumination, geometry, and material models, which can be time-consuming and challenging to obtain. Traditionally, such models have often been created manually by skilled artists, but this process can be prohibitively time-consuming and costly. Alternatively, real-world examples can be captured, but this approach presents additional challenges in terms of accuracy and scalability. Moreover, while realism and accuracy are crucial in such processes, rendering efficiency is also a key requirement, so that lifelike images can be generated with the speed required in many real-world applications.

One of the most significant challenges in this regard is the **acquisition and representation of materials**, which are a critical component of our visual world and, by extension, of virtual representations of it. However, existing approaches for material acquisition and representation are limited in terms of efficiency and accuracy, which limits their real-world impact. To address

these challenges, **data-driven approaches that leverage machine learning** may provide viable solutions. Nevertheless, designing and training machine learning models that meet all these competing requirements remains a challenging task, requiring careful consideration of trade-offs between quality and efficiency.

In my thesis, we propose **novel learning-based solutions to address several key challenges in physically-based rendering and material digitization**. Our approaches leverage various forms of **neural networks** to introduce innovative algorithms for radiance encoding, digital material generation, edition, and estimation.

First, we present a **visual attribute transfer framework for digital materials** that can effectively generalize to new illumination conditions and geometric distortions. We showcase a use-case of this method for high-resolution material acquisition using a custom device. Additionally, we propose a generative model capable of synthesizing tileable textures from a single input image, which helps improve the quality of material rendering. Building upon recent work in neural fields, we also introduce a material representation that accurately encodes material reflectance while offering powerful editing and propagation capabilities. In addition to reflectance, we present a novel method for global illumination encoding that leverages carefully designed generative models to achieve significantly faster sampling than previous work.

Finally, we propose **two innovative methods for low-cost material digitization**. With flatbed scanners as our capture device, we present a **generative model** that can provide high-resolution material reflectance estimations using a single image as input, while introducing an **uncertainty quantification algorithm** that increases its reliability and efficiency. Additionally, we present a novel method for **digitizing fabric mechanical properties using depth images** as input, which we extend with a perceptually-validated drape similarity metric.

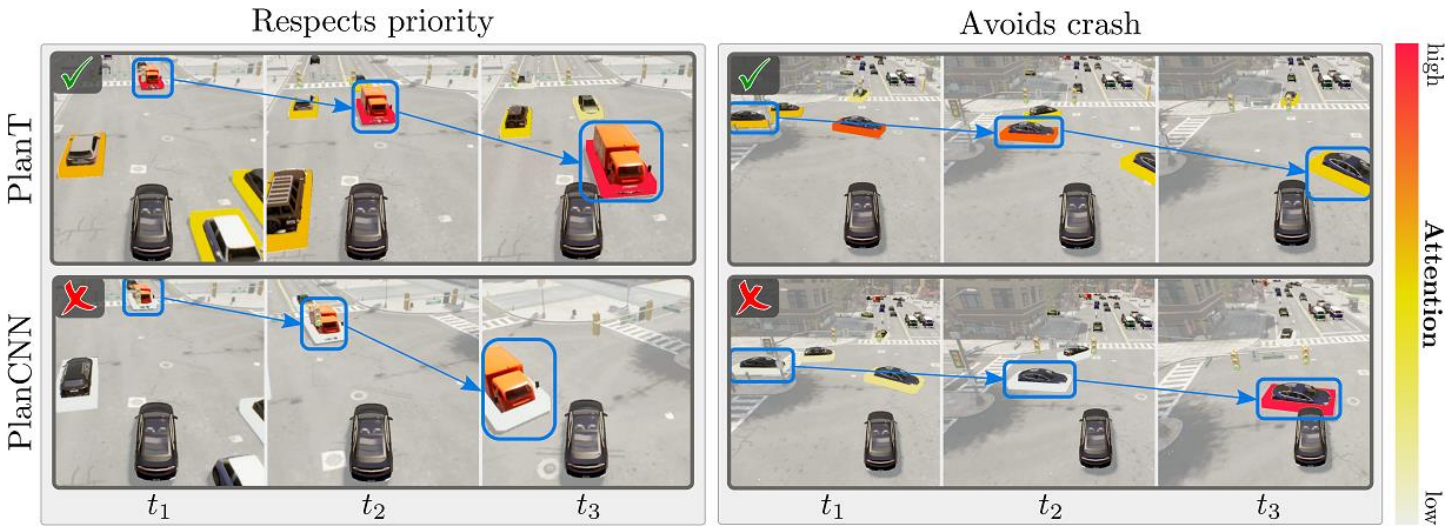
Overall, the contributions of this thesis represent significant advances in the fields of radiance encoding and digital material acquisition and edition, enhancing the quality, scalability, and efficiency of physically-based rendering pipelines.

PlanT: Explainable Planning Transformers via Object-Level Representations

Katrin Renz is a PhD student at the University of Tübingen in Andreas Geiger's lab, working on the combination of autonomous driving and language.

Last month, she won the Best Presentation prize at Sicily's International Computer Vision Summer School (ICVSS). She speaks to us about her award-winning work.

PlanT, Katrin's paper from last year's Conference on Robot Learning (CoRL), proposed **a state-of-the-art learning-based planner for autonomous driving**. Autonomous driving can involve a modular pipeline comprising sensor data, a perception module for perceiving the environment around you, and a planner. The planner considers the perception output and the 3D detection of surrounding vehicles and determines **the optimal trajectory for the ego vehicle**.



Check out our Video Interview with Katrin

max planck institut informatik
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS
EBERHARD KARLS UNIVERSITÄT TÜBINGEN
Renz K. Koepke A.S. Chitta K. Akata Z. Mercea O.B. Geiger A.

PLANT:

EXPLAINABLE **PLANNING**

TRANSFORMERS

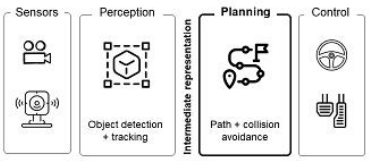
VIA OBJECT-LEVEL

REPRESENTATIONS

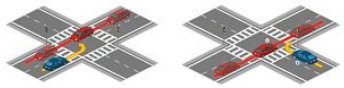
Abstract

Planning an optimal route in a complex environment requires efficient reasoning about the surrounding scene. In this paper, we propose PlanT, a novel approach for that uses a standard transformer architecture. PlanT is based on imitation learning with a compact object-level input representation. Combining PlanT with an off-the-shelf perception module provides a sensor-based driving system that is more than 10 points better in terms of driving score than the existing state of the art.

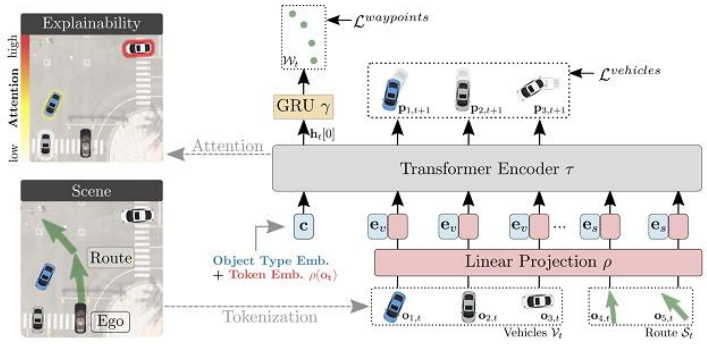
Task



CARLA scenarios

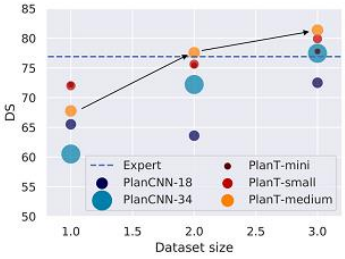


Architecture



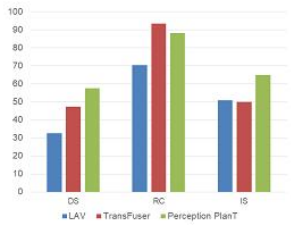
- Train **a standard transformer encoder** from scratch
- Loss on **future positions** of ego vehicle and the other vehicles

Results



- **Scaling** dataset and model improves performance
- **Expert level performance**

Perception PlanT



- Adding a perception module
- State of the art on longest 6 benchmark

Explainability

- Visualization of **attention weights** to show the **most important object**
- **Temporarily** more **consistent** than the CNN-based method + also takes **geometrically distant** objects into account

Traditionally, planning modules were rule-based, requiring **handcrafted rules for every possible scenario**. While these methods served their purpose, they lacked flexibility and scalability, making them challenging to adapt to dynamic and complex real-world environments.

Previous works perform this learning-based planning using a rendered image and a CNN or a graph or object-level representation evaluated offline. **Katrin's innovative model performs online evaluations of the ego vehicle's trajectory in real-time simulations**, a vital improvement on offline metrics.

Other key aspects that set PlanT apart are its **simplicity and extendibility**. Many existing models use complicated architectures, making implementation and understanding difficult. In contrast, PlanT offers a straightforward yet effective transformer-based approach that can serve as a baseline for various extensions. **It tokenizes the scene**, making it easily adaptable for other use cases, such as inputting language tokens for combining language and driving.

A transformer network inherently has attention weights. Katrin used this feature to enhance the **explainability of the model**. Attention weights were visualized to determine how much attention the network assigned to each vehicle in the scene when making decisions. Additionally, she proposed an evaluation scheme to assess the reliability of these attention weights to explain the model's decisions, as she tells us there is some contention in the field around whether attention is explainability.

What does Katrin think convinced the jury to award her presentation the top prize?

"*There was a first, second, and third place, and from what I saw of the others, I think we had the most non-standard poster,*" she reveals. "*The posters were not what you see all the time at poster sessions. All three had some kind of fancy layout. A third of my poster was just a red bar with a title, my logo from the project, and the QR code. I had a smaller piece for the content with the really important stuff. It was an eye-catcher. Presentation-wise, many people did a great job conveying what they did and explaining it in a clear way.*"

Check out our video interview with Katrin to learn more about this work, including the efforts that have already begun to extend it, her thoughts on modular vs end-to-end approaches, and her unconventional path to working in this domain.


GHANA
Khadija Iddrisu
Watch her in video!
AIMS
GHANA
CLASS
OF
2021

Khadija Iddrisu is a PhD researcher at Dublin City University and works with the Insight SFI Research Centre for Data Analytics in Ireland.

Read 100 FASCINATING interviews with Women in Computer Vision

Khadija, you're not Irish, are you?

No, I'm Ghanaian. I'm from Ghana.

How is it to be a scientist from Ghana?

This is actually a very intriguing question. In Ghana, there are not a lot of scientists, I would say, in the field of machine learning and AI. It's just a few people. But for now, I'm part of the Women in Machine Learning, and our agenda is to try to get so many people to pursue research in STEM and AI and related areas.

When did you understand that you were going to be a scientist?

It started in my undergraduate studies. It was when Covid started, and by then, I was a computer science student, but I didn't know exactly what I wanted to do after school, so you would find me doing graphic designing, you would find me doing web designing, and it was just like a whole lot of mess! [*laughs*]

When Covid came, and there was a long break, we were not sure when we were going back to school. One time I saw a flyer that said they are having six weeks training in AI. On the flyer, there was a robot, and I like to watch sci-fi movies, so the robot caught my attention. I said that I wanted to know how people build robots, so I would go for this training. Afterwards, I realized it wasn't just about robots, it was about AI.

At the end of it, I did a project, and this project was about trying to use computer vision to detect disease in poultry. Even though this project was not a success because by then we didn't have a lot of datasets and AI was new in Ghana as well, I was just so fascinated by the fact that we can use computer vision to solve problems in almost every field. I just became interested, and I knew that I wanted to pursue further studies in this area, and that is what got me here.

How prepared were you for the challenge that awaited you?

I've met a lot of people on my career path, and they have acted as sort of like career guides, and they are people that I can always go to for advice. Alex [*Alessandro Crimi*], for instance, he's a very great person, and he has advised me to work on several projects. With these people, I felt like it would be much easier.

Regarding your school path, was it good enough for you to get in contact with complicated technological challenges for a PhD?

My undergraduate was very challenging because, by that time, we were not doing a lot of coding, and even if we code, sometimes, during exams, we have to write programs on paper. It was really

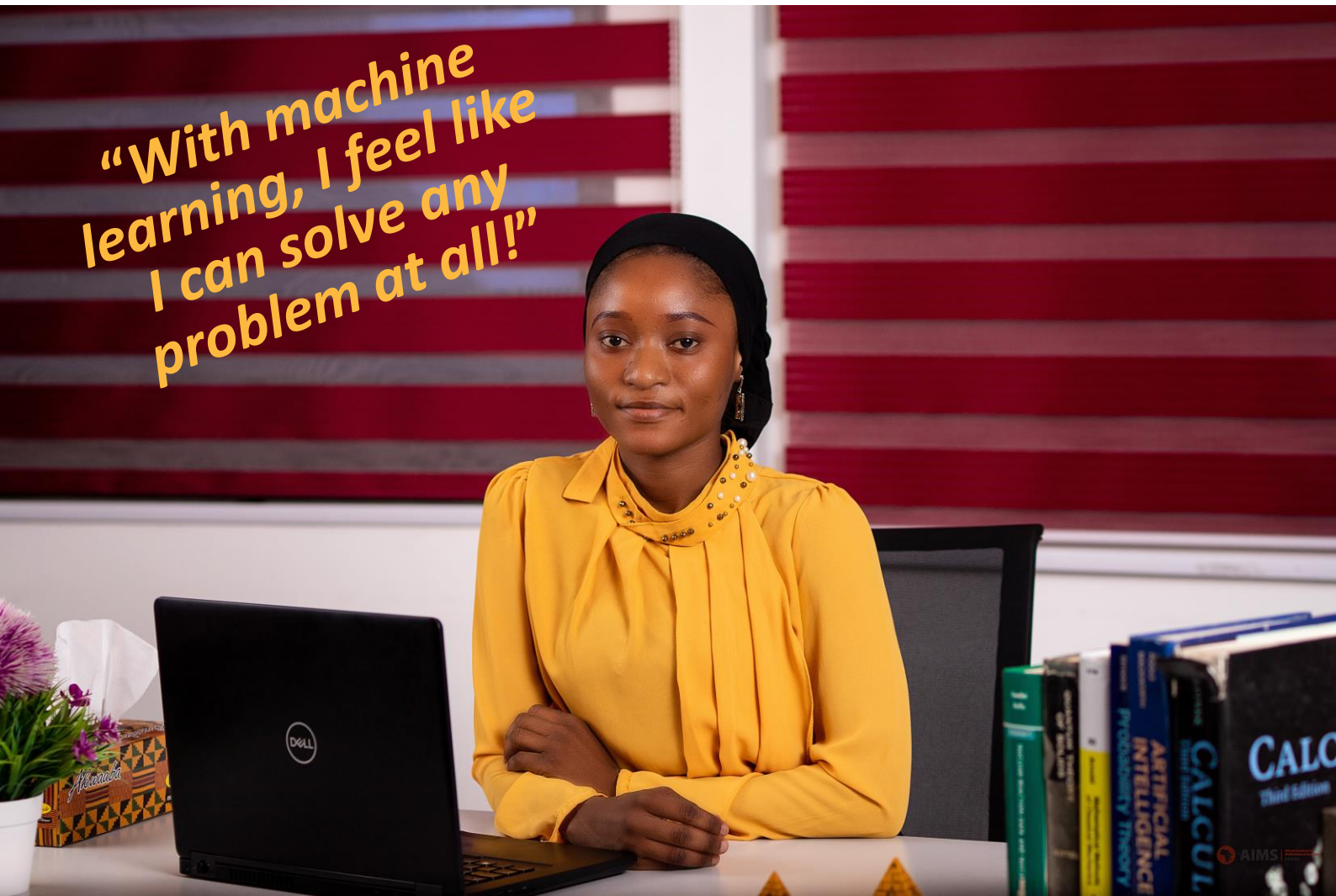
"I just have to keep doing this for the sake of my soul, for the sake of my country, and for the sake of the whole world!"

difficult, but afterwards, I had my master's at the African Institute for Mathematical Sciences, and over there, it was very intensive.

We have lecturers from different parts of the world. They would come in and teach us courses for three weeks. In my typical universities in Ghana, we would take three months to study these courses that they are teaching for just three weeks. It was quite an amazing experience. That was actually where I met Alex. He was my advisor for the first teasers that I ever did in machine learning.

Going to the African Institute for Mathematical Sciences really gave me a lot of confidence. I gained a lot of skills that I needed to use. All those technical skills that I needed were provided to me there. Afterwards, I also went to the African Masters of Machine Intelligence, which is actually the same as my Master's in Maths program, but this time, it was sponsored by Google and Facebook.

We had lecturers who were industry



workers in Google, we had people who worked at Meta, and they would all come down to Senegal and take us through courses. We would have mock interviews with them as if we are interviewing for roles in their company. We'd also sometimes have sessions with them on how to apply for PhD programs and how to choose a career path, and I think it really gave me a lot of understanding. That really was a very huge stepping stone to where I am right now.

Through all these challenging paths that you just described, how did you keep yourself motivated to overcome these challenges?

Yes, that is a very interesting question! [*laughs*] Mostly, when you're from Ghana, and from my region specifically – I am from the northern part of Ghana – people believe that girls like myself should only just go to high school and they are done, and they should get married and stuff like that. When I first got to the university, and I realized that I wanted to do something like this, I met about two women that were from the same community as myself, and they went to the same schools as myself, and they were actually doing very well. When I saw that they could actually leave that community and become great people that would be role models to other people, it sort of motivated me and gave me the understanding that I can leave as well, and I can get to motivate many young people as well. That was one thing that was just the driving force for me to go into research.

Also, when I realized that with machine learning we could do so many things and Ghana was lagging behind, it was the main driver for me. I realized that if I get a PhD and I have experience, I can always go back to my country and set up a research lab where people would learn about machine learning and AI, and that way, they can also use it to solve problems in their community as well.

What are you currently working on?

I applied to a program in machine learning in Dublin, but it was very competitive, and I was put on a waiting list. A few months afterwards, a professor reached out to me that he was impressed by my interview, and he wanted me to work with him on a project. This project is with a company called Xperi. They have offices in Europe and US. As part of the arrangements, I would do my PhD at Dublin City University, and automatically, I am part of the Insight SFI Centre for Data Analytics, but also, as part of my PhD, I would move to their company to work for two years.

Their company deals in a lot of computer vision applications for cars, for smart homes, and the current project I am working on is about trying to detect microsleep. Microsleep happens when people are driving, or you are just seated, and then you doze off for five minutes, and then you can't even believe that you slept. This has led to a lot of accidents. What I am trying to do is to use EEG signals and also just images and videos of people that have been simulated driving and sleeping, and then we try to predict when the microsleep will happen and how long it would occur. That's where we can implement a system to trigger when someone is going into microsleep. That's the topic I'm working on for my PhD.

We are trying to use a different type of data called data from event cameras. Event cameras are new in the vision system, and the type of data it provides is different from the data we have from traditional cameras. Event cameras give us more data. We can see when a change in brightness occurred, we can see the direction of the brightness change and everything. This is why we are trying to use event cameras, traditional cameras, and EEG signals to predict when a microsleep would occur.

I'm still at the early stages of my PhD, so I have been asked to just do some experiments to get used to how event camera data works. For instance, right now, I'm working on trying to replicate a paper that tries to do eye blink detection. We try to estimate attention level from eye blink detection, and the data is from an event camera.

Can you tell us more about those moments when you wanted to give up? How did you handle them?

I was telling one of my friends when we were in my master's project, he would ask me, what is my next step after here? I would tell him I'm not even thinking about my next steps. I just want to solve this maths question, and I'm happy

enough. I've encountered times when it's very difficult to keep working on research, especially when you are not getting any new results. I just felt like if I had gone to the industry, if I had just gone to get a regular job, my life would have been easier.

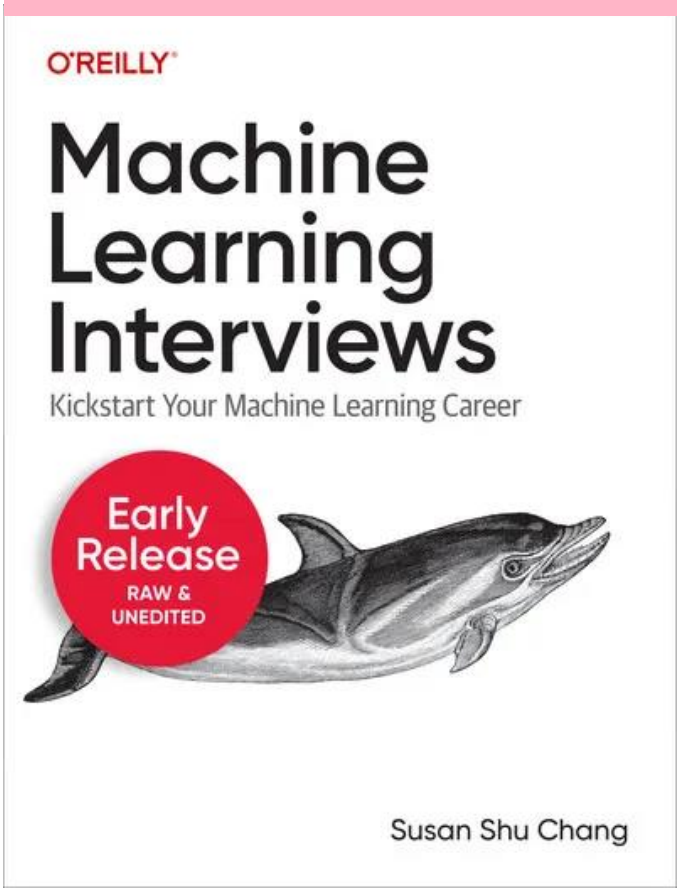
But I always refer back to the first research I did with Alex. When I did this research, I had the opportunity to present it at a conference in Tunisia. After I presented this work, a lot of people came up to me asking me questions about how I was able to do this sort of work, how they can apply it to their own research and everything. Then afterwards, I also had the opportunity to present it at a Women in AI conference in Ghana, and people asked me lots of questions about how I was able to do that. How can this be applied to our daily lives?

Anytime I'm kind of stuck in a loop where I feel like I don't want to continue with research anymore, I just refer back to that, and then I realize that this research that I thought I was doing just for fun has actually impacted the lives of so many people. Because of that, I don't want to stop. I want to keep researching, and I want to keep making the lives of people better. It's not up to me to decide when I want to stop. I just have to keep doing this for the sake of my soul, for the sake of my country, and for the sake of the whole world, and that's just what motivates me to keep on going!

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

New Chapters of Machine Learning Interviews Book!

Our readers already know about the exceptional book by awesome Chip Huyen: **Designing Machine Learning Systems**. Apparently, O'Reilly is publishing another excellent ML book, this one by **Susan Shu Chang**: **Machine Learning Interviews**. We can already read some of the chapters at the following link: Read More



Training CV to Think Like the Brain Enhances Performance and Robustness

This is an article by Cryptopolitan, but don't worry, it doesn't talk about crypto currencies. It is about an intriguing work by MIT and IBM aiming at aligning computer vision with human vision. They think this has the potential to advance the field and deepen our understanding of biological neural networks.

Science knows which part of the brain is the one humans and monkeys rely on for object recognition. The MIT researchers led by James DiCarlo have made a computer vision model more robust by training it to work like that part. Read More

How photonics is revolutionizing convolutional neural networks

This nice piece by Eurekalert relates how researchers have turned to photonics as a means to enhance convolutional neural networks in a way that consumes less power and requires less memory, which is not mean feat for current voracious CNNs. The study was originally published on Intelligent Computing journal and all the authors of this EU H2020 work come from Greek universities. Read More

How Do We Know How Smart AI Systems Are?

A fantastic article by **Melanie Mitchell** of the Santa Fe Institute, published on **Science** about the **limitations of LLMs**: it's not (only) me saying that, it's Yann LeCun, who like Melanie believes that "*Taken together, these problems make it hard to conclude - from the evidence given - that AI systems are now or soon will match or exceed human intelligence*". A must read for all AI passionates! **Read More**

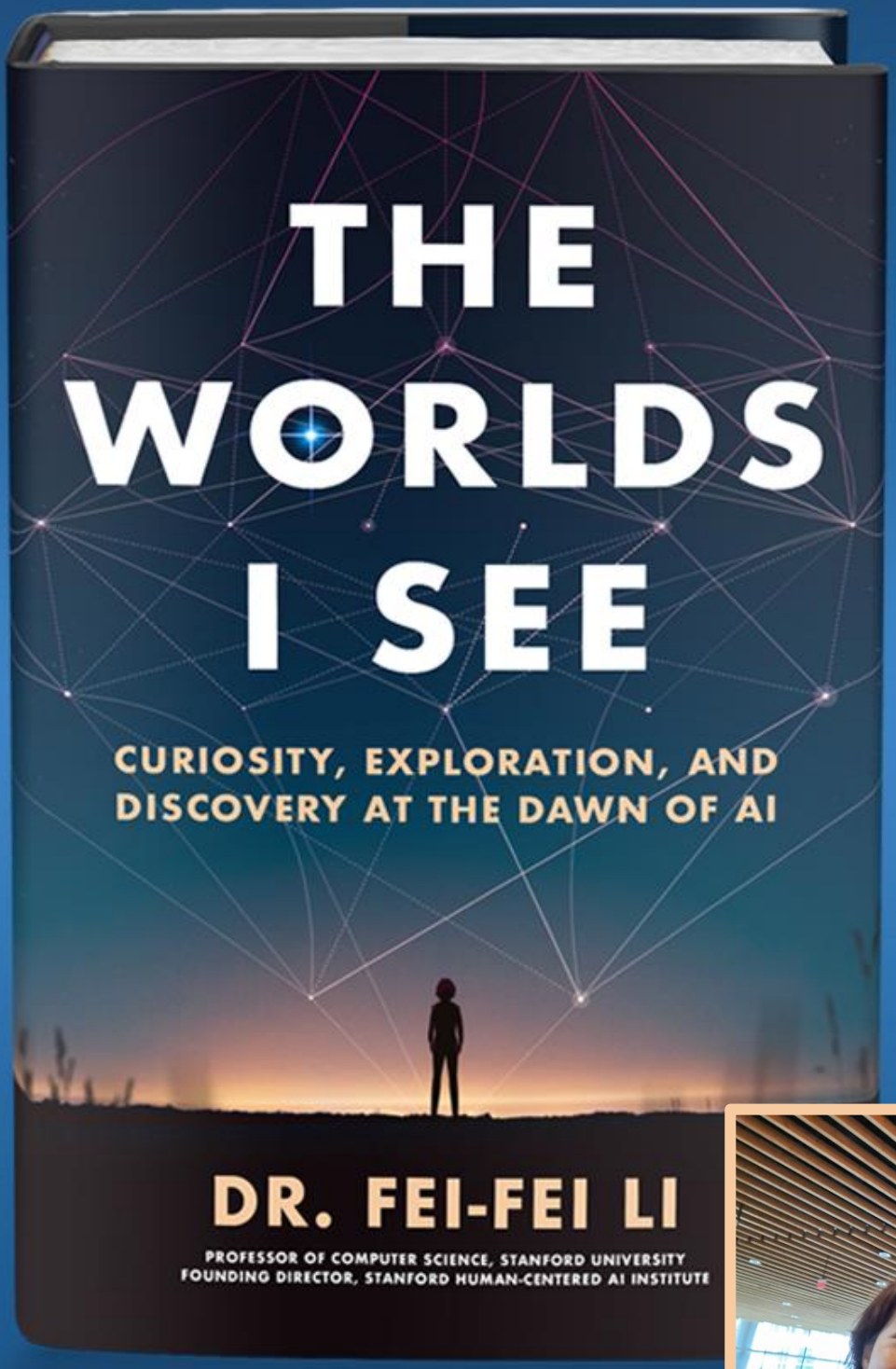
Glaze, a tool to protect human artists from style mimicry by generative AI models

Diffusion models such as MidJourney and Stable Diffusion have been trained on large datasets of scraped images from the web, many of which are copyrighted. They can be then used to copy individual artists, through mimicry. This software called Glaze claims to be able to protect human artists by disrupting style mimicry. The key lays in how AI sees visual information differently from humans. **Watch the video**

Australia Post uses computer vision for site safety

Apparently, Australia Post is using computer vision technology as a site safety initiative, to detect team members moving into "unsafe zones" at facilities. The company claims that it made significant investments to protection a workforce of almost 30,000 workers, using machine learning and computer vision technology, supported by Google Cloud Platform (CGP).

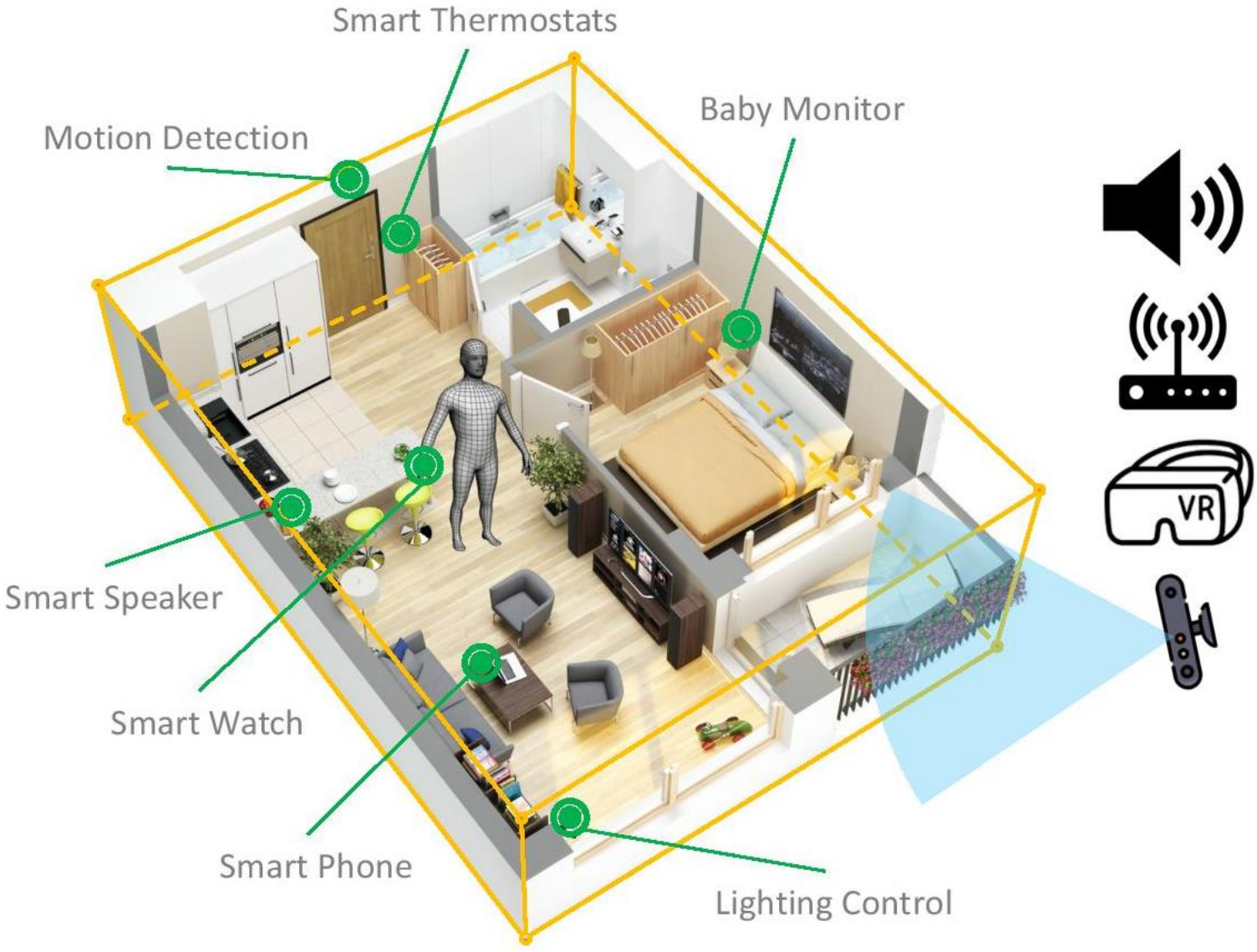
This is obviously very much needed for interacting with machines such as forklifts, trucks and sorting machines. What a shame that the post did not tell the details of their tech. **Read More**



Fei-Fei Li just announced that her book, The Worlds I See, will be published on Nov 7, at Moment of Lift Books (an imprint from Melinda Gates and Flatiron). She says that *"AI can help people and I hope you'll come along on the journey!"* Fei-Fei was General Chair at CVPR 2023 just one month ago and we even had the chance for a cute selfie and a chat about… Computer Vision News!

MEDICAL IMAGING NEWS

AUGUST 2023



Learn about Ambient Intelligence for HealthCare on page 38!

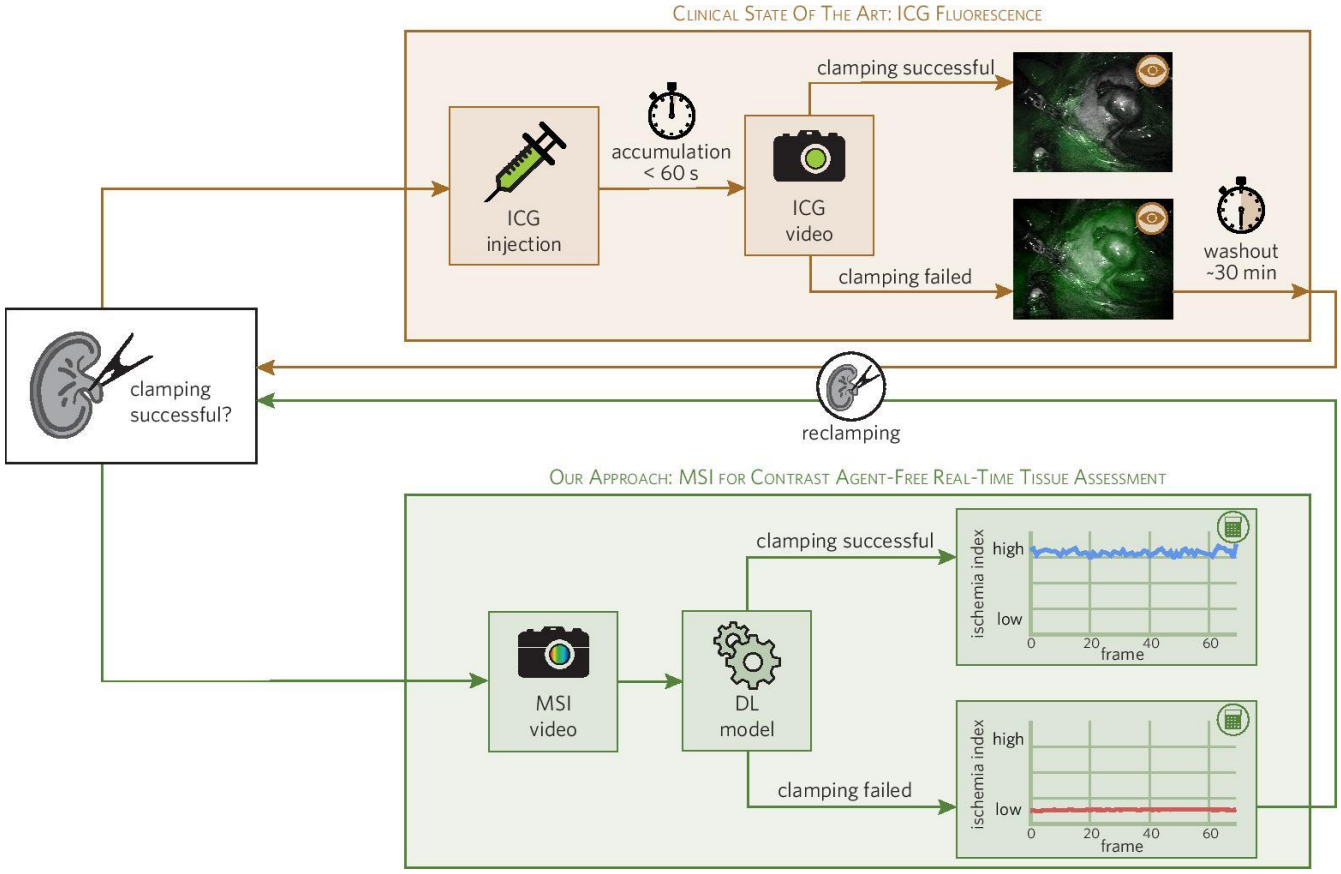
It's a MICCAI Workshop

Translational Functional Imaging Enabled by Deep Learning

Leonardo Ayala just gave "one of the best PhD defense presentations" that his supervisor Lena Maier-Hein has ever witnessed. Leo brought much joy during his PhD to the division of Intelligent Medical Systems (IMSY) at the German Cancer Research Center (DKFZ) through the "FUN ministerium" that he established. Congrats, Doctor Leo!

When I interviewed for my PhD position, Lena asked me why I wanted to change my research fields **from material science to computer science applied to medicine**, "*Because I would like to do something that has a more direct effect on helping people*" I answered. Even though at that moment I was full of doubts, I can now certainly say that I chose the right path, or perhaps it chose me.

Spectral imaging (SI) is an imaging technique that, in contrast to traditional RGB (red, green, and blue) imaging, provides much richer



spectral information by collecting light in many narrow regions of the optical spectrum. This property allows SI to encode functional properties (e.g., oxygenation and ischemia monitoring) in diffuse reflectance data. However, its translation to clinical practice is currently hindered by a number of limitations such as **image recording speed**, **controlled illumination restrictions**, and **high inter-patient data heterogeneity**.

During my PhD I worked on the development of systems and methods to translate spectral imaging into clinical practice. More precisely, my team and I addressed three main challenges: 1) **slow imaging devices**, 2) **controlled illumination restrictions**, and 3) **high inter-patient variability** and eliminating the need for **contrast agents**. Among these challenges, the later one was at the core of my dissertation.

Emerging imaging modalities such as SI innately face the challenge of limited data availability. Under conditions of data scarcity, high inter-patient data variability substantially impedes the development of clinically usable AI models. The challenge arises from the bias introduced in AI models when the distribution of the deployment population differs significantly from the population on which the models were trained.

To mitigate this bias, an **out-of-distribution (OoD) detection approach** was developed to monitor ischemia during surgery in a personalized manner, which only requires data from one single patient for **model training** without the need for contrast agents (Fig. 1). More details about this approach can be found in our **Science Advances publication**.

In summary, my work pioneered an entirely novel functional imaging paradigm based on spectral techniques and specifically removed common roadblocks to clinical translation. In doing so, it opens up new avenues of clinical functional imaging to the benefit of patients in surgery and beyond.

Multiple-instance Learning Inspired Explainable Deep Learning Network

by Christina Bornberg
@datascEYEnce

Hello, I am Christina! Welcome to the new RSIP Vision column datascEYEnce!

I am interested in deep learning applied to ophthalmology! I just finished my master's in medical image analysis and am now working at the Singapore Eye Research Institute.

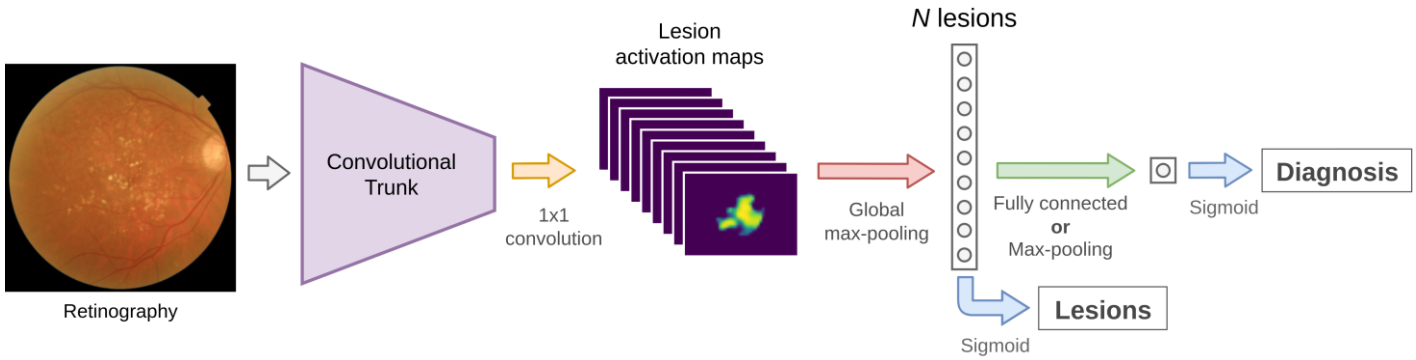
featuring José Morano Sánchez

José is currently a doctoral research scientist in the Christian Doppler Laboratory for Artificial Intelligence in Retina at the Medical University of Vienna. He received his bachelor's and master's degrees from the University of A Coruña in Spain, where he also pursued the research on weakly supervised learning which we are focusing on here today.

We're kicking off the **Deep Learning for Ophthalmology** interview series with **multiple-instance learning**. A huge thank you goes to José Morano Sánchez, who introduced me to his recently published work **Weakly-Supervised Detection of Amd-related Lesions in Color Fundus Images Using Explainable Deep Learning**.

The goal of the research was to create a pipeline that is able to diagnose **age-related macular degeneration (AMD)** which is a common cause of irreversible blindness. Instead of using a simple classification black box model for image-level labels of the nine lesion classes in their AMDLesions dataset, weakly supervised learning is applied in order to **generate one activation mask for each lesion type to enhance the explainability**.

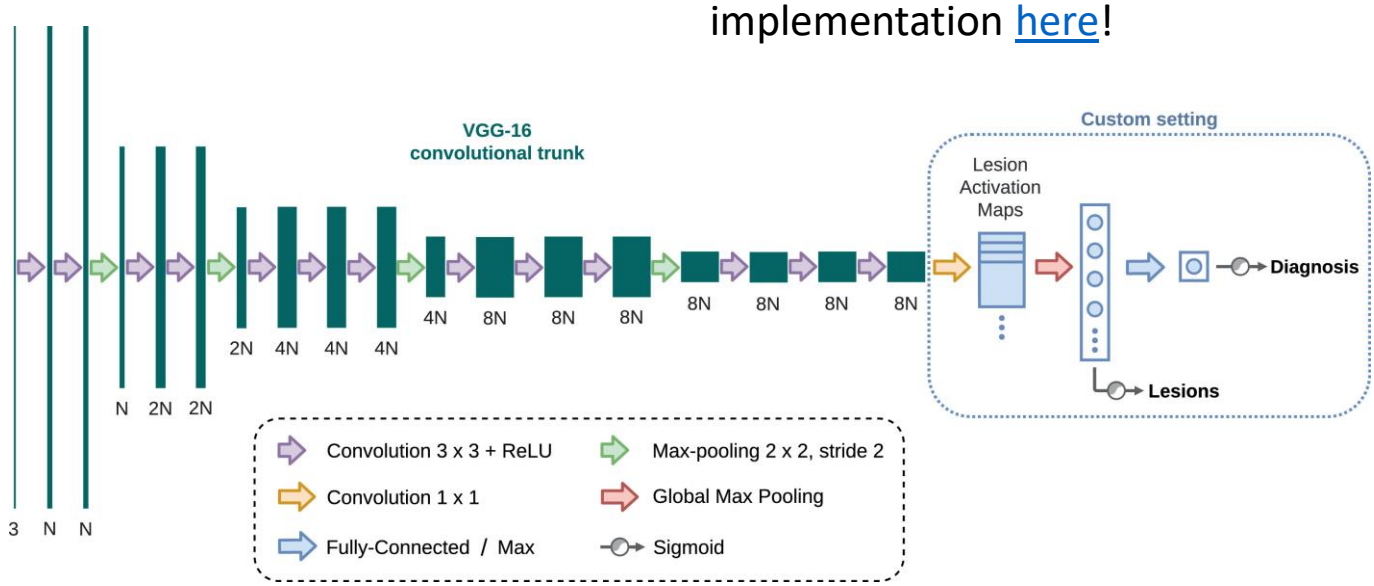
Now, how does this connect to multiple-instance learning? As you probably already know, in multiple-instance learning, we have a bag (image) of instances (pixels or groups of pixels). Multiple labels are provided for the entire bag (image-



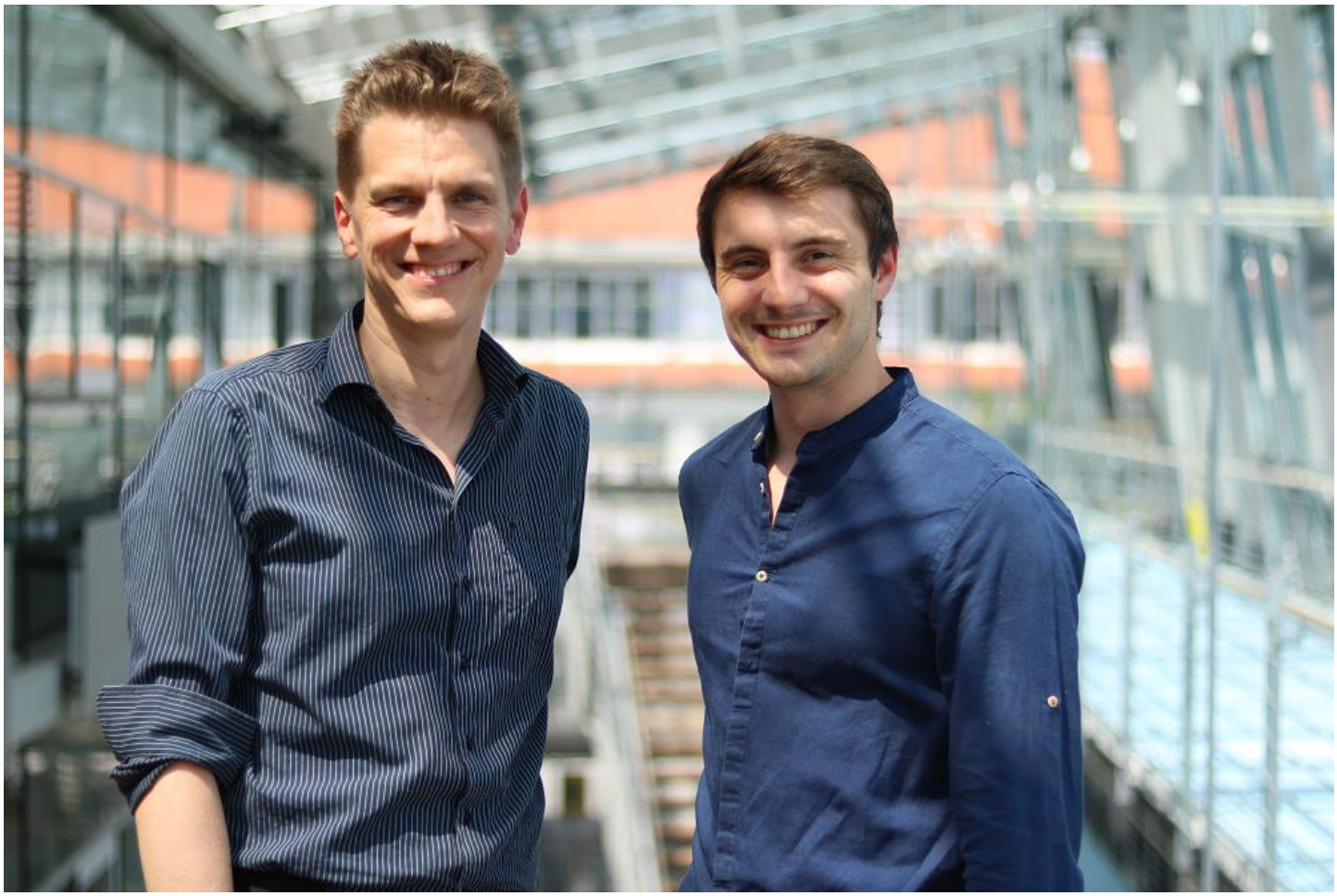
level label) instead of each single instance (segmentation mask). The approach makes it possible to generate a single activation map for each one of the nine AMD-related lesion types. Those masks then feed into global max-pooling which has one advantage over Grad-CAM approaches: it provides more intuitive explanation maps. Additionally to the lesion maps, the pipeline produces two more outputs: a vector revealing the presence or absence of a lesion and the final AMD diagnosis.

In case you want to use their method for your future work, I collected some additional technical details of their setup worth mentioning: they used a VGG-16 backbone but according to José any backbone could be used since their method is model-agnostic. Here, it is important to exclude the last max-pooling layer from the backbone which would otherwise result in a too small activation map size. The adapted backbone then feeds into a 1x1 convolutional layer with nine output channels, one for each lesion. You can get more information about their work here.

I want to thank José again for the interview and wish him the best of luck for his ongoing PhD journey where he focuses on multimodal and self-supervised learning for retinal imaging! If you are interested in his work and are attending MICCAI 2023, I would recommend to keep an eye out for "Self-supervised learning via inter-modal reconstruction and feature projection networks for label-efficient 3D-to-2D segmentation"! Additionally, you can find a publicly available version here and a code implementation here!



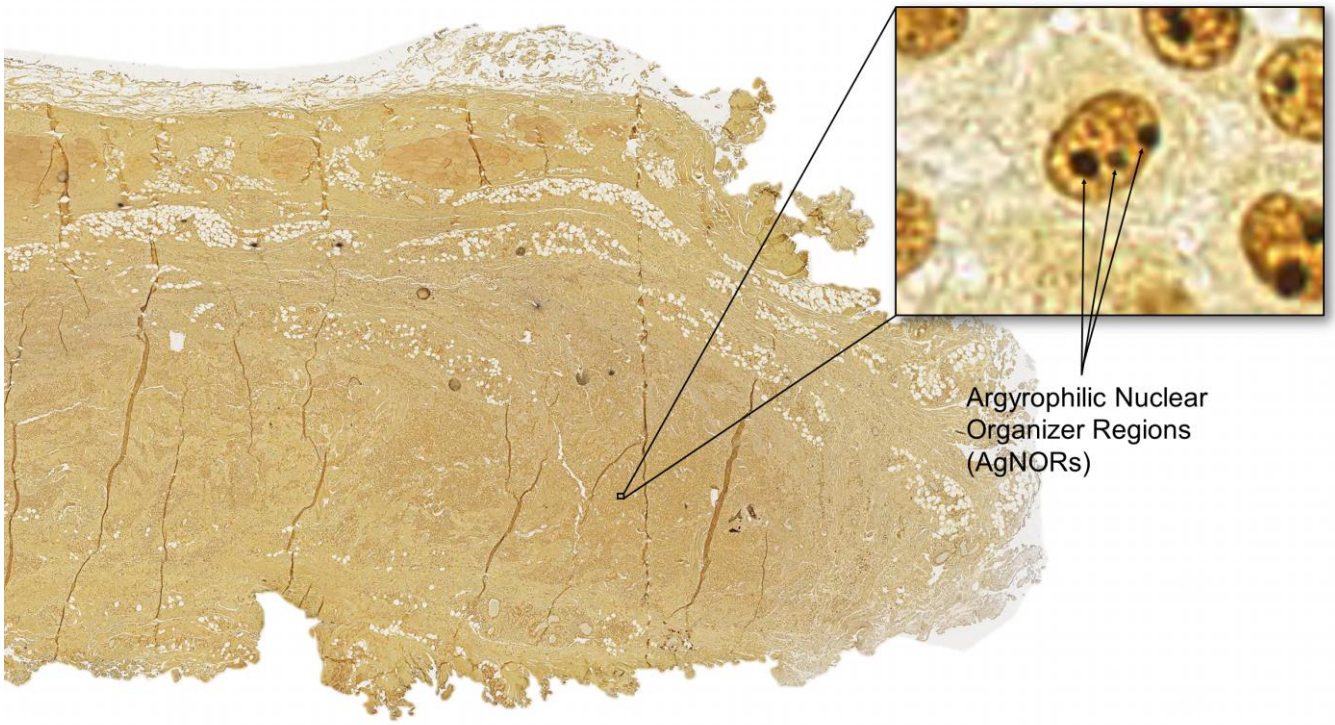
Deep Learning-Based Automatic Assessment of AgNOR-scores in Histopathology Images



Marc Aubreville (left) is a professor at the Technical University Ingolstadt of Applied Sciences in Germany, where Jonathan Ganz (right) is a PhD student, with a co-supervisor at FAU Erlangen-Nürnberg. They speak to us fresh from winning the Best Paper Award at BVM 2023 last month.

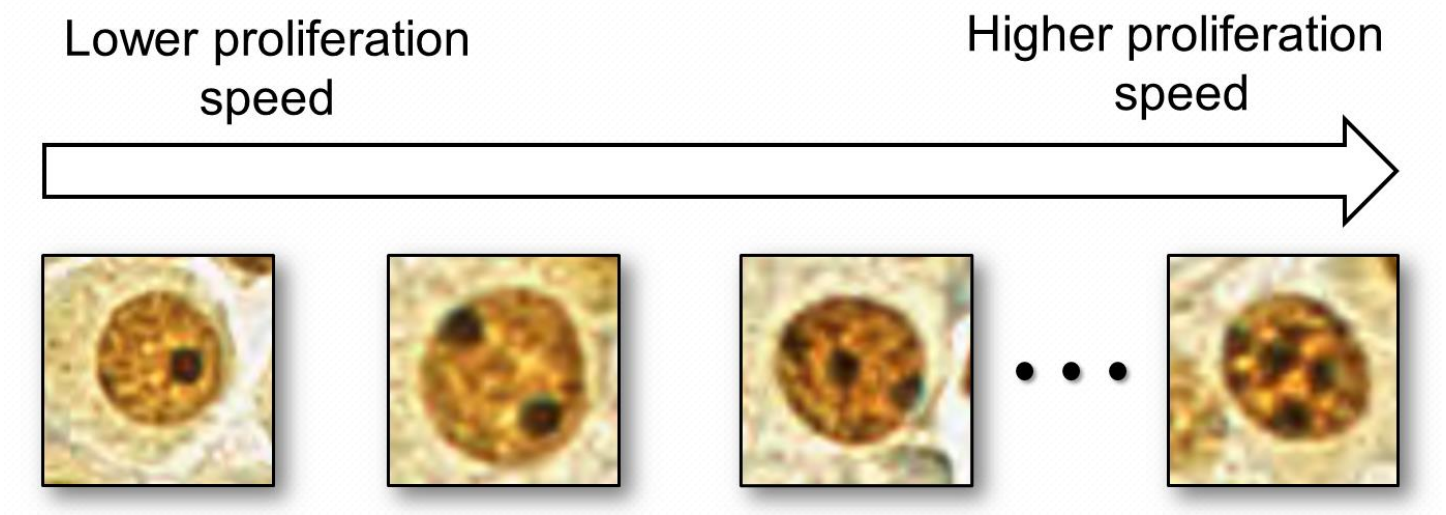
Accurately measuring **cell proliferation speed** is important for understanding the **aggressiveness of tumors**. A key element in this assessment is the **argyrophilic nucleolar organizer regions (AgNORs)** found within cell nuclei, which are correlated with cell proliferation. More AgNORs mean faster proliferation.

This paper explores the automatic assessment of AgNORs from **histopathology images**, paving the way for more precise and informed **tumor diagnosis**. Alongside other methods, such as **Ki-67** and **counting mitotic figures**, AgNOR-scores offer an additional layer of **explainability**, shedding light on the pace at which cells divide.



Exploring AgNORs as a viable assessment tool is no simple task. Jonathan and Marc worked closely with collaborators in **veterinary pathology**, including **Christof Bertram**, who was passionate about the topic and instrumental in pushing it forward. The team used **supervised learning** to establish a substantial canine dataset with roughly 23,000 annotated cells.

As AgNORs are primarily counted under **light microscopes**, limited information exists regarding how accurate humans are at carrying out the task. The researchers conducted a **human rater experiment**, enlisting pathologists for a **comparative study**, allowing a **unique assessment of the algorithm's performance against human raters**, which resulted in valuable insights into the reliability of both methods.



Check out our Video Interview with Marc and Jonathan

This research has established a new task in the domain of **machine learning**. Many studies focus on well-established tasks with good baselines and data, whereas this work tackles an essential yet unexplored task due to a **lack of prior baselines**. Also, despite its importance in predicting outcomes, the tedious nature of AgNOR assessment has thus far **prevented its integration into routine pathologist workflows**.

What do Jonathan and Marc think convinced the judges to award the paper the **top prize at BVM** this year?

"*Oh, that's a tough question,*" Jonathan remarks. "*We did science. We didn't over-advertise what we did. We know what our algorithm is able to do, and we highlighted our limitations.*"

Marc adds: "*It was just a **rock-solid science paper**. It was targeted toward **insights, not methods**. We wanted to find out stuff, see if that's possible, and how well we do against a human baseline. It was really just the science focus of the paper that the reviewers agreed was actually good. They were happy it wasn't yet another new method that does yet another 1% increase in something, which is, I think, what **many people are a little bit tired of!***"

Check out our video interview to learn more about this work, including Jonathan and Marc's ideas for extending it, their ambitions to work on different focal planes, and their regular work around tumor biology.

Best Poster Award at BVM

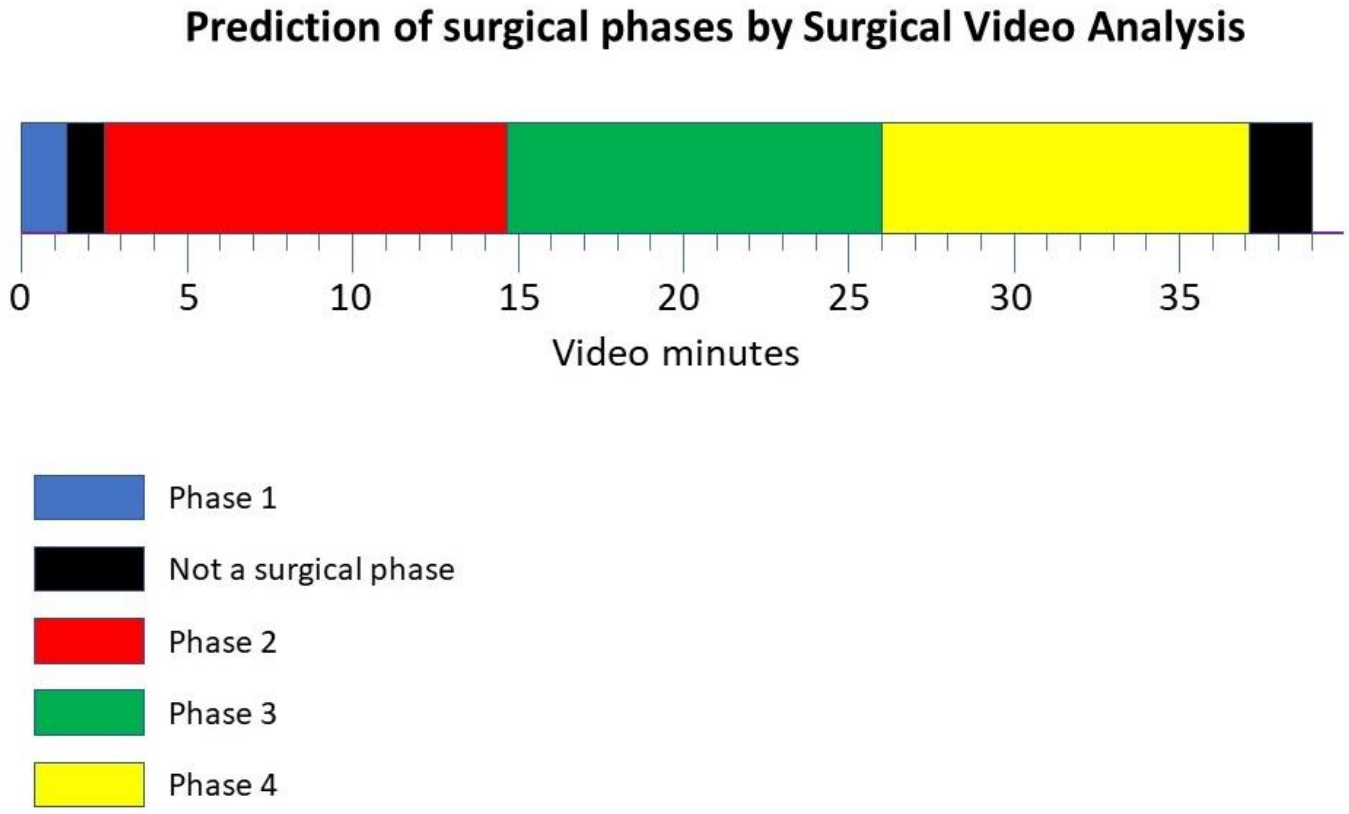
Robert Mendel of the Regensburg Medical Image Computing (ReMIC) group of Ostbayerische Technische Hochschule Regensburg was awarded by the audience the prize for the best poster at the BVM conference. Robert's poster "Exploring the Effects of Contrastive Learning on Homogeneous Medical Image Data" identified weaknesses of contrastive learning in the medical imaging domain and proposed sampling and masking strategies adapted to the domain's characteristics. Congratulations!

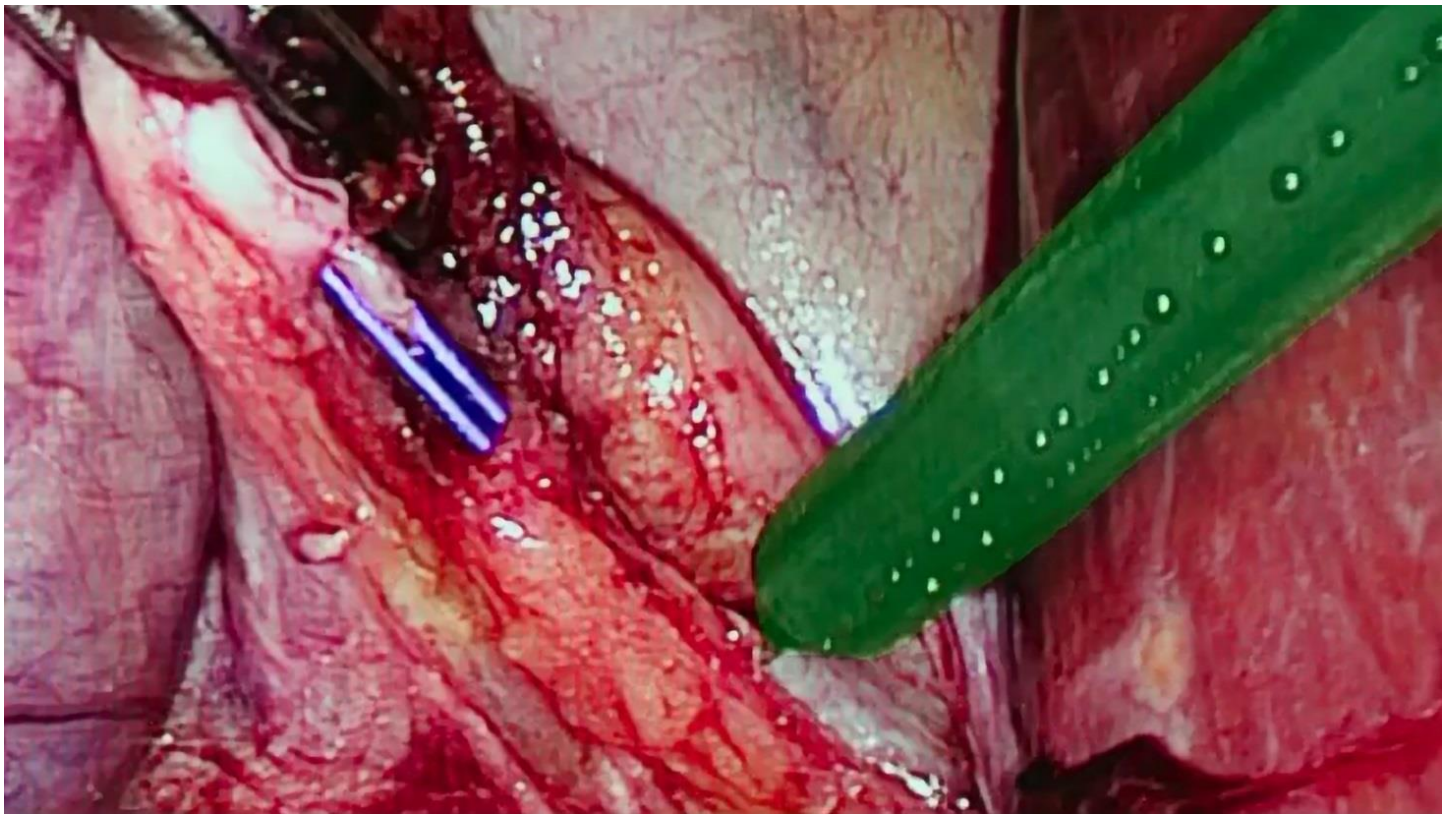
Surgical video analysis involves using artificial intelligence and machine learning algorithms to analyze surgical video footage. This practice, which includes both intraoperative and postoperative video analysis, has numerous benefits for patients, for surgeons and for other medical professionals as well. This is what we use at RSIP Vision.

We spoke with **Asher Patinkin**, one of the knowledgeable experts in this field at **RSIP Vision**. He provided a full review of the most advanced AI and Computer Vision algorithms that can be used for surgical video analysis. Depending on the specific requirements, **Deep Learning** algorithms, such as convolutional neural networks (CNNs), RNNs can be trained on large datasets of surgical video footage to perform tasks such as object detection, tracking, segmentation, and activity recognition, as follows:

Object Detection – Asher points out that Algorithms like YOLO, Faster R-CNN, SSD (Single Shot Detector) and RetinaNet help identify specific objects or instruments within surgical video footage, such as surgical tools, implants, or anatomical structures.

Tracking – Here we distinguish between traditional algorithms (like Kalman Filter, Mean Shift, Particle Filter) and more recent Deep Learning algorithms, which follow the movement of objects or instruments within the surgical video footage over time, allowing for analysis of the trajectory and motion of these objects.





Pose Estimation – Pose estimation and Mask R-CNN allow to estimate the position and orientation of instruments or anatomical structures within the surgical field, allowing for analysis of surgical technique and instrument placement.

Segmentation – Asher says that U-Net, Mask R-CNN, FCN separate objects or instruments within the surgical field from the surrounding environment, allowing for more precise analysis of their movements and interactions.

Activity Recognition – It is Asher's opinion that Transformers and 3D CNN enable to identify and classify specific surgical actions or tasks performed within the surgical field, allowing for analysis of surgical workflow and technique.

RSIP Vision's AI experts and engineers have both the knowledge and the experience to respond to your specific needs in Video Analysis of surgeries and all medical procedures with AI.

Calibration Techniques for Node Classification Using Graph Neural Networks on Medical Image Data

Iris Vos (left) is a fourth-year PhD student at UMC Utrecht in the Netherlands.

Her work on the risk prediction of aneurysm development in the brain has just won the Runner Up Best Poster Award at MIDL 2023.

Graph neural networks (GNNs) have emerged as a promising approach to enhance the **accuracy and efficiency of predicting aneurysm development in the brain**. This innovative method represents the brain's vasculature as a graph, providing a unique understanding of the structure of vessels and potential risks for aneurysms.

In recent years, **deep neural networks (DNNs)** have been shown to be prone to miscalibration, leading to some unreliable predictions. Overconfident probability estimates often cause this miscalibration. By contrast, **GNNs tend to be underconfident in their predictions**.

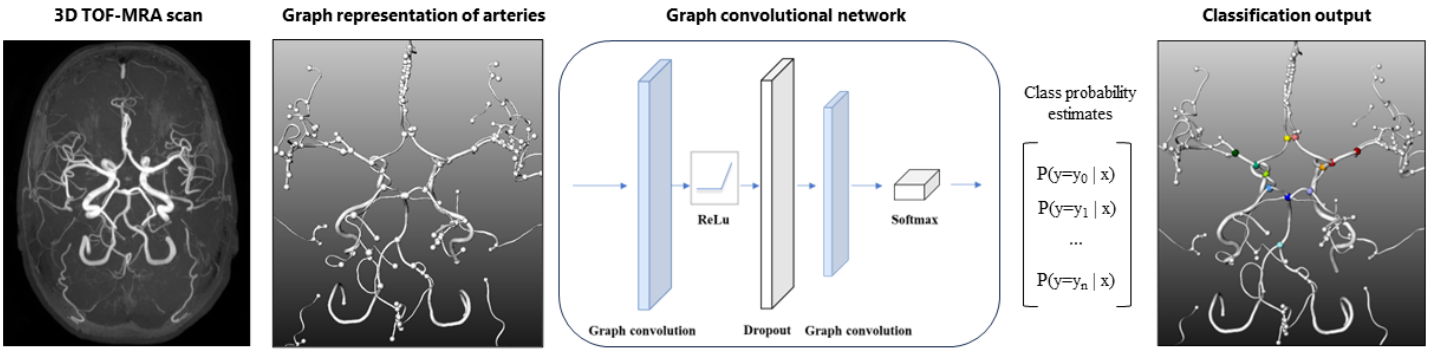
Previous research has attempted to mitigate the issue of underconfidence in GNNs, but the effectiveness of these techniques remains largely untested in the context of medical image data. This paper aims to address that by determining whether **calibration techniques applied to overconfidence in DNNs could be generalized to fix underconfidence in GNNs** trained on medical image data.

"GNNs are still quite a new topic, and especially for medical image data, it's difficult to get good graphs because they rely heavily on the segmentations," Iris tells us. *"I focus on the* ***circle of Willis, an anastomosis of blood vessels in the brain****. They have a lot of anatomical variety among healthy people as well. The brain vessels form this kind of circle. Only 30% of the people in a healthy population will have this complete circle, and 70% will have some anatomical variance. Blood vessels can be absent, or certain vessels can be duplicated or underdeveloped. To get good graphs, you need good segmentations, especially of the smaller vessels. If they're underdeveloped, it's quite difficult."*

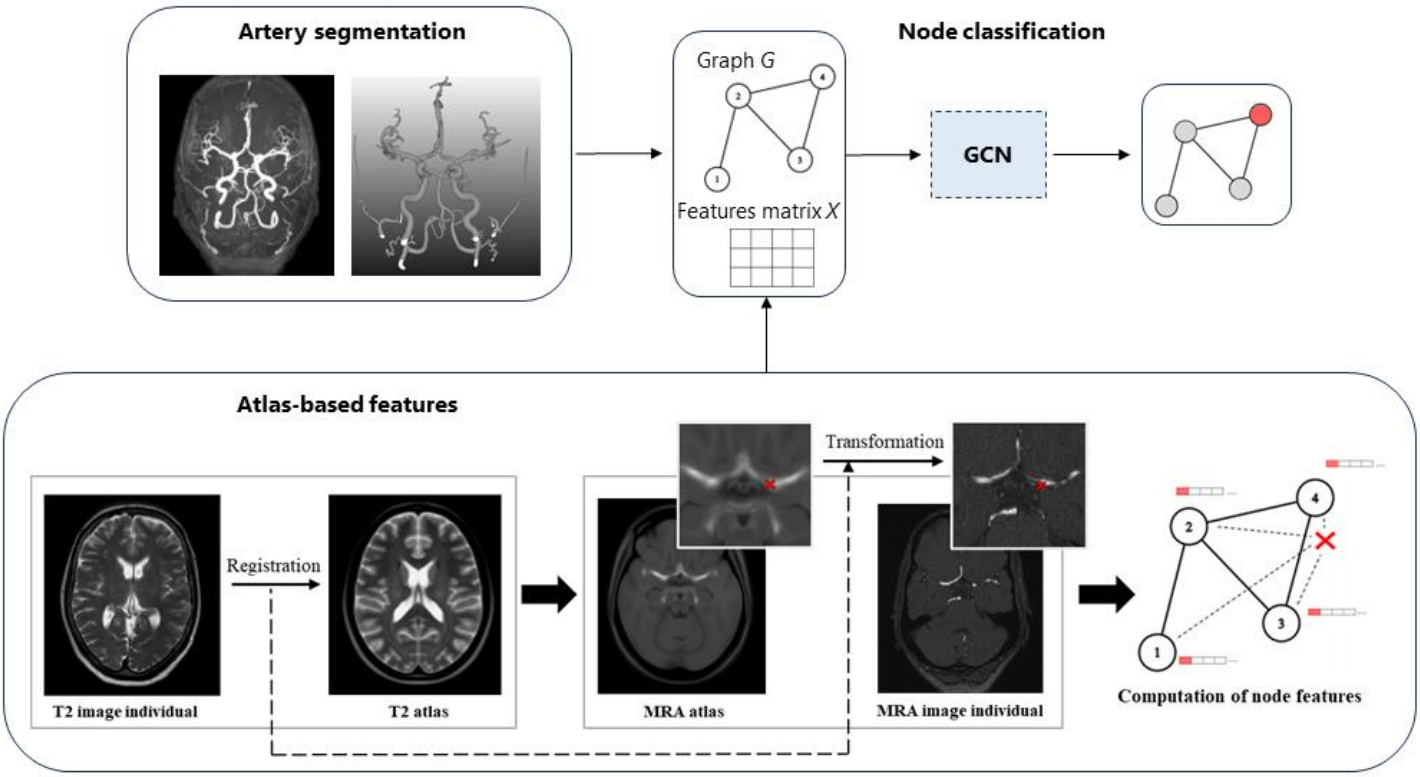
Iris's more observational study focuses on the **applicability of calibration techniques to graphs** and finds that the methods are indeed effective. However, the segmentation challenge is ongoing for researchers. Current and future efforts, including a **MICCAI** challenge, are focused on **segmenting the intracranial arteries** and developing more accurate vessel segmentation and graphs moving forward.

"We used node classification for GNNs," Iris explains. *"We looked at the most vanilla GNN that there is at the moment, but we also focused on* ***higher-order graph convolutional networks*** *because a limitation in these GNNs is that they focus on local regions. They only learn the embeddings of direct neighboring nodes of a target node. But they fail to capture global patterns in the data. The higher-order graph convolutional networks add information beyond direct neighborhoods, and* ***show better performance in both discriminative power and calibration****."*

Calibration is a crucial aspect often overlooked in pursuit of high accuracy and discriminative power in classification tasks. It becomes particularly important when the primary goal is not simply distinguishing between two classes,



Good calibration means that a model is confident about accurate predictions, while also indicating low confidence when it is likely to be inaccurate. In contrast to most deep neural networks, which are often overconfident in their probability estimates, graph neural networks tend to be underconfident.



Graph neural networks learn by exchanging information between local neighborhoods of nodes. By adding information on a global scale, using features based on a statistical brain atlas, we were able to improve the performance of node classification

for instance between malignant and non-malignant tumors. In the context of aneurysm development, the objective is to **determine subgroups within the population to identify individuals at higher risk**. The confidence estimates produced by neural networks play a pivotal role in clinical decision-making.

"*If you use neural networks, and the produced confidence estimates are over or underconfident, it can lead to real issues in the clinic,*" Iris points out. "*If we want to identify if a certain person is at risk of developing an aneurysm, and we say it's 70%, you need to know that your model is not over or underconfident.* ***We want to use these models to decide whether or not we perform follow-up screenings of at-risk individuals!***"

Although other techniques have already proved to work well, she hopes people will incorporate the essence of this research in their work. It demonstrates how relatively simple it is to calibrate a model or at least report on its uncertainty rather than reporting solely on how a model has obtained a degree of accuracy. **Evaluating the uncertainty of a model brings it closer to clinical acceptance.**

Last year, Iris worked on an interesting project using **GNNs for automated intracranial artery labelling, using an atlas to extract atlas-based features that were used as input for node classification**. It was named a finalist for an award at the SPIE conference.

Does she have any thoughts on what it was about her work this time that convinced the judges to commend it so highly?

"I don't know for sure, but I feel it's because it's not that mathematically complex or innovative; it's just an application," she ponders. *"MIDL focuses a lot on reproducibility and scientific impact, and* ***this study is quite easy for other scientists to get some information out of*** *and get some take-home message and then apply some techniques.* ***We also put the code online for the benchmark dataset****, so it's reproducible. But yeah, I was still surprised!"*

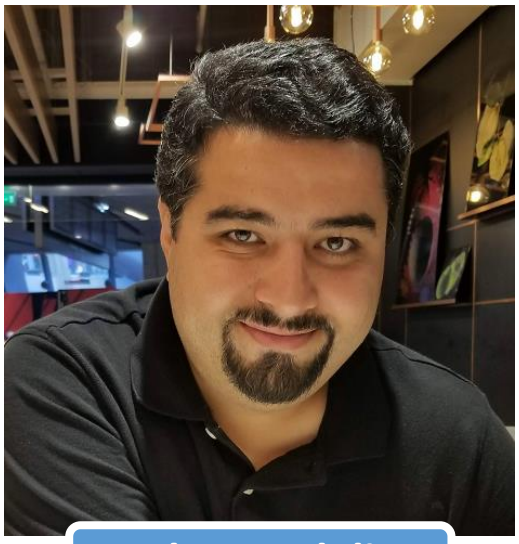
Iris's supervisor on this work, **Hugo Kuijf**, should already be familiar to readers of this magazine. She tells us his philosophy is that PhDs should be hard work but also fun.

"It felt more like a collaboration than that he was my supervisor," she recalls. *"We got along really well. He taught me a lot about supervising students and how to best motivate a student that's not doing well. He's really good in his communication, in his soft skills."*

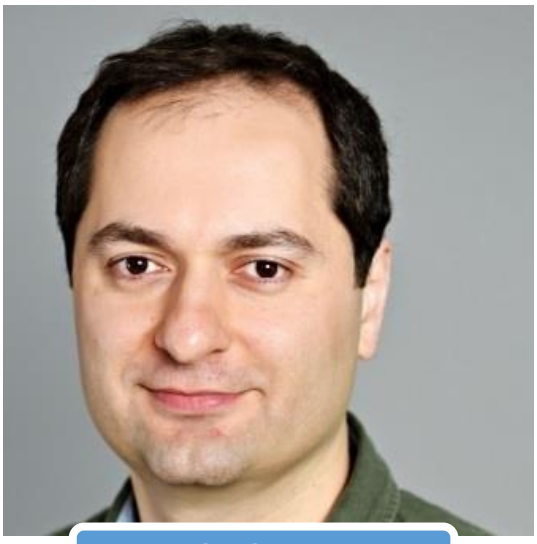
Looking ahead, she is focused on **mesh neural networks**, a novel approach in medical image data analysis, aiming to **identify patterns in vascular structures across individuals to identify different subgroups with varying risks for aneurysm development**. Also, combining different fields to obtain the best model – for example, CNNs or meshes with graphs.

"In the end, we want to obtain ***a personalized risk model for each patient****,"* Iris reveals. *"Then, based on that model, we can see if we should perform follow-up screening and how often. This way, we hope to detect aneurysms in an early stage so they can be treated before causing damage."*

Iris with the other award winners at MIDL 2023



Ehsan Adeli



Babak Taati

Ehsan Adeli is an Assistant Professor at the Stanford School of Medicine in the Department of Psychiatry and Behavioural Sciences. He is also affiliated with the Computer Science Department and works in the Stanford Vision and Learning Lab.

Babak Taati is a senior scientist at the KITE Research Institute, the research arm of the Toronto Rehabilitation Institute, an adult rehabilitation hospital which is part of the University Health Network, a network of research hospitals affiliated with the University of Toronto. He is also an associate professor (status only) in the Department of Computer Science at the University of Toronto, with a cross-appointment in the Institute of Biomedical Engineering, as well as a Vector Institute Faculty Affiliate.

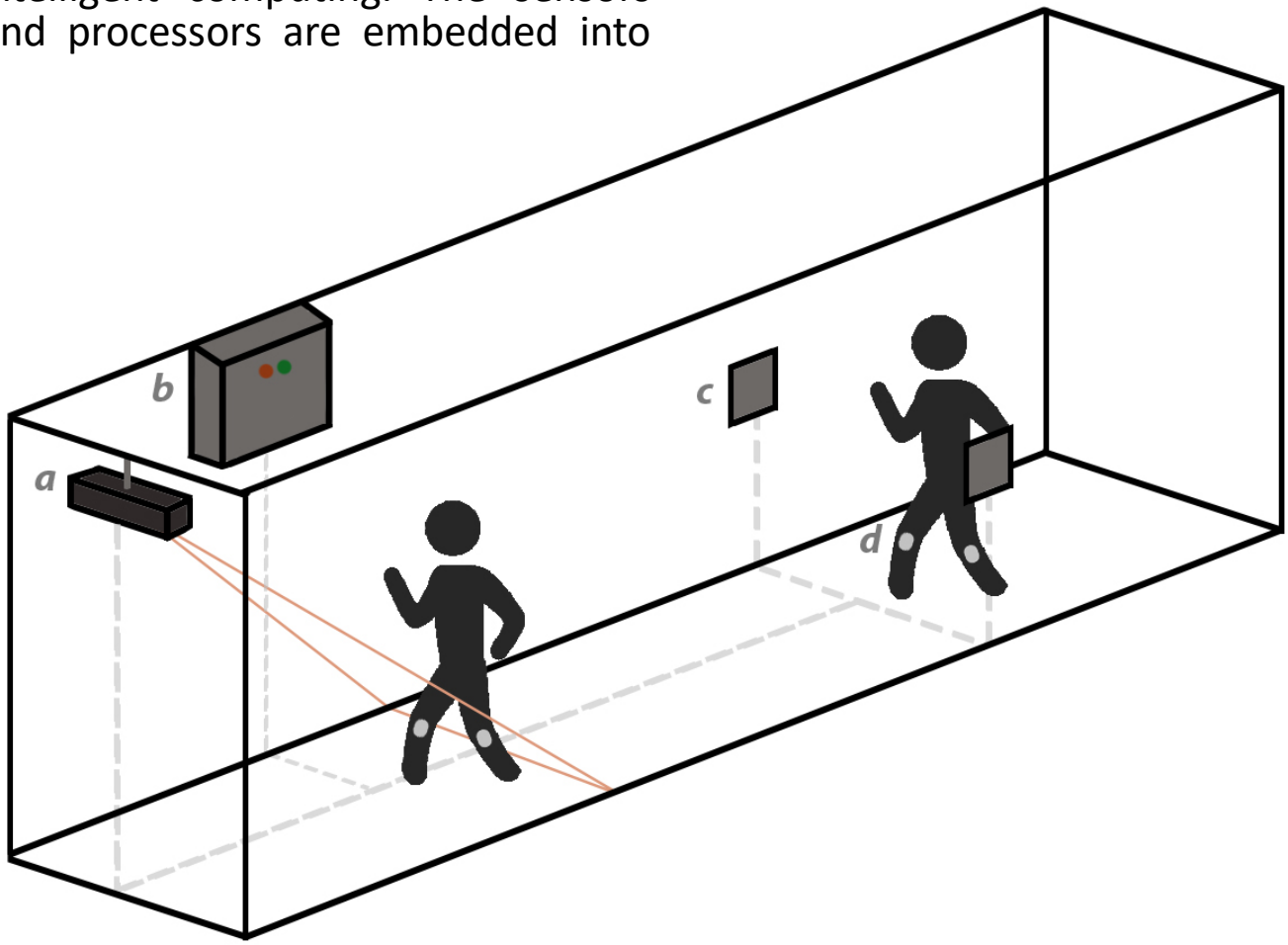
Ehsan and Babak speak to us as the co-organizers of an exciting MICCAI 2023 workshop.

Ehsan: Based on the MICCAI workshop chairs' decision, we are going to have a joint workshop with the AICAI workshop. That's the Ambient Intelligence for Healthcare and Computational Affective Intelligence for Computer-Assisted Interventions workshop.

Babak and I were discussing doing a workshop in the medical world that uses more sensor-based type technologies for healthcare applications. Both of us do have a lot of research and publications in this space. After brainstorming, we decided to propose this workshop to MICCAI 2023. The main goal of this workshop is to advance knowledge and technology of intelligent environments for healthcare. The topic is very broad. It's not limited to any specific sensor modality or any

specific disease; it's more using the integration of ambient sensors with any sort of medical imaging, radiology, neuroscience, or neurology applications, and basically understand the disease and disease progression in daily living spaces or generally in daily environments.

I'm not sure if ambient intelligence is a very clean and clear term to the MICCAI community, but as we have it in our workshop page as well, ambient intelligence defines the terms and technologies for intelligent computing. The sensors and processors are embedded into everyday devices and everyday environments and will seamlessly adapt to the users' requirements. This ambient intelligence technology is being used in retail spaces, education, office spaces, and so on, and it's finding its applications also in healthcare for quality improvement, senior care at home, ensuring clinician and patient safety, rehabilitation, and so on. The goal of this workshop is to shed light on this novel technology and how it could be used for the healthcare space in their skill space.



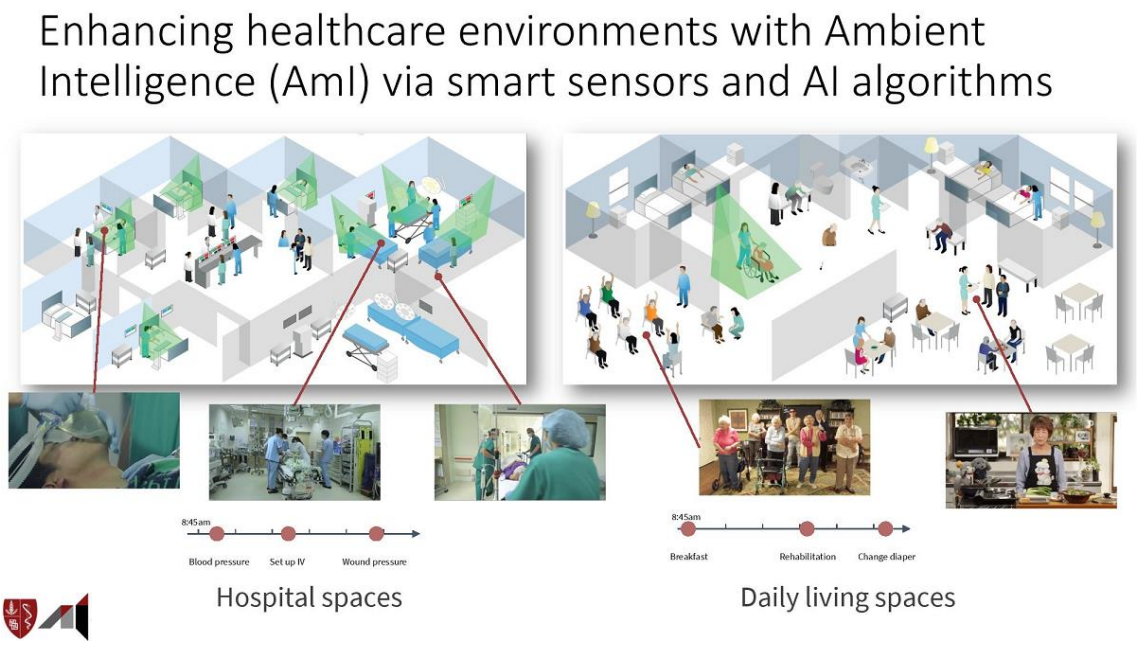
The ambient monitoring system includes (a) a Kinect v2 sensor installed in a hallway, (b) two RFID antennas mounted on the hallway walls, (c) RFID tags attached to participants' clothing at knee level, (d) and a control box consisting of a laptop and an RFID reader. Image courtesy of Vida Adeli et al. [Ambient Monitoring of Gait and ... <https://ieeexplore.ieee.org/document/10102699> reproduced with permission].

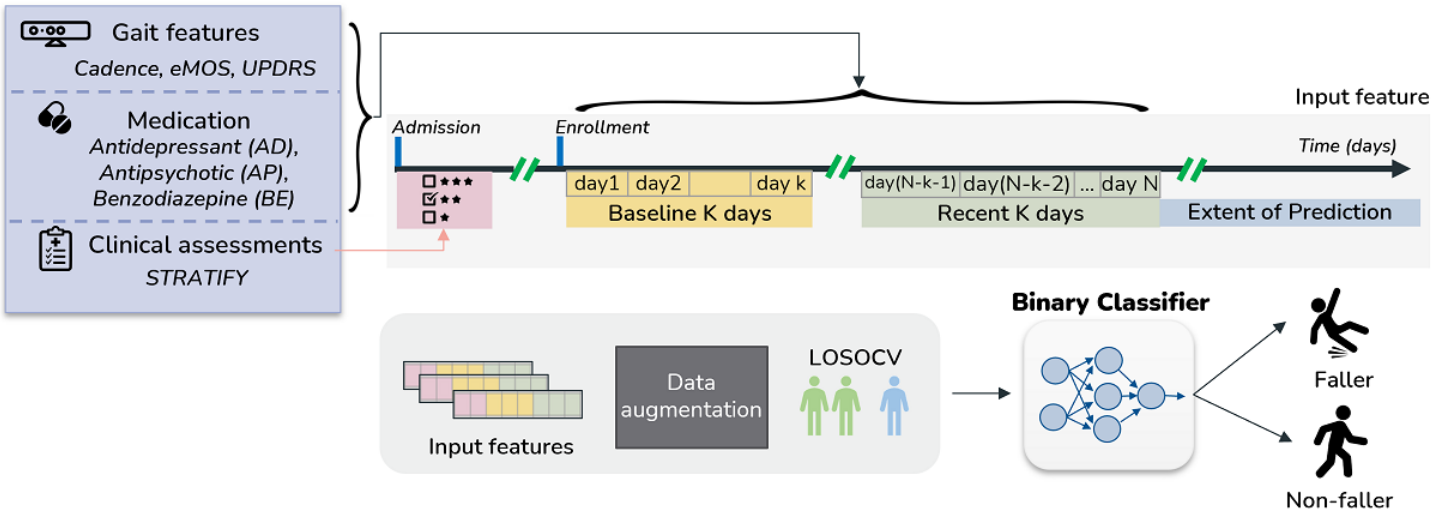
"I think healthcare is the new funder. I think the next multibillion-dollar company is probably going to be in the healthcare space!"

Babak: The areas of work where ambient technology can be used are very diverse. It can be in the home, but it can also be in hospital settings or other clinical settings. For example, long-term care homes or retirement homes, where individuals living are typically a lot older or may have comorbidities. Also, one of the papers accepted to the workshop is ambient sensing in the operating room, looking at how the surgery is going or which areas doctors and clinicians in the operating room are paying attention to. There are a couple about gait analysis using ambient sensors to monitor gait, and that could be for gait assessment, fall risk assessments, and things like that. The areas in which ambient assessment can be used are pretty diverse.

There are so many workshops at MICCAI, what would be a great reason to join yours?

Ehsan: I've been involved in the MICCAI community for several years as an organiser, area chair, and also organiser of several workshops. Most of the technologies and methods presented in MICCAI are around medical imaging and some interventions using other sensor modalities, videos, and so on, from mostly radiology perspective, or the CAI component also looks at operating rooms and so on. This workshop is planning to look at the medical and healthcare applications from a different perspective using more recently developed types of sensors and sensing technologies that could be used for other applications, but in this workshop, we are exploring their application in the healthcare space.





Having the capability of having similar types of infrastructures and set up to do monitoring and quality improvement in hospitals, ICUs, operating rooms, and so on, and at the same time at homes, gives us the capability to broaden the development of healthcare-related applications and algorithms. One very important aspect of this workshop would be exploring this multi-interdisciplinary field from novel sensor technologies to pervasive computing, AI algorithms, and AI applications. It's a wildly interdisciplinary field. We hope to start this in the MICCAI community. Our goal is to actually take it out as well and probably do this same workshop in broader like NeurIPS types of communities as well, but our goal was to start it within the medical and MICCAI community.

One thing I want to add, and this is my personal interest in this field, is trying to build and develop what we call infrastructure-mediated sensing for healthcare applications. This is something that I'm doing a lot of research on right now. We are applying those to mostly neurological diseases. I do have some projects at Stanford working with the hospital building ME intelligence in ICU rooms at Stanford Hospital, but generally, building this infrastructure component of it is basically what is currently of my interest and hopefully the interest of a lot of young researchers who will join the workshop.

Babak: Ehsan has also done some really interesting work on multimodal analysis between ambient sensing, for example, video ambient sensing data and medical imaging data, so looking at brain scans of individuals with Parkinson's disease, for example, and at the same time looking at their activities or how they ambulate or walk or move their body parts. That's a very interesting area of research as well, and Ehsan has done some really interesting work.

Another area of interesting multimodal analysis is combining ambient sensing with electronic health data. For example, if you want to analyze a person's health,

you can look at how they're walking and what their facial expression is, but also look at their electronic health record to see what conditions they're diagnosed for, when was the last time they fell, what medications they're on, and so on. That's good information to include in machine learning models to improve accuracy and performance.

Do you have any suggestions for researchers who are looking for a new field of research and would like to make progress in this field?

Babak: There are lots of interesting things to work on. What I like about this field is that it is very interdisciplinary, so it's a lot of collaboration between computer scientists and engineers and clinicians of different fields of practice, but it also brings about lots of really interesting computer science problems. For example, in computer vision, human movement tracking, human gait analysis, pose tracking, facial expression analysis, facial recognition, privacy, fairness, and things like that. A lot of the challenges are in moving technology that is developed in the lab and trying to make it robust and reliable enough to be deployable in real life in the wild applications. That's often quite challenging but also rewarding to work on to try to get something that is sort of working in the lab but try to make it reliable and robust enough for deployment.

Also, be aware of things like fairness. Does the algorithm work equally well for men and women or people with different skin tones? Does it work well for old versus young? If it works well for the faces of young people, does it work well for people who are old and have facial wrinkles and things like that?

Is there anything else you think our readers should know?

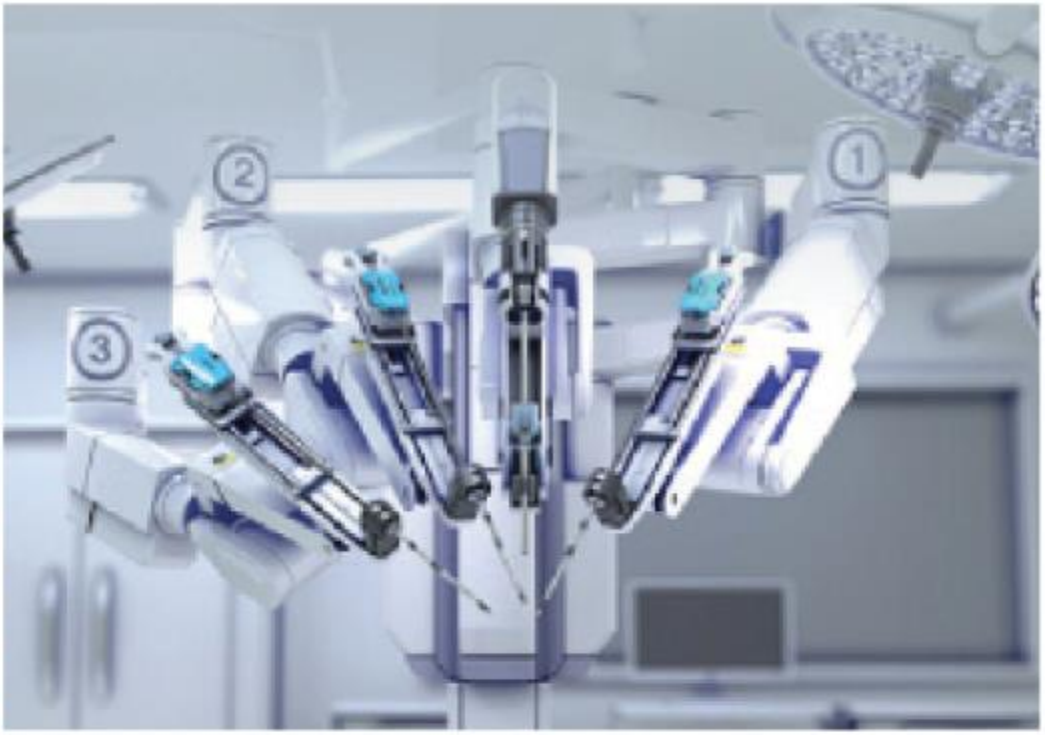
Ehsan: I think healthcare is the new funder. I think the next multibillion-dollar company is probably going to be in the healthcare space. Because of that, exploring these interdisciplinary and multidisciplinary fields and their application to healthcare spaces is something to be very much aware of. That's basically one of the goals of this workshop.

In terms of research, I would say one of the things that I think is going to be explored widely in the next couple of years or so is the use of multimodal sensing technologies and foundation models. Building large-scale foundation models with this multimodal sensing that ambient intelligence can provide. That's the infrastructure, that's where the data is coming from, and now we can build these large-scale foundation models and use them for downstream tasks and applications of healthcare. In addition to what Babak said, this is another very interesting research field.

Beyond that, this is going to be a very interesting workshop with a number of very good speakers in the field. I'm inviting everybody to join us in MICCAI, October 8, in Vancouver, in person, and hopefully, there will be a hybrid component as well.

And the winner is...

Our community admires awesome Emma Robinson for many reasons. The Faculty of Life Sciences & Medicine at King's College London too! So they gave her the Research Guidance & Mentorship Award because she "*fosters a culture of collaboration and positivity and provides exceptional pastoral care to her students and junior staff*" and she is also "*a true role model*" in her field. We couldn't say it any better!



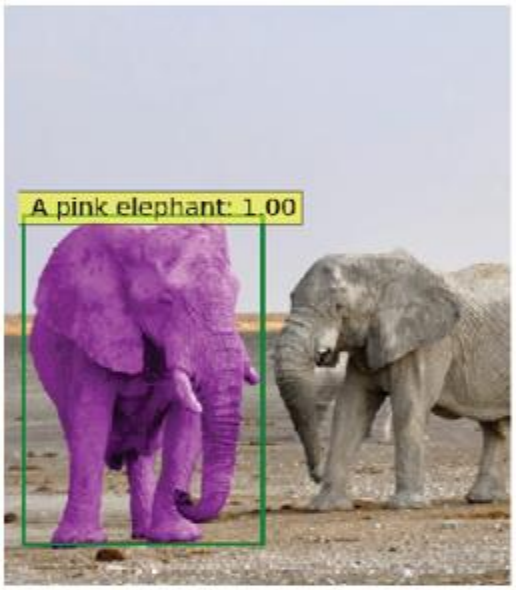
IMPROVE YOUR VISION WITH

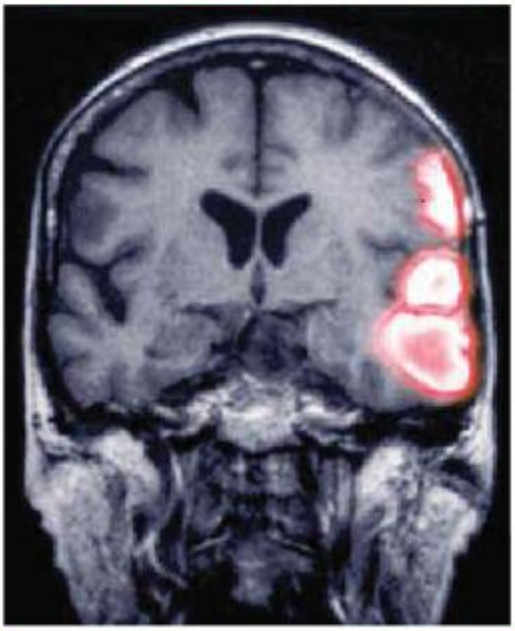
Computer Vision News

SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!





RSIP VISION