September 2023

Computer Vision News

& Medical Imaging News

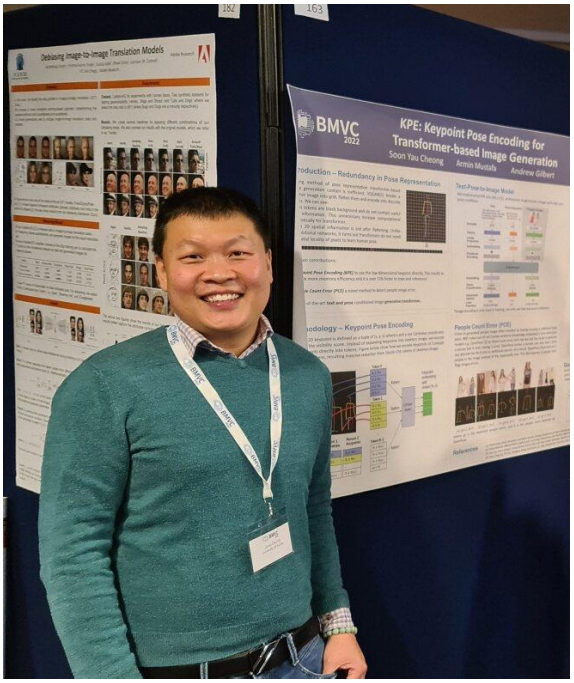
The Magazine of the Algorithm Community

A publication by

RSIP VISION

pixee medical

UPGPT: Universal Diffusion Model for Person Image Generation, Editing and Pose Transfer



Soon Yau Cheong is a second-year PhD student at the University of Surrey, under the supervision of Andrew Gilbert and Armin Mustafa.

He tells us about his novel multimodal diffusion model for text, pose, and visual prompting.

This work will be presented next month at ICCV 2023 during the Computer Vision for Metaverse workshop.

In this paper, Soon proposes **a new diffusion model that can fuse three distinct modalities simultaneously to generate and edit images of people**. Previously, separate models have been used to generate images from text or transfer a person's appearance from a source image to the pose of a target image. In a departure from those models that operate in isolation, **UPGPT represents the first-ever integration of text prompting, pose guidance, and visual prompting**, unlocking new creative possibilities for image generation and editing.

The purpose of integrating text, pose, and visual prompting under a single unified model is one of efficiency. Why rely on multiple disparate models when a comprehensive model can effectively harness these diverse inputs?

"*It makes sense,*" Soon affirms. "*Why do we have so many different models if one can do all? Even if you have different models, the base model already has a rich understanding of what a human should look like.* ***We have one model that can learn everything!***"

Achieving this was not without its difficulties. **Data availability** was a challenge Soon faced. He used the **DeepFashion** dataset, but the 3D pose information the model required was not inherently present. Ultimately, he had to employ other software to extract it, adding an extra layer of complexity to the process.

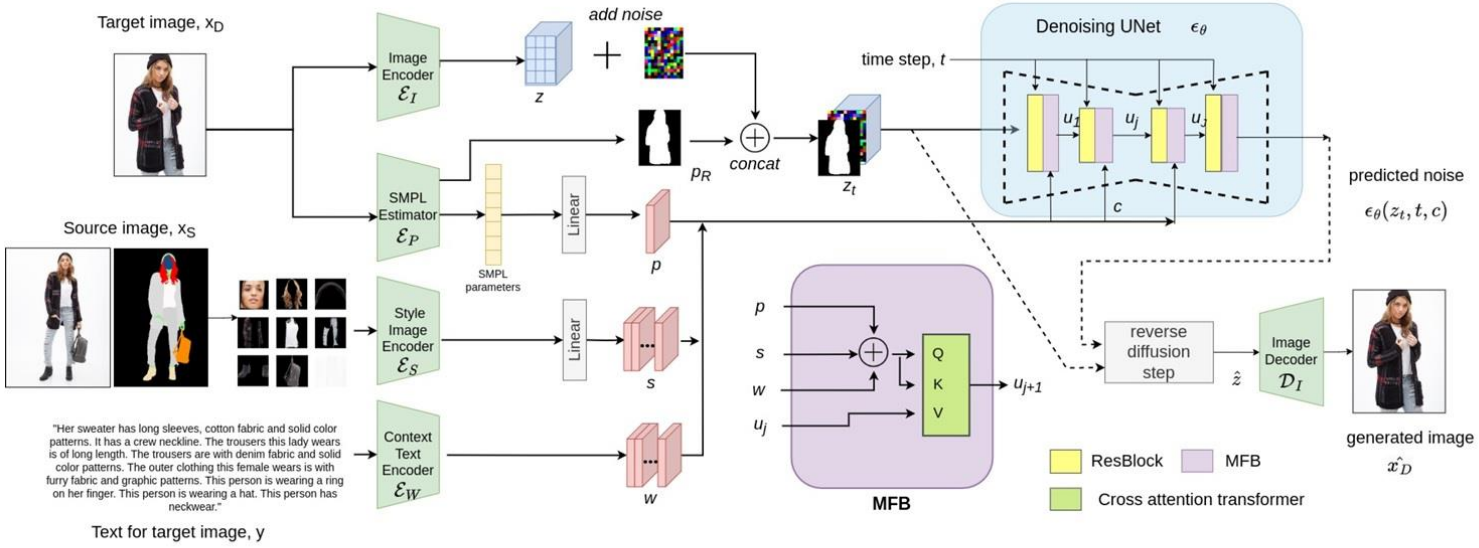


Figure 3: Overview of our proposed UPGPT architecture. In training, we encode pose, style image, and context text into embeddings that go to the Multimodal Fusing Block (MFB) for fusing. The output of MFB is used as a condition in UNet to predict the noise needed to denoise the image's latent. In sampling, the image encoder decodes the denoised latent $\hat{z}$ into pixel space.

Another challenge stemmed from the **computational demands of training the model**. The iteration was long, and Soon was not working with a big budget or a large GPU, so with limited resources, innovative solutions were necessary.

"The obvious choice is you get a bigger GPU, but we didn't have a bigger GPU, so I had to be smart," he recalls. *"I modified the model a little to make it a bit smaller.*

*Also, we used **a 3D body model**, unlike conventional methods that use the whole 3D body map. If you have to map out every single point on your body surface, that takes up a lot of memory, but we used the **SMPL model**, which contains 10 parameters for body shape and 72 for the rotation of your joints. Instead of trying to run every single thing, you just need to run: What is the angle of this rotation? **We were able to reduce the computational requirements a lot!**"*

The extraction of 3D pose information from given images involved using a library to estimate the 3D body shapes and poses. **Deep learning methods**, including segmentation, were harnessed to predict silhouette masks, enabling precise positioning of elements within the generated images. Conventional approaches require **fine-grained semantic segmentation**, where you must segment every pixel of the different body parts.

"The off-the-shelf semantic segmentation models were not very good," Soon explains. *"It's very difficult to get the segmentation correct, especially if a person faces backward. Our methods don't use semantic segmentation, so we're able to avoid that problem. We just need **a very simple silhouette mask**."*

The integration of visual prompts adds another layer of sophistication

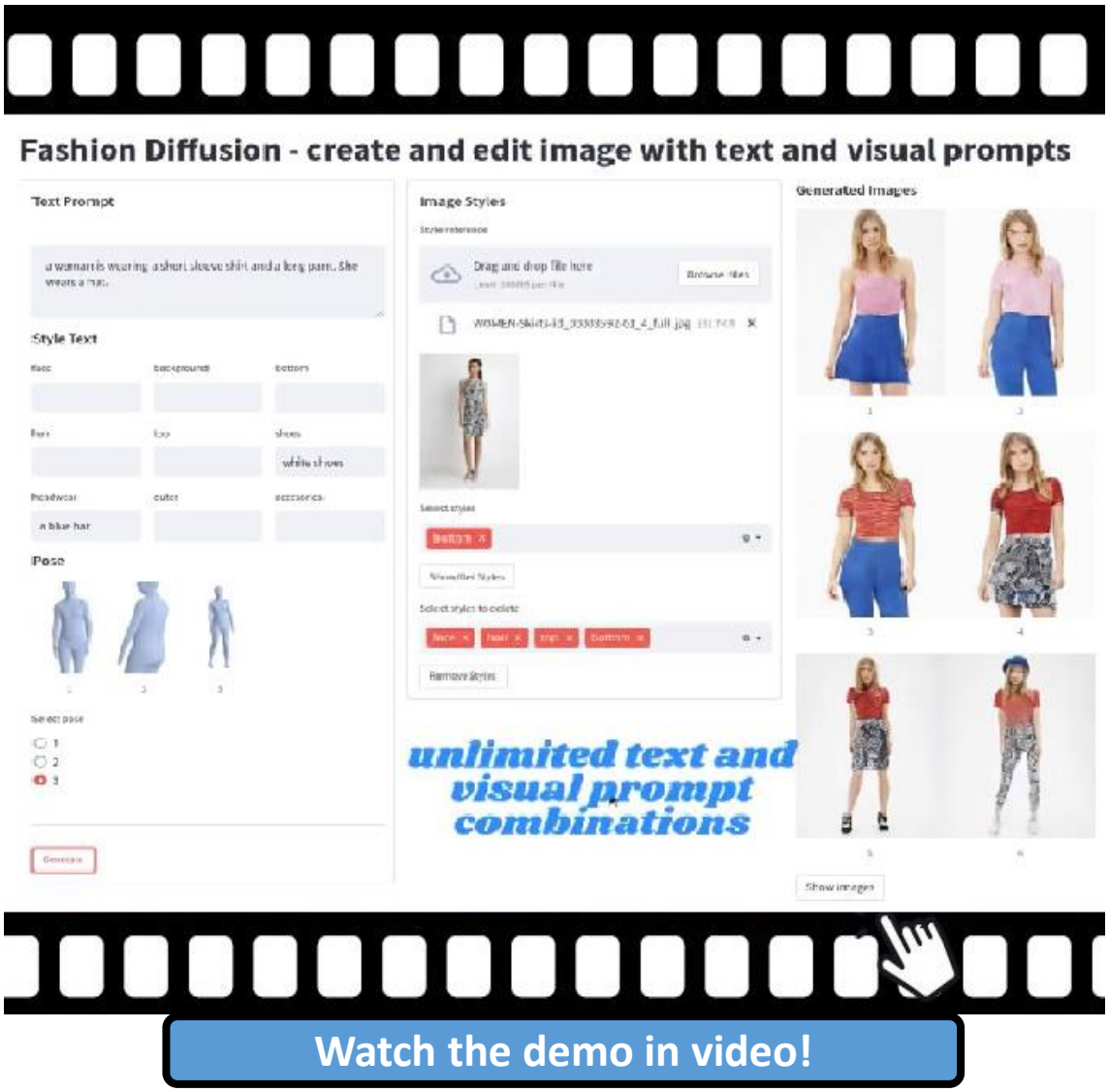
to the model. It can crop out people's clothes and send them into a new network to extract the features. That is used as conditioning so the model understands color and style preferences. For example, a text prompt to create a person wearing a blue shirt begs the question, what shade of blue? If it is light blue, how light is it? It is not easy to put that into words, but **a visual prompt enables a faithful representation**.

Another exciting aspect of this work is its ability to perform **pose and camera view interpolation**. Given two pictures, it can interpolate anything in between. **This groundbreaking feature, the first of its kind, facilitates the seamless generation of images that transition between poses and camera perspectives**.

What are the next steps for this work?

"Currently, we use a small data set with a plain background," Soon tells us. "Our next step is to apply that to

UPGPT represents the first-ever integration of text prompting, pose guidance, and visual prompting!



the real world. **Real people with real backgrounds**. I hope in my next work, I'll be able to create **high-resolution real-world images that can be used for commercial applications**."

This work will be presented next month at **ICCV 2023** during the **Computer Vision for Metaverse workshop, or cv4metaverse**. Its rights have now transferred to ICCV, and Soon will also be presenting a live demonstration as part of the event in Paris, which will be a unique opportunity to showcase it to attendees.

"*I have a **video demo** you can download and use on **GitHub**,*" he adds. "*The tech capability is particularly good. You don't need artistic skills to change the picture; you can just type in the words and change it. It's an all-in-one solution for people image generation and editing, and it's fun!*" The demo is on top of this page for you to watch.

Yunlu Chen has recently completed his PhD at the University of Amsterdam, under the supervision of Efstratios Gavves, Thomas Mensink, and Arnold Smeulders.

His research explored how 3D deep learning can effectively leverage the advantages of continuum and achieve better generalisation. Yunlu is now a postdoctoral researcher with Fernando De la Torre at Carnegie Mellon University.

Congrats, Doctor Yunlu!

Continuity and discreteness are long-standing characteristics when juxtaposing natural and engineered systems. The natural world exhibits apparent analogue and continuous material and energy flows at our observable scale. In contrast, computing systems are based on the **discrete nature of information processing**, employing binary digit values and quantized domains.

Traditional discrete 3D representations faces limitations in high memory and computation costs for high fidelity due to the need for elevated sampling rates as required by the **Nyquist-Shannon theorem**. To this end, the recent trend of modelling 3D signals with implicit neural representations is a new paradigm that utilizes continuous coordinate-based neural functions which are not bounded to any certain resolution. To understand how these representations encode 3D shapes, Yunlu's paper at **ICML 2021** investigated the **hidden-layer features in latent-coded implicit representations**, from which the emerging hierarchical structure is observed in the implicit network layers, such that the earlier layers encode coarse shape outlines, while deeper layers encode fine shape details (Fig. 1). In addition, this research suggests the representations' unsupervised acquisition of correspondence and semantic awareness, which facilitates generalization to a collection of 3D shapes. Furthermore, Yunlu's work at **ECCV 2022** extended the representation by injecting the design of equivariance and graph embedding, allowing high-fidelity encoding of 3D signals in local details, and generalization to unseen geometric transformations including (continuous) rotation, translation and scaling (Fig. 2).

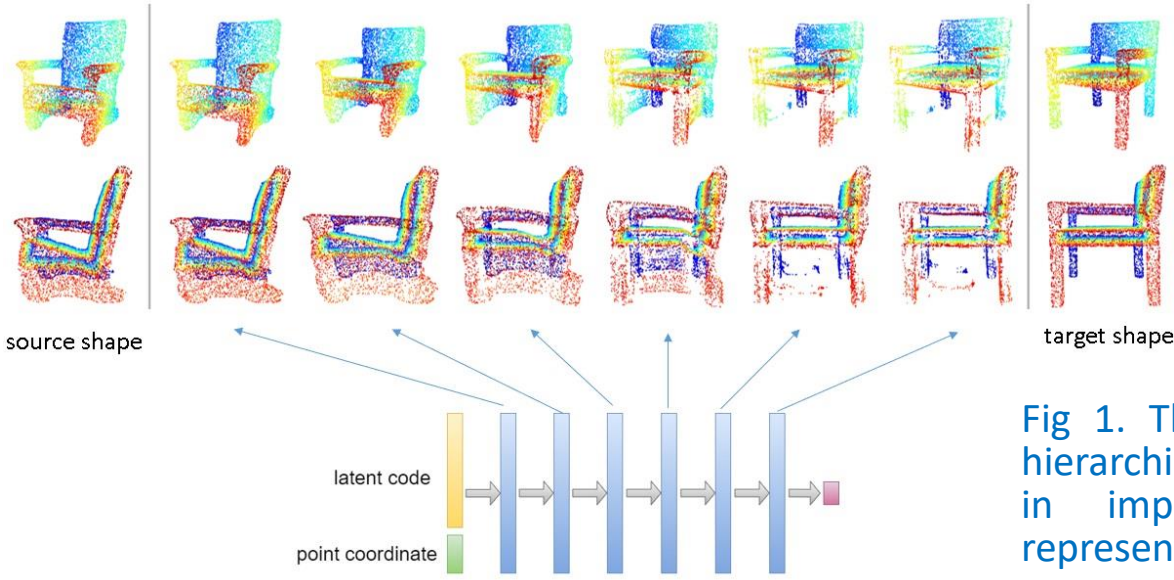


Fig 1. The emerging hierarchical structure in implicit neural representation layers

In addition to the data representation, **continuity also benefits learning models by aligning with the manifold hypothesis**, suggesting that high-dimensional data resides on lower-dimensional latent manifolds. This principle guides the success of **deep learning** by enabling smoother and more continuous latent representations. As such, Yunlu introduced the Mixup augmentation to point cloud recognition, which jointly interpolates input data and corresponding labels. The paper that appeared at **ECCV 2020** developed the **PointMixup** strategy based on the idea of the optimal transport, with mathematical justifications of the effectiveness.

The application of **continuity in 3D deep learning** has shown promising results in terms of improving the efficiency and accuracy of 3D data representation and learning algorithms. Future research can focus on addressing the limitations of oversmoothing and lack of efficiency for downstream tasks, as well as on incorporating designs that consider the continuous and discrete nature of real-world data.

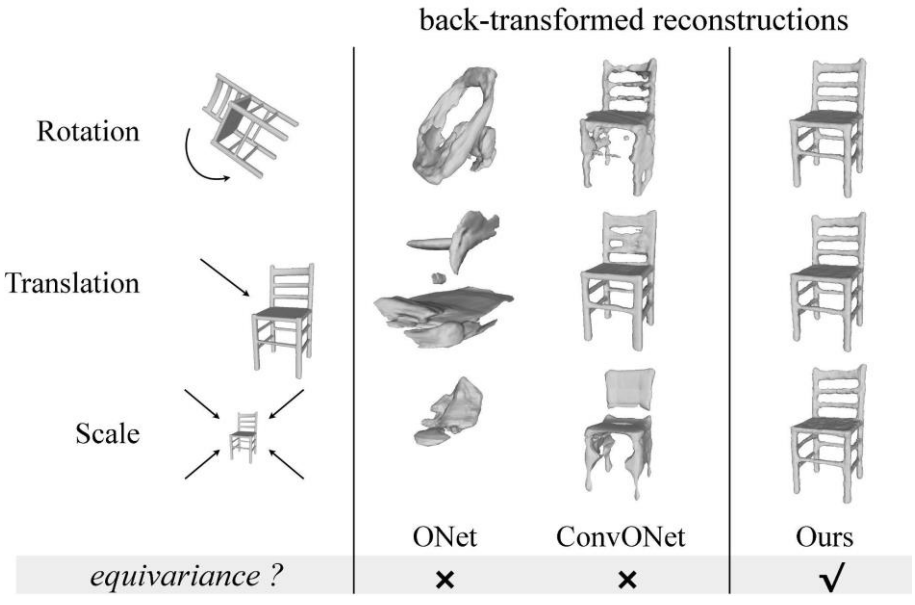


Fig 2. The equivariant implicit representation generalizes to unseen transformations

Gitta Kutyniok is a professor at LMU (Ludwig Maximilian University) Munich in Germany, where she has a Chair for Mathematical Foundations of Artificial Intelligence.

Read 100 FASCINATING interviews with Women in Computer Vision

First of all, thank you so much for interviewing me. It's a great honor for me!

Gitta, what is special about LMU that we don't know?

LMU Munich is the top university in the EU, actually, and is already 551 years old now. It has a long history, and it's a place where I very much enjoy working. Amazing research environment, particularly in AI.

We are not in competition, but I am Italian, so we also have very, very old universities!

I know, I know! Yeah, I know, I didn't want to start any competition! [*she laughs*]

You have recently given a keynote speech about the foundations of deep learning: what are you telling people about it?

[*Listen to Gitta's awesome answer in the video interview on the next page at 0'21"*]

Is this good enough to dedicate one's own professional life to it?

[*Listen to Gitta's inspiring answer in the video interview on the next page at 3'29"*]

You have a very rich experience,

Watch her interview in video!

Gitta, but you also have many, many years in front of you of teaching and researching. What would be an ideal goal to attain for you?

One direction where I'm working seriously is in making deep neural networks and AI reliable. It turned out, and this relates to what I said about limitations, that one main limitation is the hardware we train it on. We use digital hardware like CPUs and GPUs, so you can show that there are actually problems concerning the computability of these types of algorithms. You can model digital hardware as a Turing machine, and you can get rigorous results. What you also see is that, for instance, in self-driving cars, we are not as far as we thought. These are also indications of this problem.

One future direction we also work in is to augment digital hardware by analog hardware. Analog hardware like neuromorphic computing, quantum computing, biocomputing, and developing algorithms for those, and then showing that those are reliable. In that sense, that's one key goal that many people in my research group work on. If you ask me what I hope to achieve within the next 5-10 years, this is something I would like to achieve. For different application areas like medicine, robotics, and telecommunication to build these augmented hardware platforms and algorithms and show that this is reliable and satisfies the constraints given by the EU Act and by the G7. This is something which is, I'd say, a vision for the future.

What has been the biggest eureka moment in your career so far?

Wow, that's actually very, very, very difficult. For me, when I entered the area of explainability, at first, it seemed like there was no mathematics possible in some sense. It's an area which has a lot of social and psychological components. It's not clear what an explanation is. But now, also doing the CVPR paper, we have some first mathematical results. That was something which really surprised me and, in that sense, was maybe something like such a moment.

What is the latest CVPR paper?

The latest paper was about a novel explainability approach using ideas, many from mathematics like applied harmonic analysis, using wavelets and shearlets, and getting

a high-level explanation. A pixel-based explanation but in a very natural way. Combinations of pixels showing how important they are for a certain decision.

That is a very recent success. I'm very happy for you that your strongest eureka moment is something very recent! [*Gitta laughs*] You are a modern researcher in a centuries-long chain. Of the scientists that preceded you, which one you admire the most and why?

Maybe I can mention two. One is Ingrid Daubechies – she introduced wavelets. Before, I said I worked on imaging sciences and applied harmonic analysis was one of my main areas, so Fourier analysis, and then wavelets, and she developed this. Compactly supported wavelets is a beautiful mathematical theory. She also got hired at Princeton for that, and I admire her greatly.

Another one I admire a lot is David Donoho, who introduced compressed sensing. That is a technique with which you can acquire data in a very compressed form. You have also beautiful mathematical theory. I think what he usually does is he solves concrete problems, but at the same time, develops a deep mathematical theory for those like, for instance, compressed sensing. These are two people I would like to mention. I also had the honor to be a postdoc with them for some time, so this was very much a great experience.

Am I right that the most important search in your professional life is to explain things and to understand how they work?

That's very correct, yeah. I would like to understand things in their

depths. For instance, for deep neural networks, I would like to understand how they work and why they work, and in that way, make them also trustworthy and reliable.

I don't know how my washing machine works, but it works for me. It does the laundry. Why do you dedicate so much effort to something that, for whatever reason, works?

Yeah, that's a very good question, but your washing machine does not pose any threat to you, I imagine. However, if you have a self-driving car that doesn't work properly and makes wrong turns, it could cause serious harm to humans, animals, and so on. In that sense, I think it's a very different situation. I also don't understand my washing machine. I mean, to a certain extent, I think I know how it works, but I don't have any, let's say, worry about it. But AI is also used, for instance, for reaching decisions like, is this person sent back to prison or not? I mean, really key decisions. In that sense, it needs to be absolutely reliable. We need to understand how it reaches decisions.

Also, we need to understand how it relates to the training data, which is usually something not that much put into focus. The training data needs to be clean, and it needs to be fair, and so on. From the data science side, there needs to be a lot of research in this direction. In that sense, from my perspective, it's absolutely key. Also, this EU AI Act correctly requires that we understand it in depth. There's this right to explain, a right to explanation. If AI technology is approved, we need to understand how it works. The customer needs to know or has the right to get an explanation for this decision.

Gitta, you have succeeded in scaring me a little bit that AI could find things that may not be exactly in the direction we want. Has this already happened, or are we not there yet?

If you think of self-driving cars, they were already involved in accidents, so it's known that they sometimes brake strangely. I think if there is a car on the right-hand side at the

" ... they said there are these great methods now, which often easily outperform the state of the art, but we can't really understand them. They are like a miracle, and the mathematics is missing!"

curb, which has a blue light, then sometimes it reaches the wrong decision. There were already people harmed. In that sense, we already have some situations where self-driving cars or, in general, robots, because self-driving cars are nothing else than a robot, cause harm to humans.

Is it worth it to incur all these harms and, at the same time, get the benefits?

I think it's definitely worth it. There are a huge amount of benefits which we see. If we then have self-driving cars which are reliable, I think it will be a great help. Also, think of robots in general. I mean, in the whole area of logistics. During the pandemic, it could actually substitute people. Otherwise, companies like Amazon could not have delivered their packages. There, it was a great advantage. Also, if you think of the elderly at home, in the demographic change, robots could be a great help to keep them in their homes and assist them.

Think also of ChatGPT. It's discussed very diversely, but for instance, a lot of my colleagues right now use ChatGPT to write a draft of a report, and then they personalize it a bit. It helps people in various different ways. If we make it trustworthy, then it will be a tremendous help also in very sensitive areas. Think also, for instance, medicine. This is already used for detecting skin cancer. If you go to a doctor, typically, they take images of your skin and the different spots, and then it's analyzed by an AI-based algorithm, which is often much better than the human doctor. The advantage is that the human doctor looks at it and then decides based on that. It's not only the AI that decides, which I think is a great advantage, but there it's a great help to the doctor. In that sense, I think it's definitely worth it.

If I hear you correctly, we will take the risk and run on the runway and see what happens.

Yes, exactly.

Do you have any advice for younger scholars? They are starting their careers in confusing times.

Their first conferences, where they were supposed to meet their lifetime friends, were virtual because of Covid, and AI is quickly bringing dramatic changes.

Yeah, sure, I understand that it's extremely difficult. Finally, fortunately, we're back to conferences also on site. I would recommend everyone to take that advantage because nothing beats meeting people in 3D. If I would advise a young colleague, first of all, go to conferences if you have the chance. Don't take part virtually. Go to them. Meet as many people as you can. Talk to them. Network. That only works in 3D; it doesn't work if you move around on your monitor.

Then concerning AI, it now affects every area of science, basically, even the humanities. For instance, at LMU, we now have AI as a minor for all different subjects. Take advantage of courses in AI so that at least you learn the basics of it because I can imagine, in your research, you will need it in one way or the other, and so then it's good to know the basics. I mean, to start from there.

Also, it's usually a good idea at your university to get in contact with people from computer science, mathematics, and statistics, who have a bit more knowledge. Maybe team up with them if you have interesting research questions from an applied area. For me, it's very exciting to work in an interdisciplinary way. I'm always happy if people approach me with problems from other areas of science. For instance, since I came to Munich, I work now in robotics, which I never imagined before, and it's very rewarding and exciting. Then if you work in a different area and you're not familiar with AI, reach out to your colleagues and fellow PhD students, and talk to them.

Congratulations to Alaa Bessadok for receiving the award for the best women's scientific research of the year 2023, titled "Innovation and Renewal in the Fields of Technology, Digitalization, and Agriculture". The ceremony took place on August 11 in Tunis. Congratulations also to Alaa's brilliant mentors: Mohamed Ali Mahjoub and Islem Rekik. In Alaa's words: *"Wishing Tunisian women 👩‍👧‍👩 continued brilliance and leadership across all fields, year after year!"* We couldn't say it better! Other exceptional Tunisian women: Aïcha and Rawia.

ARE YOU GOING TO MISS ICCV 2023?
YOU CAN BE ALMOST THERE ANYWAY!

Here's the trick to follow ICCV day by day: click here and subscribe for free to ICCV Daily (4-5-6 October) with awesome highlights from ICCV in Paris!

Feel at ICCV,
As if you were at ICCV!

Subscribe here for free

See You - for real - in Paris!

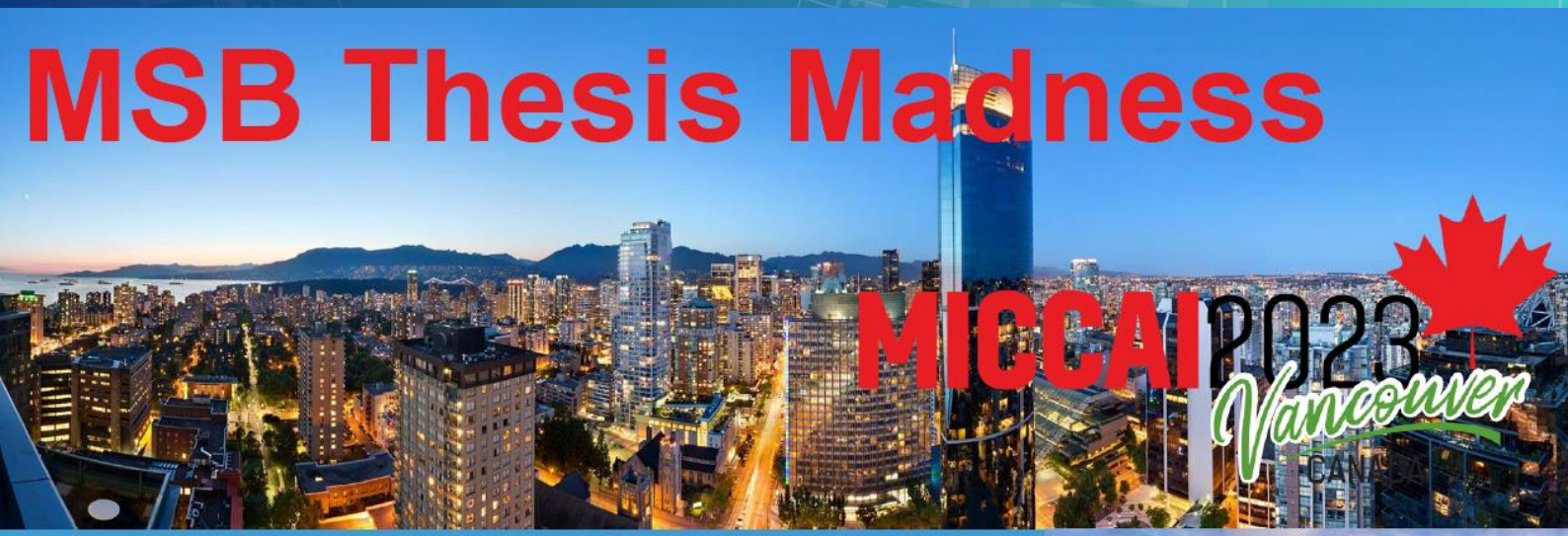
Hey, you! Yes, you! Come to ICCV 2023 in Paris! Please! It will be glorious! For real!

Awesome Yael Vinker (left) is a PhD student at the Tel Aviv University, and also a celebrity at SIGGRAPH, being the winner of the Best Paper award last year, here it is if you forgot! This year she let someone else win the prize and instead she sent us these stunning photos from SIGGRAPH 2023 in Los Angeles. The lucky gentleman smiling at her right is Moab Arar, a PhD candidate, also at TAU.

SIGGRAPH
ELECTRONIC THEATER
BEST STUDENT PROJECT

Swing to the Moon

MSB Thesis Madness



Final-year Ph.D students present your thesis in 3 minutes

Organized by:
MICCAI STUDENT BOARD

Send application to miccai.sb@gmail.com

Deadline: September 11th

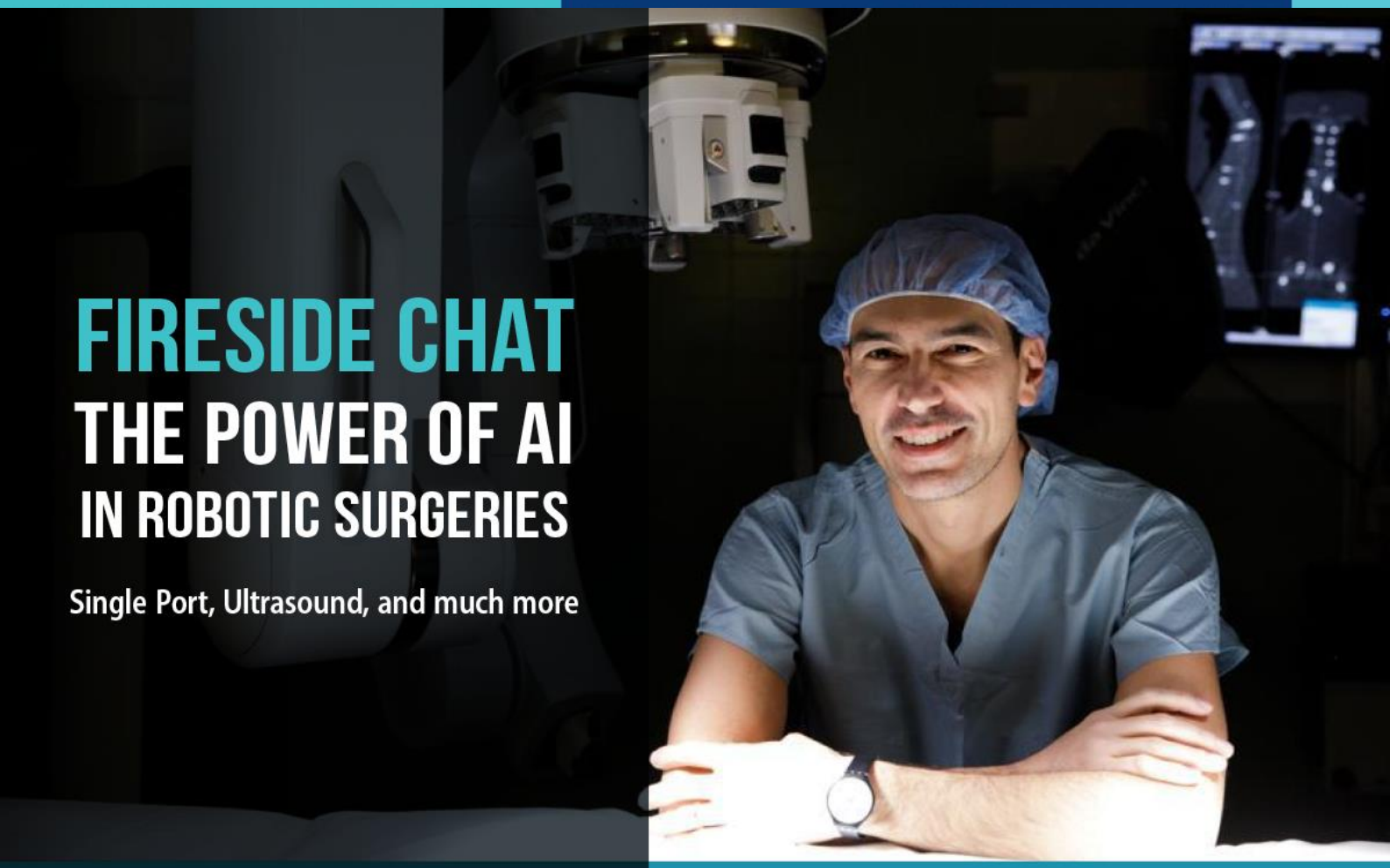
Would you like to present your thesis to the MICCAI community?

To participate, you must:

- **Be a final-year Ph.D. student or recent graduate - no more than 6 months after finishing your PhD.**
- **Attend the conference in person.**
- **Prepare a 3-minute presentation summarizing your thesis for a broad audience.**

MEDICAL IMAGING NEWS

SEPTEMBER 2023



FIRESIDE CHAT

THE POWER OF AI IN ROBOTIC SURGERIES

Single Port, Ultrasound, and much more

Tuesday, September 26th, 2023
10 am PT / 1 pm ET

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

CO-HOST
Shmulik Shpiro
EVP Global BizDev
RSIP Vision

CO-HOST
Moshe Safran
CEO
RSIP Vision USA

RSIP VISION Visual Intelligence for MedTech

UI Health Urology | UIC

Generalized 3D Medical Image Registration with Learned Shape Encodings

BEST PAPER MIUA 2023

Christoph Großbröhmer

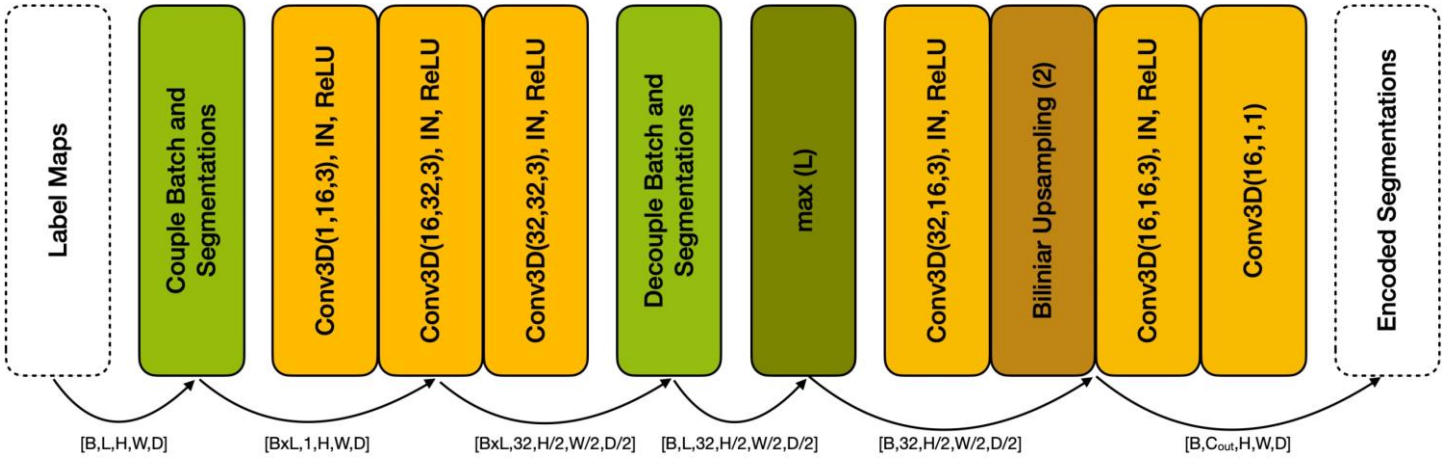
Mattias Heinrich

Christoph Großbröhmer is a PhD student at the University of Lübeck under the supervision of Associate Professor Mattias Heinrich. Fresh from winning the Best Paper prize at the 27th UK Conference on Medical Image Understanding and Analysis (MIUA), Christoph joins us with Mattias to discuss his award-winning work.

This paper proposes a novel approach to generalized image registration that leverages learned shape encodings. Traditional image registration methods often demand tailored solutions for specific tasks. However, this innovative approach uses medical image segmentations to perform registration across different domains and datasets, promising a solution capable of addressing diverse challenges.

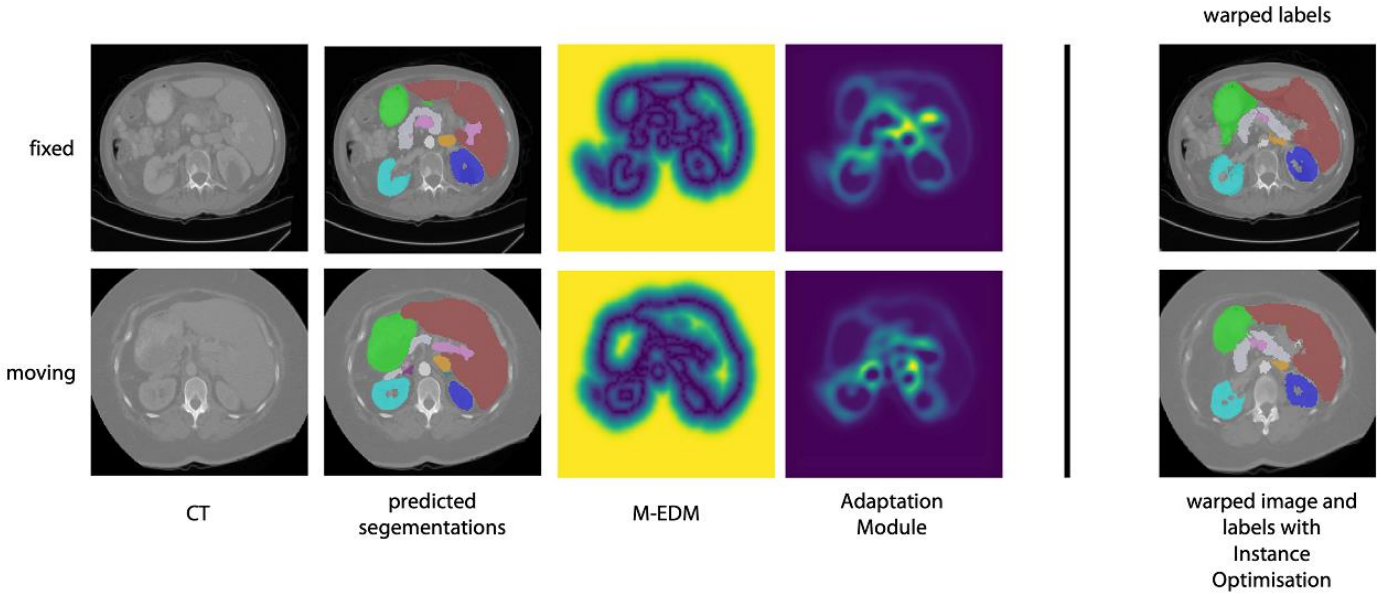
The researchers needed to find a common representation between different datasets. They grappled with the task of training a model on one instance of data and then inferring it on something completely different.

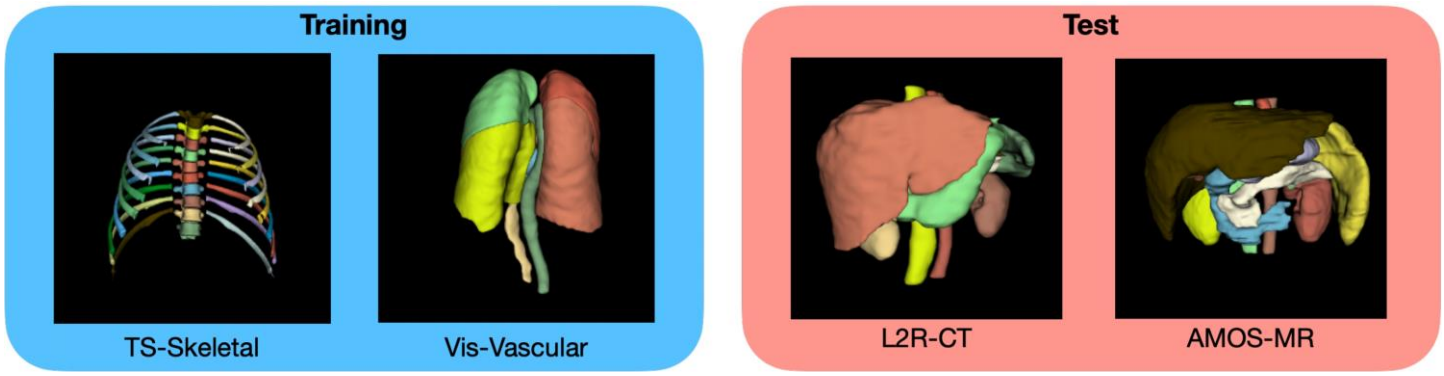
"The idea to leverage other advances in medical image analysis, like segmentations, lets us use a bridge between these different problems," Christoph explains. *"We can use advantages from one field to solve or find some kind of solution in another field. That's pretty nice, in my opinion."*



Previous research has shown that **semantic segmentations** can be inferred across different medical data types. The team faced the challenge of using this prior work for another task. Finding a generalized solution for different problems was difficult for the registration part. However, by proxying the problem to another field in the segmentation part, they could formulate a generalized approach to image registration without the limitations of traditional methods.

*"Our motivation stems from our experience evaluating and running the **Learn2Reg** challenge that we organized at MICCAI,"* Mattias reveals. *"We saw that we have this great variety of tasks. We have abdominal image registration, lung, or thorax. We have different modalities: MRI, CT, ultrasound. But so far, all the participants had unique and different solutions for each of them. This requires a lot of training, retraining, designing, and redesigning different methods for each challenge task. We thought having one joint approach would be a good step forward."*



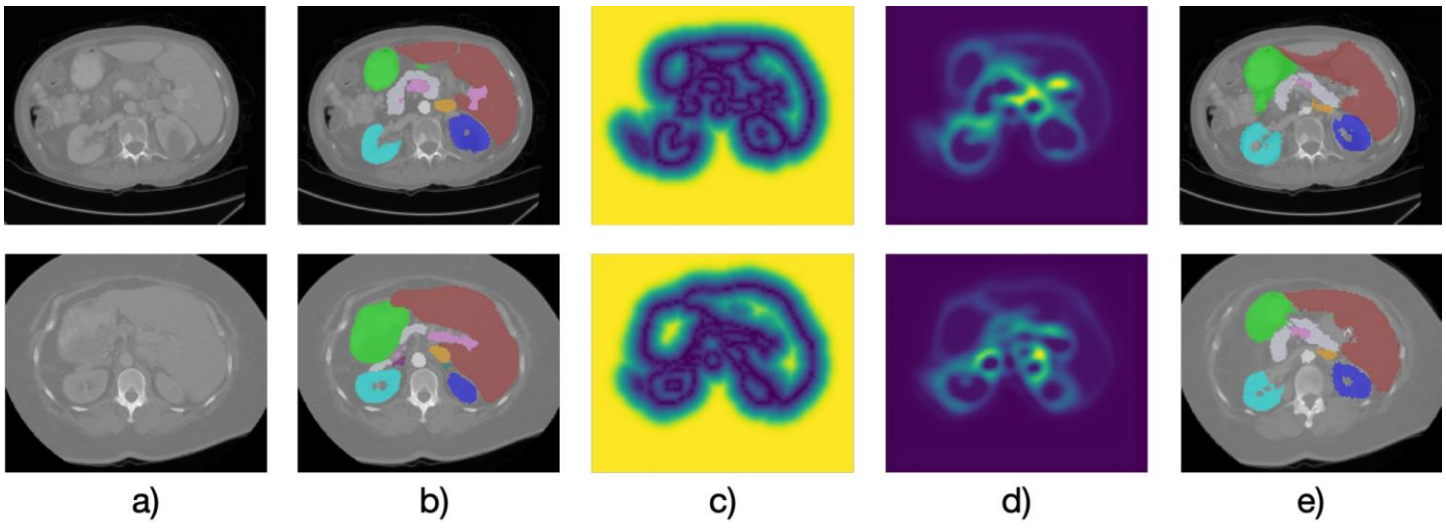


"We can use advantages from one field to solve or find some kind of solution in another field!"

Now, with **an approach that works across different modalities and anatomies**, the team modularized it so that representation learning with segmentation can be combined with any registration network. Testing it on two networks, a very basic **CNN and a transformer architecture**, showed that it generalized better in both cases.

The paper's success marks a significant step forward, but the researchers acknowledge that more work remains to be done to realize the approach in a real-world medical setting. They point to the need for further research into generalized solutions, such as **nnU-Net** and **TotalSegmentator** for segmentation.

MIUA is a modest but popular event on the conference calendar. Christoph and Mattias are keen to thank the organizers and tell us how much they enjoyed their time in Aberdeen and hope to return for future editions. Having scooped such a prestigious award at the event, has Christoph considered what made his paper a winner?

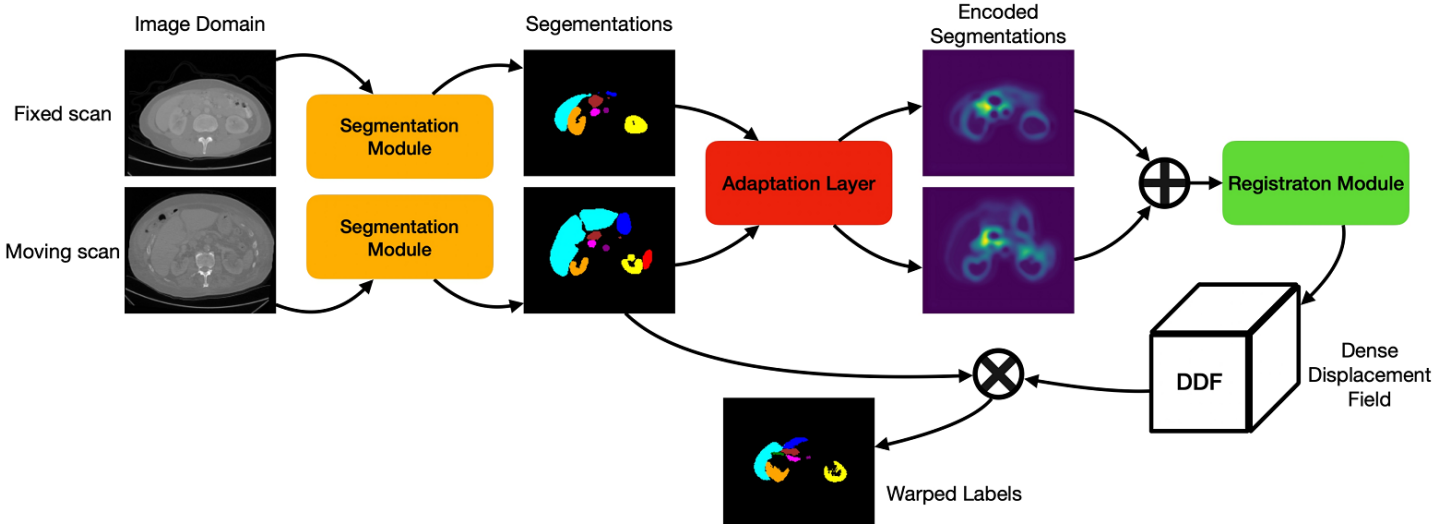


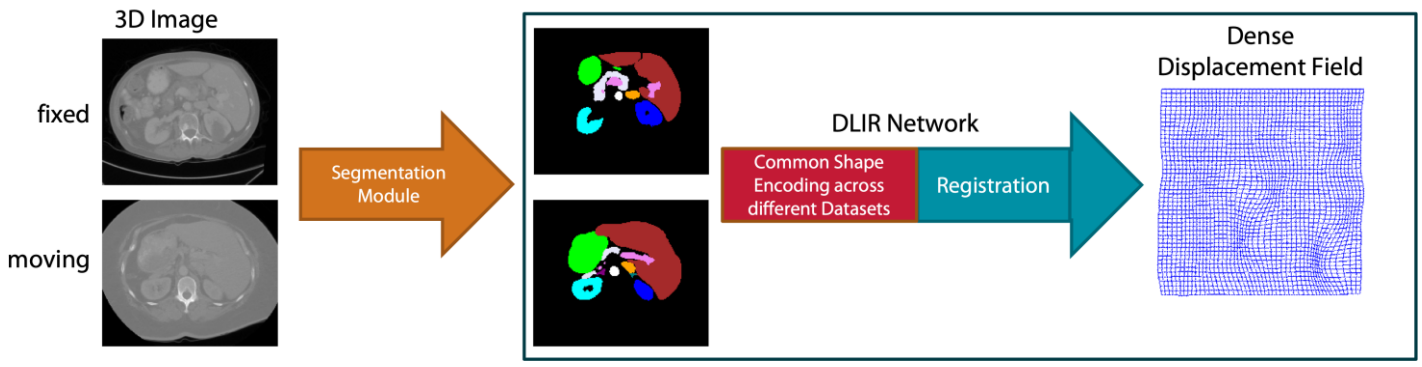
"There were a lot of great papers at MIUA, so I have asked myself the same question, of course," he ponders. *"While I'm very happy that I achieved the* ***Best Paper award****, there were many other papers I really liked. I think one of the strengths of my paper is that* ***the solution has a very modular approach****. It works jointly with prior work in other fields. It's modular for different registration frameworks. The evaluation is quite sound. Also, maybe it's a topic that everybody can visualize for themselves. At least in our community, everybody knows what image registration is and that the main idea of shape matching could be one of the basic approaches for generalized registration. I think that's a very nice but comprehensible idea as well."*

Mattias adds: *"Christoph also made a good effort in making the paper reproducible and open.* ***All the datasets are publicly available!*** *His code is on GitHub and can be reused by other researchers."*

Can Mattias give us any final insights into the man behind the work?

"Christoph is really good at interacting with the scientific community," he tells us. *"He's done a lot of great work in supporting the Learn2Reg challenge and has interacted with many other researchers in the field of medical image registration. Even though he's not necessarily met them all in person, he's provided a lot of support. I think this is a great strength!"*



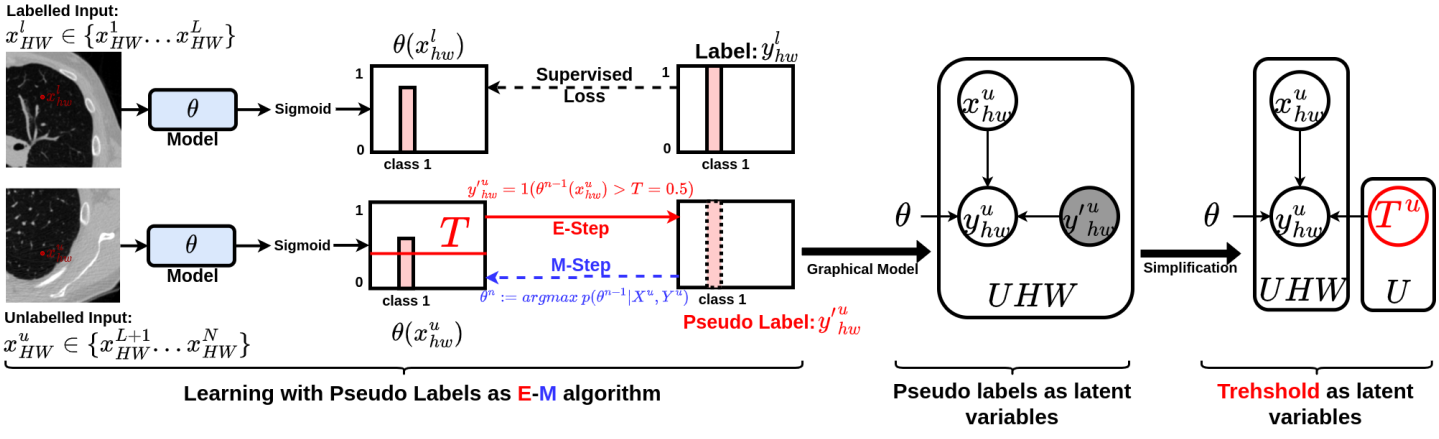


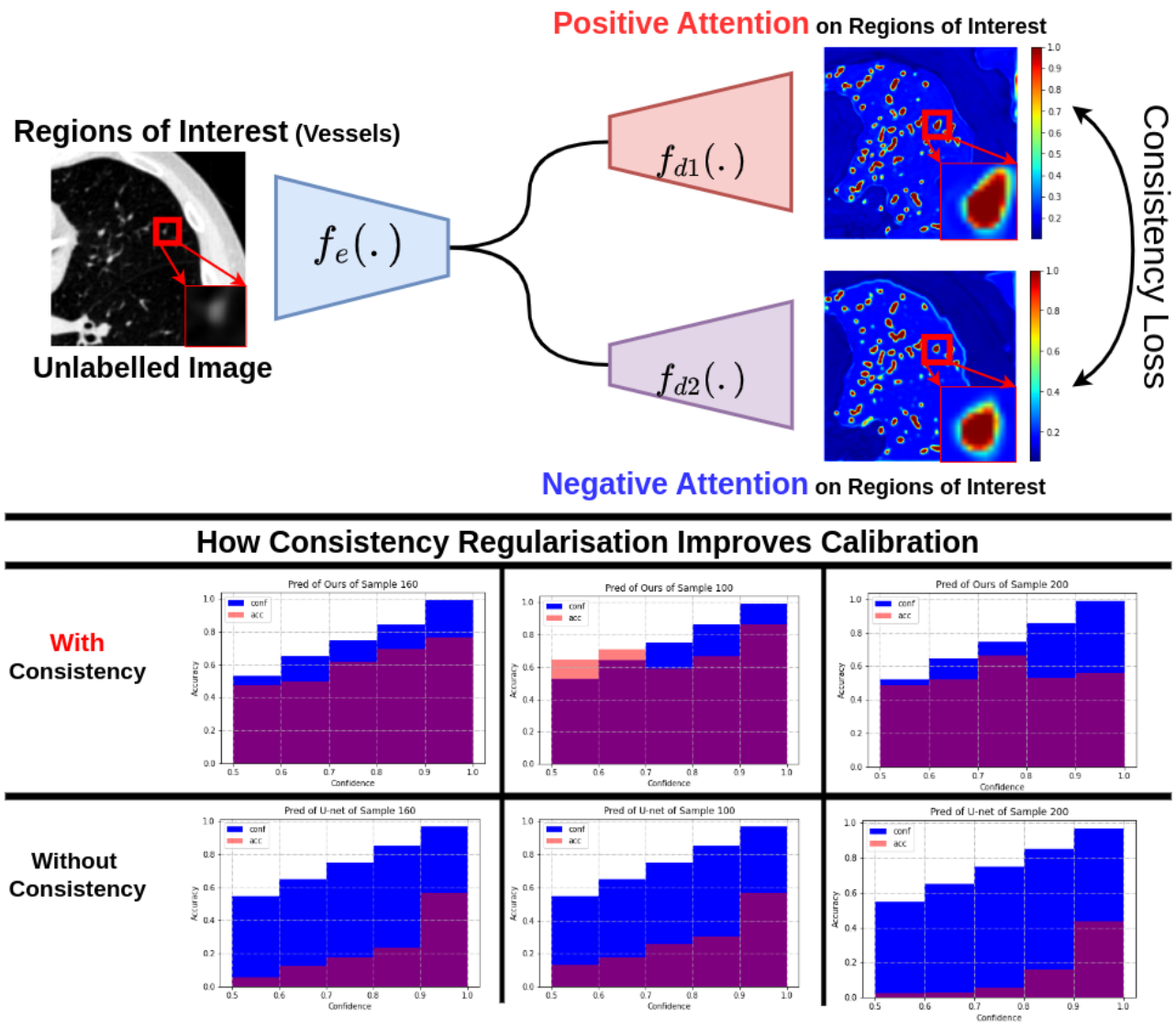
Moucheng Xu has recently completed his PhD from the Centre for Medical Image Computing at the University College London. His research focuses on deep learning with limited supervisions in medical imaging. He has now moved to industry in MedTech to continue to pursue his passion for creating intelligent medical imaging systems.
Congrats, Doctor Moucheng!

Deep learning has now become the de facto data-driven approach for medical image analysis in the digital era. A well-executed deep learning model requires a massive amount of data and their corresponding labels. However, acquiring labels in medicine is extremely challenging due to the high costs in both time and money.

Moucheng's thesis seeks the answer to the following question: How can we train a deep learning model without sufficient labels for medical image analysis?

In the first work of his thesis, he presents a new perspective on the visual attention mechanism as learnable morphological operations at the feature level. It also introduces a new consistency regularization on dilated feature perturbations and eroded feature perturbations for semi-supervised segmentation of medical images. Furthermore, his work reveals an empirical link between the consistency regularization and the model calibration for the first time. This work also led to publications at MIDL 2022, IEEE TMI, as well as a patent filing.





His second work of the thesis defines a new formulation of pseudo labelling as the expectation maximization algorithm, providing an interpretable perspective of the empirical effectiveness of the pseudo labelling in semi-supervised learning (Fig. 2). Based on this newly proposed formulation, his work extends the original pseudo labelling towards its generalization and presents an approximation of its generalization using a variational inference. This work resulted in a publication at MICCAI 2022, which was fortunately shortlisted for the Young Scientist Award, and another patent application.

The final work of his thesis leverages the latest advancements in variational unsupervised clustering techniques with discrete priors and demonstrates its first application on model parameter estimation of MRI signals. This pioneering approach challenges the traditional assumption that all of the voxels are treated independently in MRI parameter estimations. Additionally, the new approach enables a new generation of dMRI and qMRI with enhanced anatomical structures while reducing noises. For more details, reach out to him.

Diffusion Models for Sample Synthesis

by Christina Bornberg
@datascEYEnce

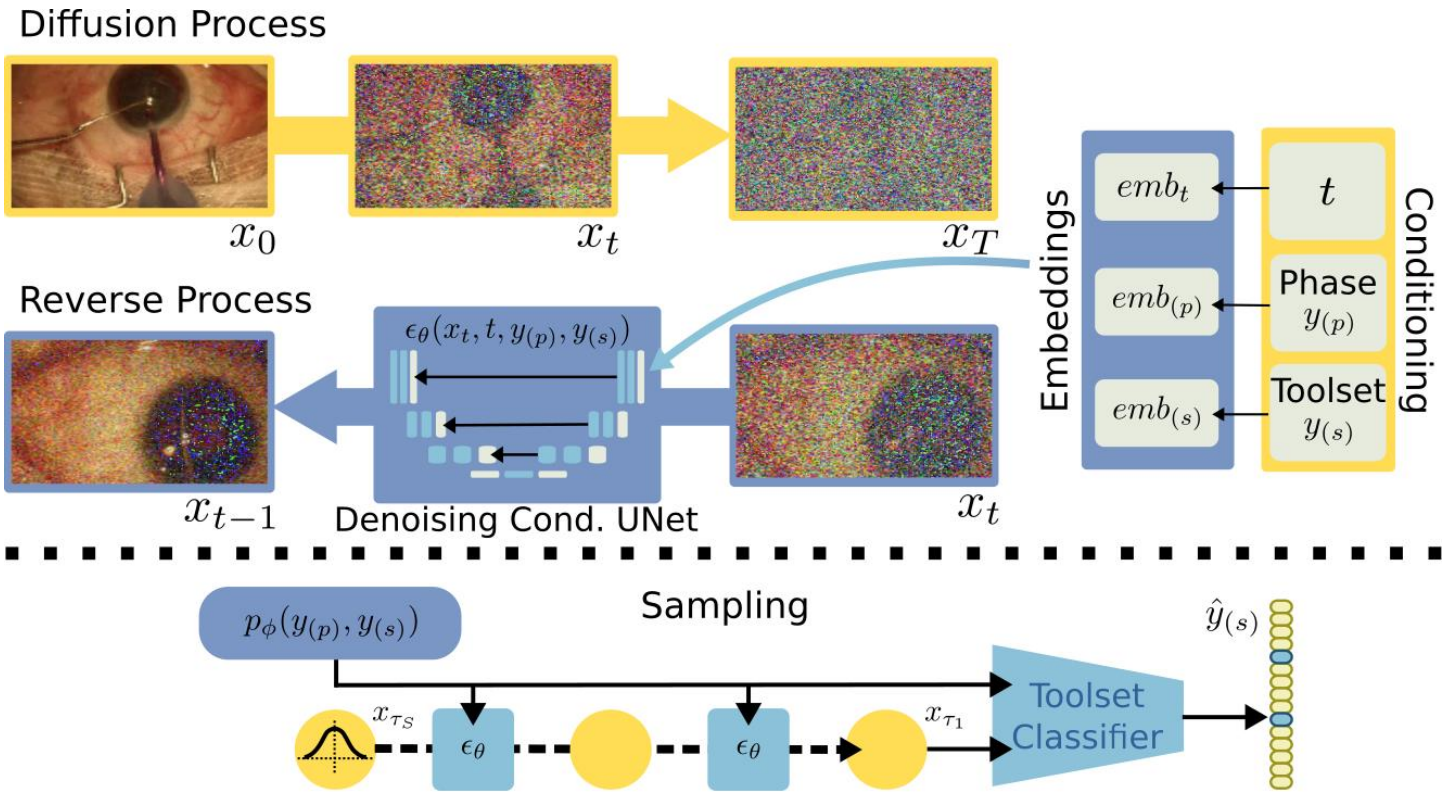
Welcome to the datascEYEnce column! I am Christina and I currently do research in deep learning for ophthalmology at the Singapore Eye Research Institute. I enjoy doing STEM outreach and thanks to Ralph, I can do this now here in Computer Vision News!

featuring Yannik Frisch

In this edition, I would like to introduce you to the work of Yannik and his colleagues at TU Darmstadt! Yannik is a PhD student at the Interactive Graphics Systems Group focusing on generative models and representation learning applied to surgical data - specifically to the rather neglected field of cataract surgery.

During our conversation, Yannik introduced me to his upcoming publication at MICCAI 2023, titled "Synthesising Rare Cataract Surgery Samples with Guided Diffusion Models". This work presents promising results in enhancing toolset prediction by synthesizing rare phases as well as tool combinations using the cataract surgery video dataset "CATARACTS".

So, how can image synthesis contribute to a classification task? There are actually two motives. The first one is rather intuitive: reduce the class imbalance by sample generation. Naturally, some surgical phases are shorter than others which commonly results in the network favoring the majority classes. The second objective in their ongoing work is quite interesting: they want to synthesize frames showing human error. While they clearly cannot tell a surgeon to make a mistake on purpose for the sake of a dataset, generating frames for wrong surgery steps seems like a great alternative. A possible application here is training novice surgeons.



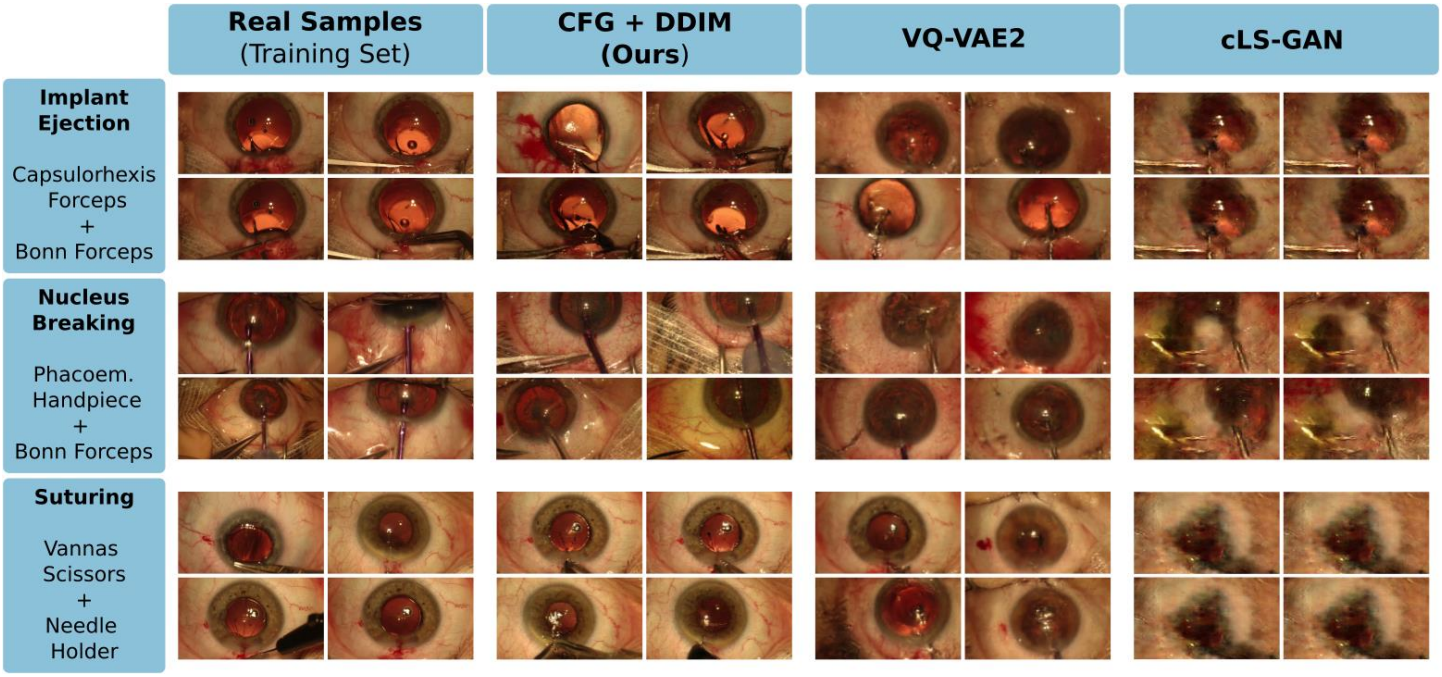
Now, let's delve into the implementation of the pipeline for frame generation, conditioned on surgical phase and tools! Yannik applied two main concepts: Denoising Diffusion Implicit Models (DDIMs) and Classifier-Free Guidance (CFG).

DDIMs are an adaptation of Denoising Diffusion Probabilistic Models (DDPMs) which follow the underlying idea of gradually adding noise to an image as part of the **forward diffusion process**, and later reverting the process with a learned model as part of the **reverse process**.

In more detail, first, Gaussian noise with a pre-defined variance schedule is iteratively added to an image, resulting in image xT.

In the next step, the goal is to reverse-predict this process iteratively with a denoising UNet. This means the noise at each timestep t will be predicted by the model. After removing the predicted noise from an input image at a certain timestamp, we get xt-1, a slightly less noisy image than the previous one. Finally, after iteratively applying the denoising process, we receive a new synthetic frame. In this reverse process, DDIMs have an advantage over DDPMs: as their sampling strategy is more efficient, the number of inference steps needed can be reduced by roughly 80%.

CFG is a technique whereby an additional model is used to influence the generation process. We can now



use the denoising UNet in both a conditionate state or a vanilla state. In the conditionate state, according to the CFG approach, phase and toolset embeddings from a separate linear model are used as additional input to the network and hence make it possible to generate an image in specific (rare) phases or tool combinations.

How can we now qualitatively assess the synthesized image?

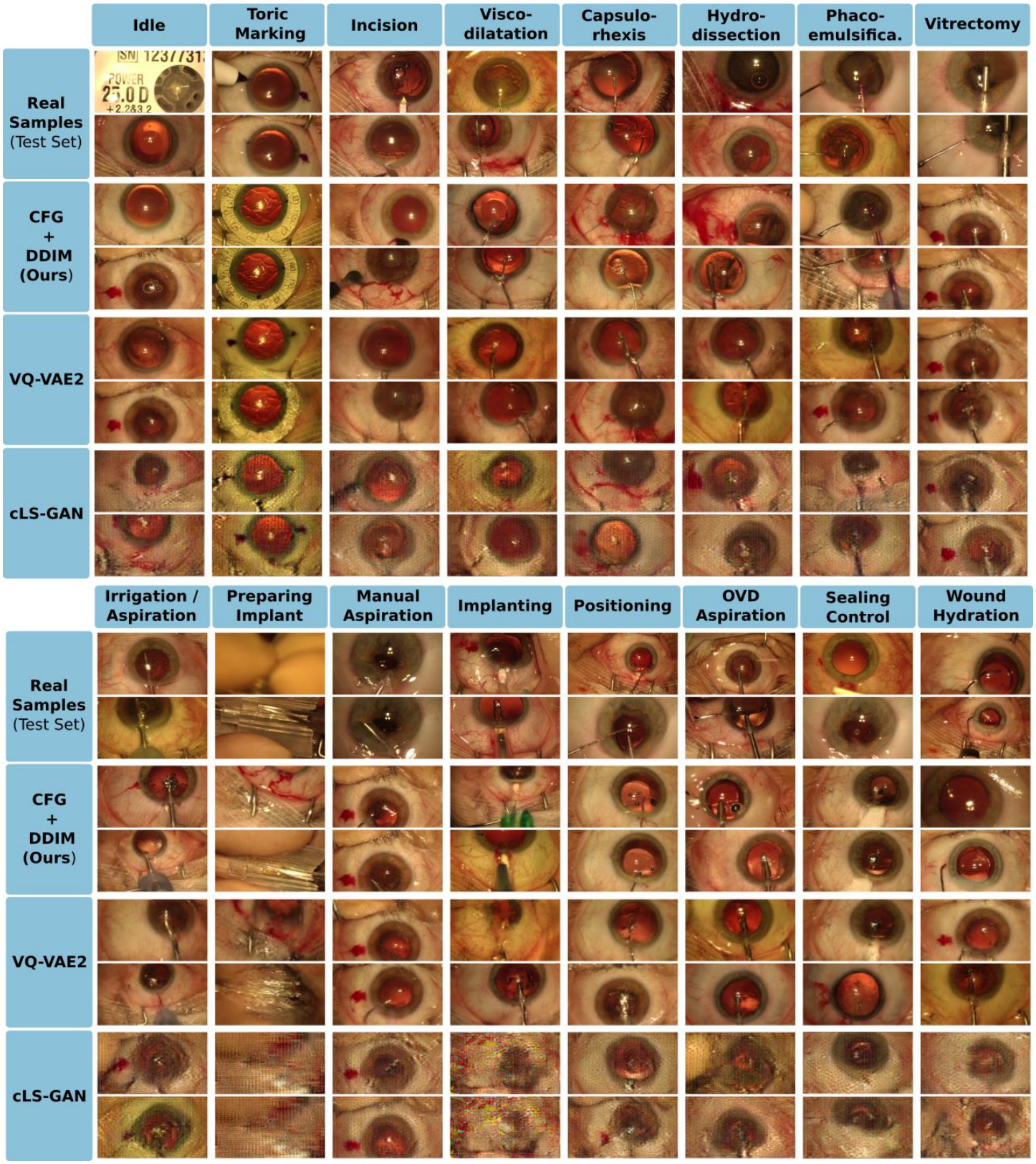
Firstly, he used metrics such as Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) to compare the distribution between generated and real images which is a common approach in testing generated image results.

In another experiment, Yannik utilized the Inception Score (IS) and F1 score to evaluate the conditionally generated images with a pre-trained model designed for multi-label, multi-class tool classification.

The last experiment we will cover here, and probably the one with the most interesting outcome is the Downstream Tool Classification. In an ablation study, a classifier trained on only real images, only generated images, and a combined version made clear that critical phases have an increase of up to 10% in F1 score when training on both real and and synthetic data combined!

In closing, Yannik shared some insights into the training procedure of the code available on GitHub: When training the denoising U-Net, you will see a strong decrease in the loss right at the beginning and close to no decrease for further training. While performing their experiments it became clear that the FID and KID scores do benefit from longer training.

If you want to learn more about the project, I would recommend reading the paper here and **keeping an eye out** for Yannik's work at MICCAI 2023, next month in Vancouver!



Cancer Prevention through early detecTion (CaPTion) @ MICCAI2023 Workshop

Bartek Papiez is an Associate Professor at the Big Data Institute in Oxford, leading the Medical Image Analysis and Machine Learning group. He is co-organizing an innovative MICCAI workshop on early cancer detection and speaks to us ahead of next month's main event in Vancouver.

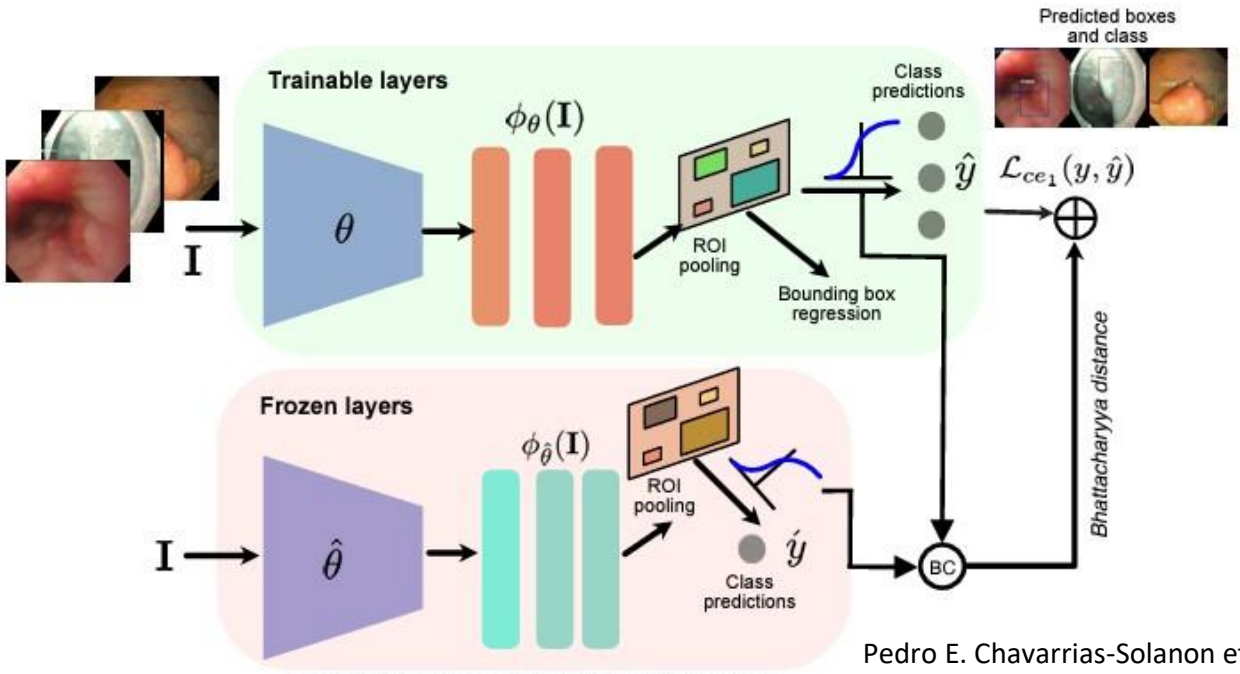
In recent years, much discussion and research funding has been invested in **early cancer detection**. The **Cancer Prevention through early detecTion (CaPTion) workshop** aims to bring this topic to the forefront of the **MICCAI** community.

Before the workshop's first edition last year, organizers noted a shift in what they perceived as a very technical conference, with MICCAI seeing a growing focus on interaction with clinicians and even introducing a clinical day.

"*Early detection is not only detection; there is a strong imaging component, and this fits very well with MICCAI,*" Bartek explains. "*There is lots of **preclinical research** which uses microscopy and all of the variants of **preclinical imaging**. They process and analyze images, not necessarily to detect cancer, but to understand the mechanism of how cancer develops and progresses. We wanted to bring this to the MICCAI community and focus people around the application rather than technology.*"

The CaPTion workshop is critical to this mission. Revolving entirely around early cancer detection, it covers various facets of the application, from biology and screening to integration. It brings together **clinicians, researchers, and**

Ziang Xu et al.

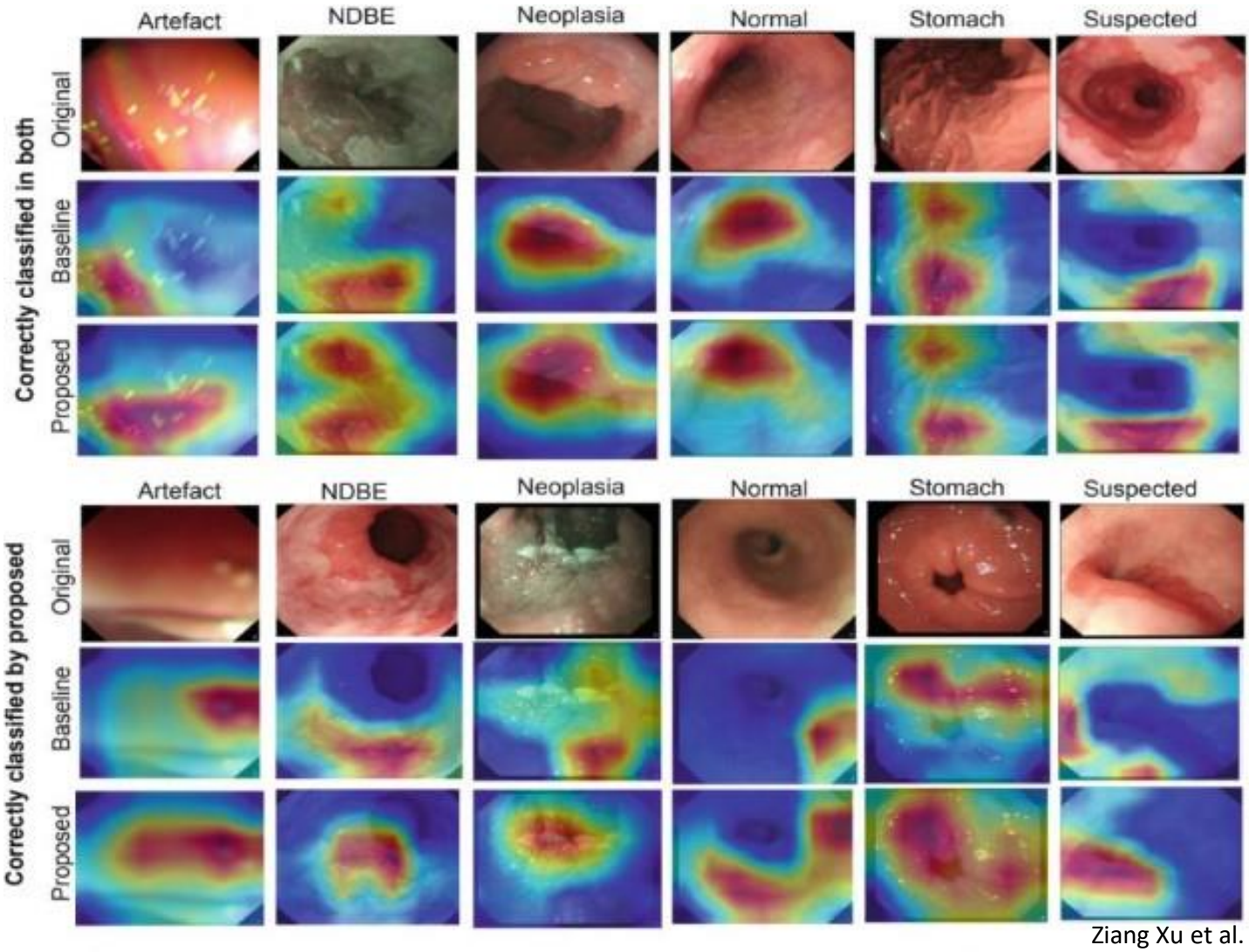


Pedro E. Chavarrias-Solanon et al.

industry players, emphasizing the potential for translating groundbreaking research into real-world solutions.

"We wanted kind of a pitch to all of the MICCAI people – maybe you already have interesting technology which can be applied to early cancer detection," Bartek says. *"Just come, show us, and show the other people who might be interested. This may be a small change, but the new technology can make very rapid developments in* ***actual detection****!"*

A crucial aspect of this workshop is the synergy between the right people and the right datasets. **Creating data pools or lakes** has historically been pivotal in research and development. These resources



Ziang Xu et al.

are precious for smaller labs that might lack the capacity to gather extensive datasets independently. By fostering interactions and networking through these datasets, the ambition is to expedite progress in the field. Ultimately, this collaborative push holds the promise of translating advancements into tangible benefits for patient healthcare.

With CaPTion scheduled for the last day of MICCAI, organizers are keen to encourage conference attendees to put off their sightseeing for one more afternoon and come to what promises to be a fascinating event. But what makes CaPTion stand out from the other workshops, all vying for attention simultaneously?

"*There are multiple reasons,*" Bartek reveals. "*We've already completed the paper selection process, and I'm very excited that we've selected a number of very, very good quality papers, which I believe will be of interest to everyone. Come and hear about the new methods developed, the new datasets collected, and the new clinical evaluations conducted.*"

As well as being a gateway to groundbreaking research, the workshop promises a unique networking opportunity and the chance to continue building a vibrant community. It has a multi-directional approach, and organizers

aim to bridge the gap between medical imaging and clinical science with a diverse lineup of distinguished keynote speakers at the forefront of their fields. **Sravanthi Parasa**, **Anne Martel**, and **Sir Michael Brady** will deliver enlightening talks covering technology and methodology core to MICCAI's mission, including a personal journey of translating technology from the university lab to **real-world patient care via an innovative spinout**.

"It's possible, it's doable, and it can make a real impact to the patients!"

"I think that's going to inspire people to pursue such research – pursue more like a commercial side of the research, basically," Bartek tells us. *"It's to emphasize that* ***it's possible, it's doable, and it can make a real impact to the patients!***"

Bartek's group at the Big Data Institute has around 12 people working on different biomedical imaging problems. It works with clinical imaging and large population studies and supports preclinical scientists and other research groups with image analysis.

For Bartek, the ideal impact of CaPTion this year and in future years transcends academic achievement. Instead, it is the realization of technologies and methodologies presented at the workshop being picked up, translated to the clinic, and **going on to save people's lives**.

"It may sound a little weird when a researcher says that the main impact is not only the paper, but the ***long-term patient benefit****,"* Bartek points out. *"A paper describing a certain mechanism for how cancer progresses would also be an excellent outcome because we can learn from it and use it to develop a better screening methodology. The big dream is that one of the papers is a technology that will improve the detection or understanding of cancer. It'll make a real impact on those who are the most important here: the patients."*

In the short term, CaPTion is about uniting a community with a common goal: revolutionizing early cancer detection. The workshop aims to attract fresh minds, encourage collaboration over competition, and lay the groundwork for an enduring movement. The hope is that this collective spirit will spur more research, new technologies, and ultimately a more significant impact on the lives of those affected by cancer.

"It's the same for MICCAI," Bartek adds. *"We hope there will be more and more papers submitted to the main conference on early cancer detection because if people become more interested, more work will be done. It's like building a backbone, building infrastructure, building a movement so that people want to work and invest time in that area."*

The workshop will be giving awards for **Best Paper and Best Poster**. Once again, it has funding from **Satisfai Health**, whose CEO and founder was a keynote speaker last year. Satisfai will also participate in the workshop, demonstrating a solid orientation toward translation.

"I'd like to invite everyone to participate in the workshop," Bartek declares. *"I would love to see as many people as possible, despite the fact we're on the very last day of MICCAI. Stay one more day in Canada and visit beautiful Vancouver, but before you leave the main MICCAI conference, please come and join our workshop and hear the exciting keynote speakers. We hope it will be another successful event for the CaPTion community, the MICCAI community, and the patients!"*

The Cancer Prevention through early detecTion (CaPTion) workshop will take place on October 12 at MICCAI 2023 in Vancouver, Canada.

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

RSIP VISION Visual Intelligence for MedTech

Honoring Anne Carpenter

Awesome Anne Carpenter, a professor at the Broad Institute of MIT and Harvard, and a pioneer in phenomics, was one of three influential figures honored for setting off Recursion's journey - Recursion is a startup decoding biology to radically improve lives. Anne's lab co-invented with Stuart Schreiber's team an image-based profiling assay called Cell Painting, which uses cells' morphology features extracted from images as a readout of the impact of a disease, drug, or genetic anomaly. More about Anne. Kudos!

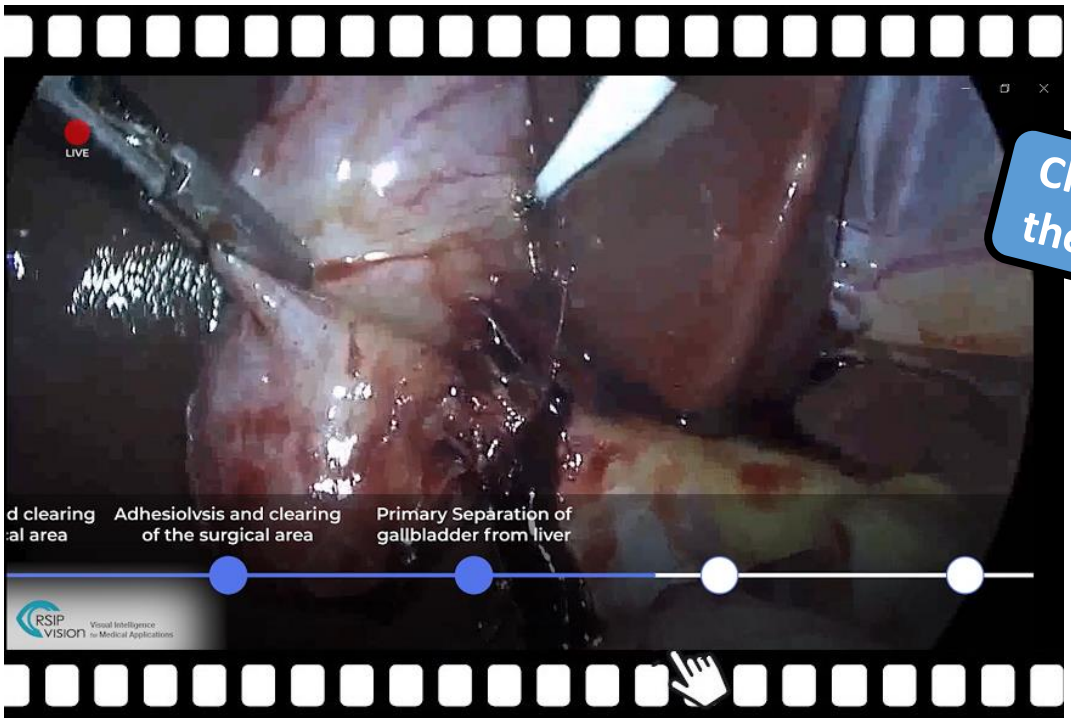
Imagine you're a surgeon in the operating room, with high-tech equipment all around you, while cutting-edge artificial intelligence (AI) works behind the scenes to analyze every movement, every decision. This is the realm of Surgical Video Analysis, a burgeoning field that's taking the healthcare world by storm. It's a place where computer science meets medicine, where AI algorithms sift through surgical video footage, both during and after operations, to uncover priceless insights.

When we sat down with **Asher Patinkin**, a knowledgeable expert in this field at **RSIP Vision**, he was brimming with enthusiasm about the potential of AI in this domain. "*Think about it,*" he said, "*imagine if we could routinely* ***improve surgical outcomes, provide invaluable training opportunities, and even drive down healthcare costs****. Wouldn't that be revolutionary?*"

He wasn't just speaking in hypotheticals, either. With **AI-driven surgical video analysis**, real-time feedback becomes a reality. Surgeons are guided in the midst of procedures, refining their techniques and securing better outcomes for patients. "*It's like having an extra set of eyes, always learning, always improving,*" Asher explained.

But the benefits don't stop at the operating table. This AI has a second life as a tireless, ever-watchful teacher, reviewing surgical footage to help **train new and experienced surgeons alike**. By learning from past procedures, surgical teams can uncover their mistakes, discover new methods, and refine their skills.

Furthermore, the analysis of surgical videos allows trainees and novice endoscopists to learn from experienced practitioners. For instance, in a procedure that requires advanced endoscopic skills such as ERCP (Endoscopic Retrograde Cholangiopancreatography), they can study the steps involved in accessing the bile duct, navigating through the duodenal papilla, and performing interventions such as



stone extraction or stent placement. The phase detection can help the trainee to easily locate and focus on the most complicated task – the cannulation – instead of browsing over hours of footage in search for this phase.

Perhaps surprisingly, the savings extend beyond the operating room, too. "*By enhancing surgical techniques and reducing complications, surgical video analytics can drive down the associated costs,*" Asher told us. The more efficient a surgeon becomes, **the shorter the surgeries and the less they cost**.

Let's use the same ERCP example: besides achieving successful cannulation and helping navigate through challenging anatomical structures, surgical video analysis provides a precious help in recognizing and preventing complications associated with the procedure - such as pancreatitis, bleeding, perforation, and infection. Surgical video analysis allows for a retrospective review of ERCP procedures, enabling endoscopists to identify factors that may contribute to complications and learn ways to avoid them.

"*So yes, it's a lot,*" Asher admitted with a smile, "*but at the end of the day, it's about more than just technology. It's about* ***better surgeries, healthier patients, and a brighter future for healthcare. That's the true promise of AI in surgical video analysis!***"

The research prospects in this rapidly expanding field are packed with potential, covering everything from shot boundary detection to video summarization. It's an exciting set of opportunities, and **RSIP Vision's AI experts and engineers stand at the forefront**. With their extensive knowledge and experience, they can produce **the right algorithms to unlock the full potential of AI in surgical video analysis**. **Contact us and we'll talk about it!**

Surgical Navigation Solutions Using an Innovative Tracking Tool

Sarah Benzouai (left) is the Marketing and Clinical Affairs Director, and Romain Fissette (below) is the Director of Technology and Innovation at Pixee Medical.

They speak to us about the startup's vision for computer-assisted orthopedic surgery using augmented reality (AR).

For decades, surgeons have relied on **navigation systems** to assist them during complex orthopedic procedures. However, these systems were often bulky, expensive, and cumbersome, occupying a substantial operating room (OR) footprint.

Now, the status quo is being challenged by the dynamic team at **Pixee Medical**, who are making real-time surgical navigation systems more affordable, straightforward, futuristic, and less invasive with **augmented reality (AR)** technology. **Knee+ is its AR navigation tool for total knee arthroplasty**.

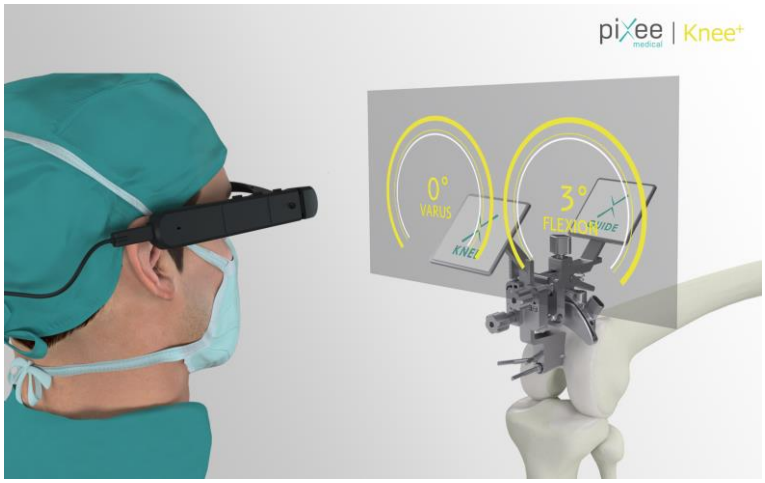
"The product itself is AR glasses plus instruments," Sarah tells us. *"That's all. That's the solution. It allows the surgeons to be assisted while wearing the glasses for surgery. With compact instruments and a small tray, they're assisted to put the*

prosthesis. It's simple because ***you're guided while using the glasses****. It's like navigation. You have to follow steps and protocols. It makes the surgeon's life easier."*

Romain adds: "*We made an effort to have everything reusable, which is not a standard in the industry. To achieve that, our technology uses* ***visible light instrument localization****. Almost every competitor product uses infrared-based technology, which so far can't be embedded in a tiny, lightweight head-worn device."*

Although other companies work in this area, the competitive edge of Pixee's solution is multi-faceted. One significant factor is in its **universal compatibility**, supporting any prosthesis brand on the market today. Also, the team was the first to believe that achieving **submillimetric precision** on off-the-shelf devices was possible.

"*Many were doubtful about that six years ago, but we proved it was possible,*" Romain points out. "*Our closest competitors still had to develop their own AR platform to achieve that. What makes us special is that we achieve this level of performance without developing our own hardware platform!*"



The success of the knee arthroplasty procedures isn't solely reliant on the technology but also the surgeon's expertise and experience. Pixee's innovation complements this by empowering surgeons to implement their skills more precisely, but does it improve the success rate?

"*Something funny we say is that our system allows a bad surgeon to be precisely bad!*" Romain laughs. "*That's not a correct answer to your question, but success rate depends on the surgeon's knowledge, skills, and experience. Our system allows them to achieve what they've been planning.*"

Pixee's localization algorithm is based on the **ArUco fiducial marker library**, and the marker pattern is designed to ensure submillimetric precision at a range of 40 centimeters, with a 4K RGB camera embedded in off-the-shelf smart glasses. Whilst the specifics of the algorithm remain under wraps, Romain shares that the pattern contains more anchor points than traditional markers. To his knowledge, this is the first time anyone has performed these kinds of in-house improvements on the ArUco library.

"*At some point, we hit a ceiling in terms of the precision we could achieve,*" he recalls. "*It pushed us to improve the ArUco concepts. We're based on ArUco technology but slightly modified the marker pattern to get more anchor points, allowing us to reach the precision we need for knee surgery and other joint replacements.*"

One of the biggest technological challenges was **defining a reproducible model for the low-cost, low-quality sensor embedded in the smart glasses**. It is not professional or industrial-grade technology, and the team struggled to get it to produce reliable results. Each pair of smart glasses has slightly different sensor parameters that must be recalibrated over time.

"*It's a lot of trial and error,*" Romain reveals. "***This is mostly a lot of mathematical and computer vision tricks to get better contrast and edges on the pattern we want to detect***. *There are a lot of layers in our image treatment to get very precise corner positions from our markers. What makes our technology unique is the combination of every little trick we have to do to achieve this precision, but mostly, it's a tremendous amount of trial and error and testing.*"

The startup has already transitioned from ideation to execution. Multiple versions of its products have been developed and are being sold worldwide. Its collaboration with **Vuzix, a major player in AR smart glasses**, has been critical to its success. But while the startup has made significant strides, uncertainty always lingers on the horizon. Romain is keen to keep their progress in perspective and maintain an unflinching eye on the future.

"… we've performed more than 4,000 surgeries with our system!"

"We are a tiny French startup competing with huge multinational super corporates that have unlimited money to develop competitor products," he tells us. *"Success is relative because we have sold systems, have some incomes, and are still alive, but success is having a reliable roadmap and future relevance. It is still a struggle to find money, sell more products, and develop new features. We have relative success as we've managed to develop, qualify, and sell products. So far, it's far from being safe and certain and reliable. Many challenges still lie ahead of us."*

Romain has worked for Pixee since it was founded in 2017 in various positions, including mechanical engineer, software engineer, graphic designer, and everything in between needed in a startup. Sarah, who joined earlier this year, is confident that the startup's workforce is central to its journey, with their dedication and expertise driving the company's evolution: *"People make the difference,"* she asserts.

Romain concludes: *"We started as three people with a crazy idea, and now **we've performed more than 4,000 surgeries with our system**! To me, it looks quite incredible that it became real, and it's working, and people are using it. This is the joy of trying to achieve challenging projects like this!"*

Inferring Morphological Patterns of EGFR Gene Mutation in Lung Cancer Tissues Using Large-Scale Architectures!

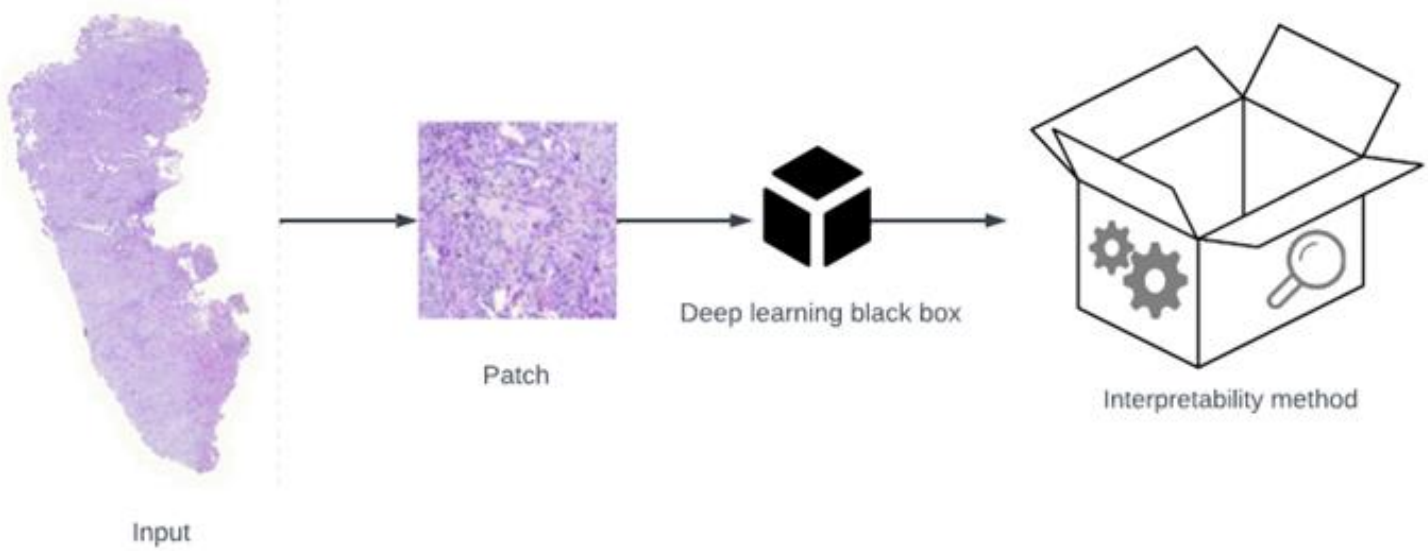
Hello, I'm Nisma! Welcome to my world of data-driven healthcare! I recently completed my masters in medical imaging and its applications. Currently exploring the intersections of digital pathology and data science to unlock new possibilities in healthcare.

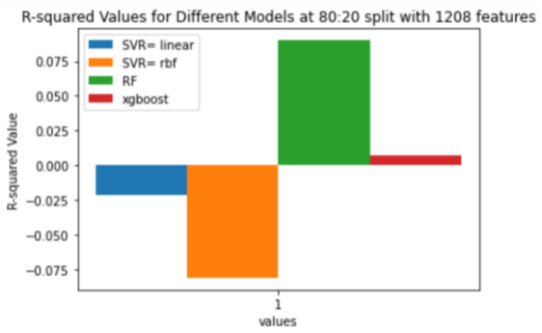
by Nisma Amjad

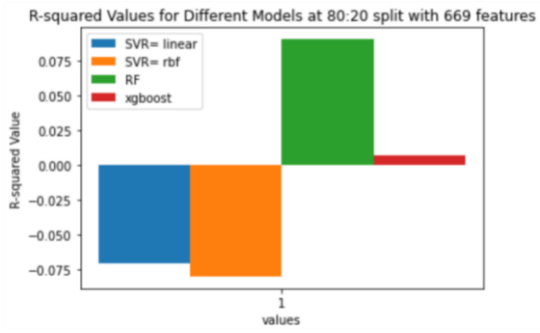
Welcome to the world of lung cancer research, where we explore the dynamic fusion of cutting-edge technology and the intricacies of the human genome. In this article, we'll dive into the exciting research done at **Ummon Healthtech**. Our goal? To uncover the secrets surrounding **EGFR gene mutations in lung cancer tissues**.

Lung cancer poses a significant global health challenge, ranking as the second leading cause of cancer-related deaths worldwide, as per the World Health Organization. To address this crisis, it's crucial to understand the different subtypes of lung cancer. Two primary types stand out: **non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC)**. Of these, NSCLC is the most prevalent, comprising 85% of cases and including subtypes like adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC).

Recent advances in deep learning have ignited hope in the field of lung cancer diagnosis and treatment. These cutting-edge technologies offer the ability to automatically extract critical histological features from lung cancer tissues, including tumor morphology, cellular architecture, and tissue characteristics.





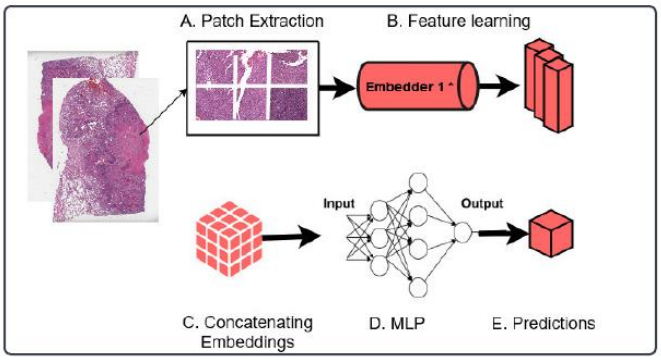


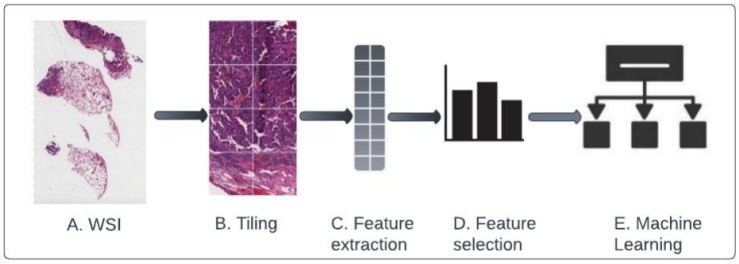
This automation translates to faster and more accurate diagnoses, a crucial factor in improving patient outcomes.

My research project at **Ummon Healthtech** centers around **EGFR gene mutations in LUAD, one of the NSCLC subtypes**. The primary goal is to unearth the underlying patterns associated with EGFR mutations, enhancing the interpretability of these complex processes and, in turn, revolutionizing patient care.

To achieve this, the research follows a systematic methodology that combines the strengths of deep learning and machine learning:

Generation of Prediction Scores: This step involves the extraction of patches from Whole Slide Images (WSI). These patches are then processed using an EfficientNetB7 neural network, which generates embedding vectors. The resulting prediction scores are crucial for the subsequent analysis.





Machine Learning Prediction Analysis: Handcrafted features are extracted from the WSI images. These features undergo random forest feature selection, and various machine learning techniques, including classification and regression, are applied to predict labels.

The results of this study are promising: in regression analysis, the Random Forest model achieved notable R-squared scores, indicating its efficacy in understanding EGFR mutations. In terms of classification, the Random Forest model outperformed other models, particularly when utilizing a subset of informative features such as GLCM, LBP, pixel intensity, and color moments.

As we conclude this journey, it's important to note that this research is just the beginning. Future work in this domain may explore advanced clinical and shape features and leverage a combination of deep learning and machine learning techniques to build even more robust models for lung cancer diagnosis.

Hans-Peter (Pitt) Meinzer (7.11.1948 - 18.8.2023)

Prof. Hans Peter ("Pitt") Meinzer led the division of Medical and Biological Informatics (MBI) at the German Cancer Research Center (DKFZ) until 2016. He was the founder of the open-source Medical Imaging Interaction Toolkit (MITK) and co-founder of CHILI GmbH, Mint Medical GmbH and mbits imaging GmbH.

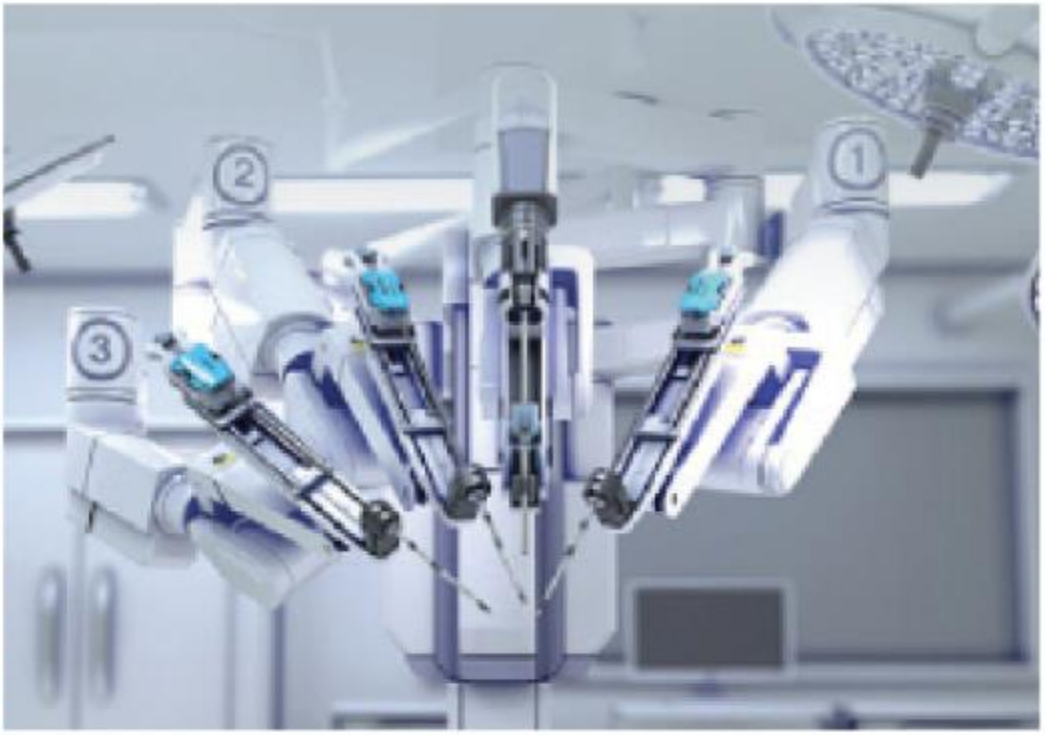
IF WE HAD KNOWN THIS WAS OUR LAST TIME TOGETHER…

We would have told you
that we wouldn't be where we are without you,
that it was never boring with you,
that we admired your courage to polarize, your love of life, your charisma.

We would have remembered the good times together;
the MICCAIs, SPIEs and BVMs,
the Christmas and bridge parties,
the freedom and support you gave us,
your advice on politics, leadership, and life,
your seminal presentations on how to eat with fork and knife.

We would have said THANK YOU!

Lena & Klaus Maier-Hein



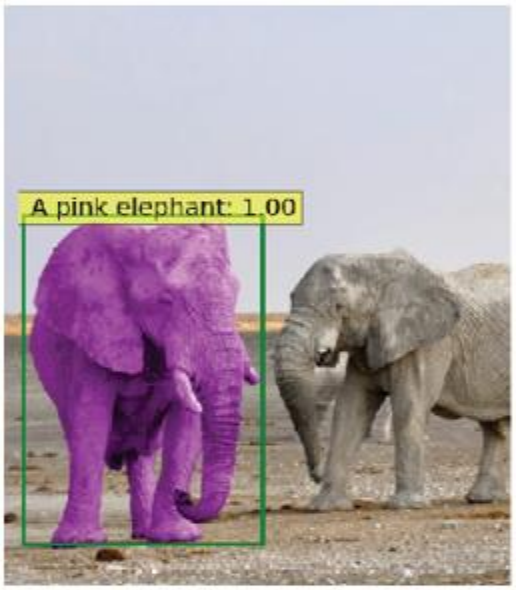
IMPROVE YOUR VISION WITH

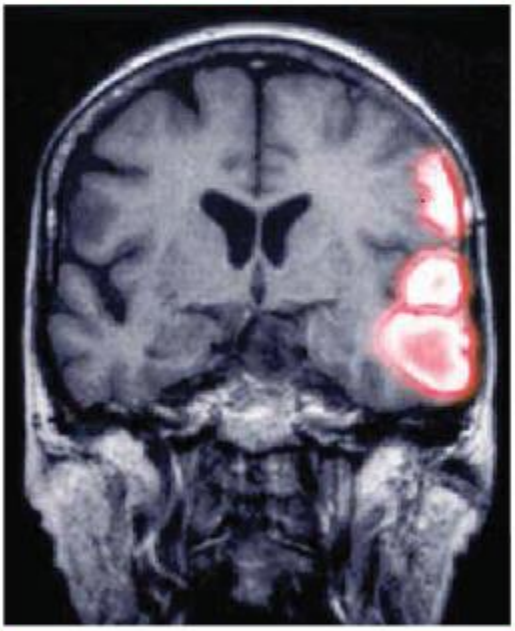
Computer Vision News

SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!





RSIP VISION