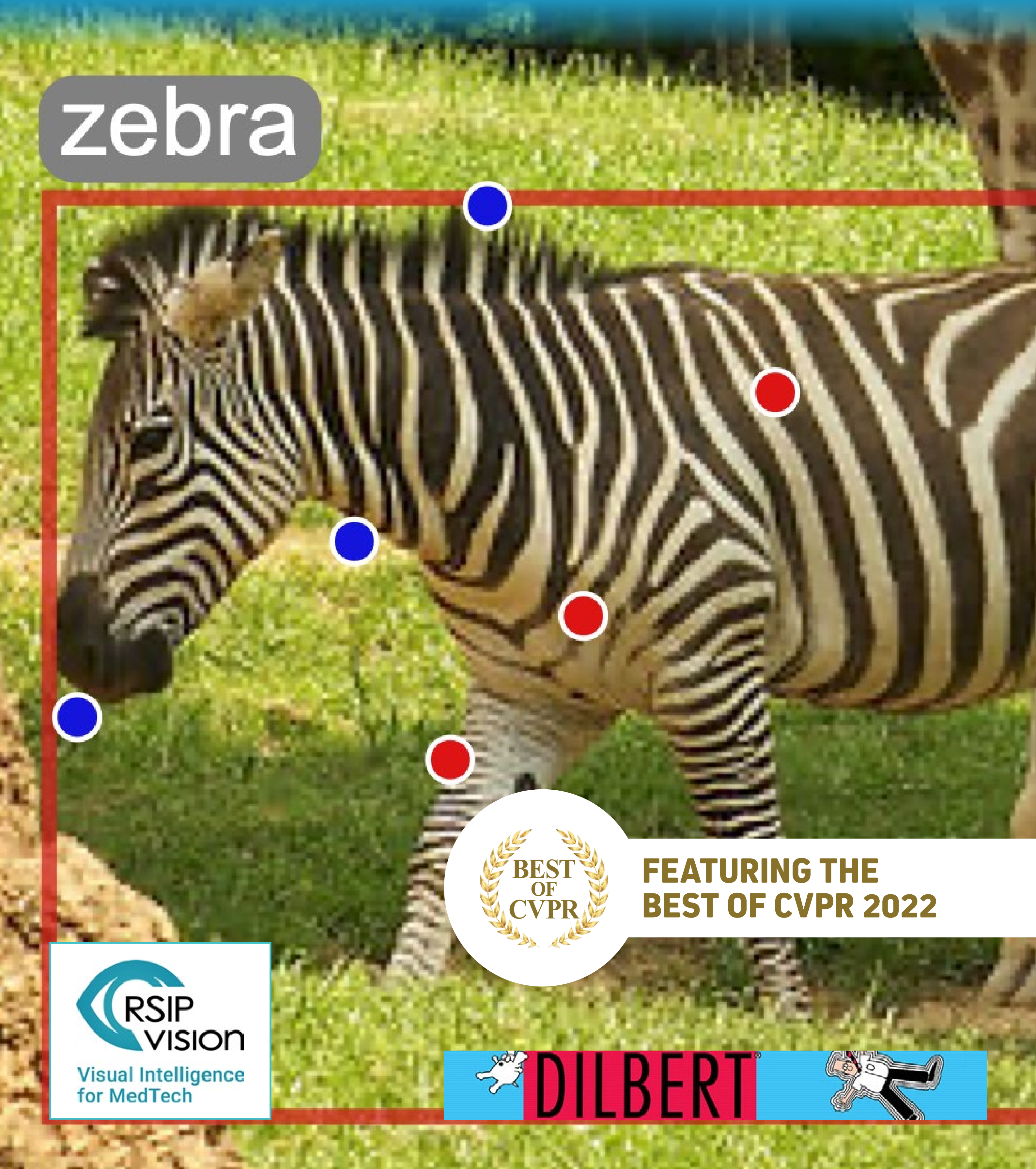
JULY 2022

Computer Vision News

& Medical Imaging News

The Magazine of the Algorithm Community



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

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editors:
Marica Muffoletto
Ioannis Valasakis

Publisher:
RSIP Vision

Contact us

Free subscription

Read Previous Magazines

Dear reader,

As one busy month closes, another one begins. Many of us are only just recovering from this year's outstanding edition of **CVPR in New Orleans**, which saw scientists from all over the world coming back together in person after two years of virtual events. It was a joyous occasion for all and proved that while we've all been through challenging times recently, **our community is stronger than ever!**

The scientific endeavor on display, with over 2,000 papers presented, was truly inspirational. If you couldn't make it to the US and missed out on attending virtually, have no fear! We have selected the cream of the crop for you in our **BEST OF CVPR** special beginning on page 4.

Away from CVPR, we've just held an exceptional webinar about the **current challenges for machine learning in medical imaging**, featuring **Klaus Maier-Hein**. If you missed it, you can catch up here.

This month, the medical imaging community is gearing up for the **Medical Imaging with Deep Learning (MIDL)** conference in Zürich, a hybrid event with many people attending in person for the first time since 2019. We'll once again be reviewing an impressive set of papers for you. Look out for our **BEST OF MIDL** feature next month!

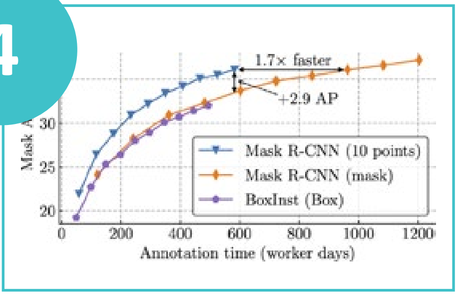
Enjoy the reading, don't forget to tell your friends about us, and subscribe for free!

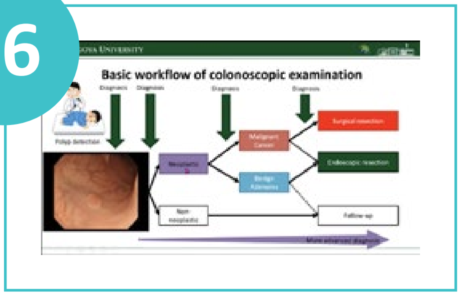
Ralph Anzarouth
Editor, **Computer Vision News**
Marketing Manager, **RSIP Vision**

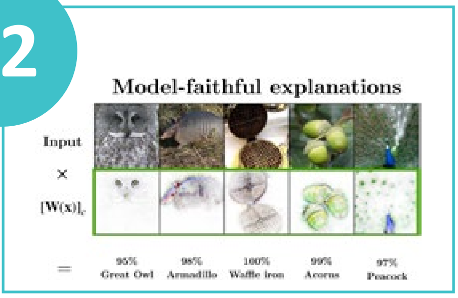
Follow Us

Computer Vision News

Medical Imaging News

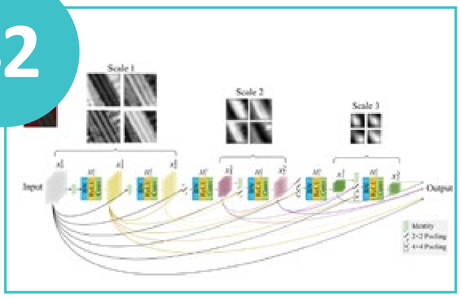
4


36


12


40


16


42


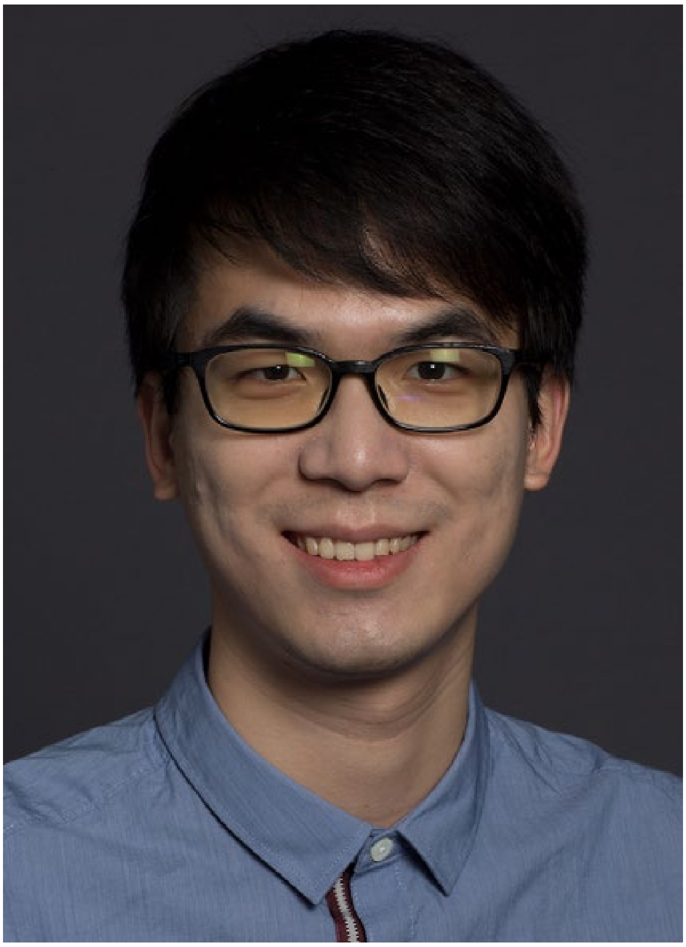
28


48
docker

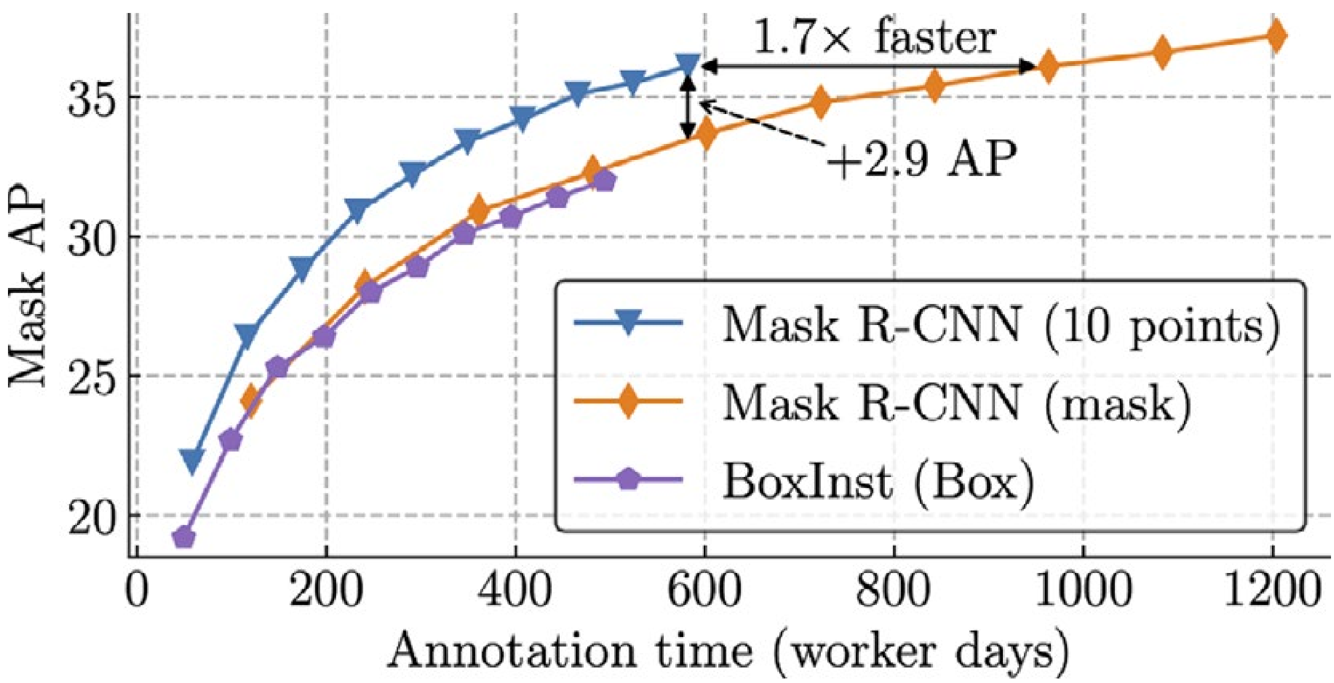
POINTLY-SUPERVISED INSTANCE SEGMENTATION

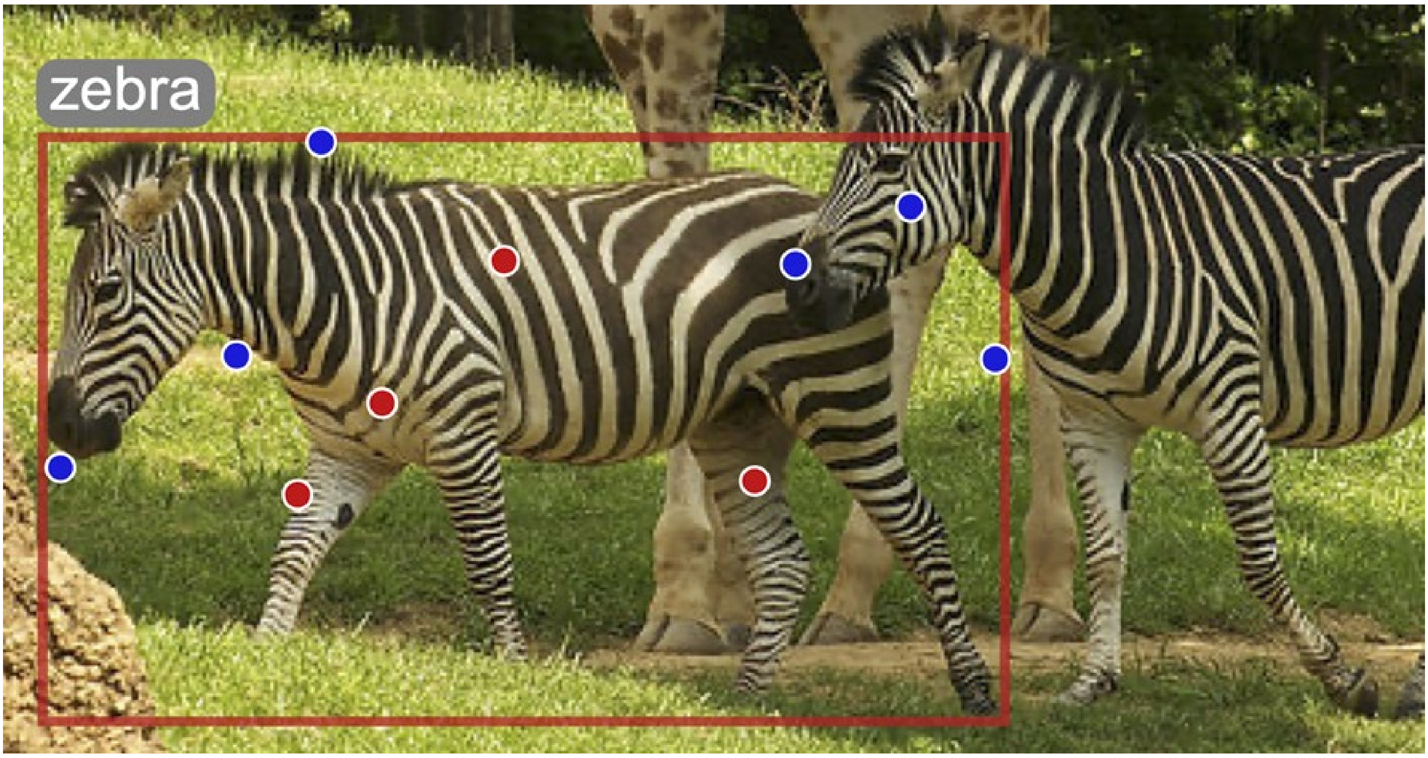
Bowen Cheng is a PhD student at the University of Illinois Urbana-Champaign under the supervision of Professor Alex Schwing. In August, he joins the Autopilot team at Tesla as a Research Scientist.

He speaks to us ahead of his oral presentation.



Instance segmentation is a computer vision task that locates objects in an image and delineates them with binary masks. When people train an instance segmentation model such as the Mask R-CNN model, it usually requires a data set with annotated masks of every object. You can spend around 80 seconds annotating a mask of an individual object, so the whole process takes a great deal of time.





"We try to reduce this annotation time by introducing a ***different form of annotation,"*** Bowen tells us.

"We're not trying to annotate the mask. Instead, for each object, we first annotate its bounding box. Doing this can take only seven seconds, which is more than five times faster than annotating the mask of every object. Then within this bounding box, we randomly sample a few points. We present each point to a human annotator and ask whether it is on the object or the background."

This reduction in annotation time provides the opportunity to collect more data for training these instance segmentation models. Also, this point-based annotation is compatible with existing instance segmentation algorithms. You can replace the mask annotation with this point-based annotation without changing the architecture, the training algorithm, or the loss.

When he started working on this, Bowen says another paper, **NeRF: Neural Radiance Fields – which received a Best Paper Honorable Mention at ECCV 2020** – had just come out.

"NeRF is used for 3D representation, and at that time, we were wondering if NeRF can represent an object in 3D, can we use NeRF to also represent an object in 2D and the mask of every object in an image?" he ponders.

"We found that when we trained this model, we didn't need the entire mask; we can sample several points inside this object to train the model, which is how NeRF works. We tried to take this one step further. If it works for this NeRF-style model, ***can these points be used to supervise any instance segmentation model?*** *Surprisingly, it works well for any arbitrary instance segmentation model."*

There are two ways to describe the

coordinates of a point, either continuous or discrete, but what happens when you are in a pixelized setting that in the real world is continuous?

"When you change from the ***continuous coordinate*** *to the discrete coordinate, you need to do rounding,"* Bowen explains.

"Then, when you try to transform back from the discrete coordinate to the continuous coordinate, you need to consider this rounding. We added a 0.5 offset because the discrete coordinates are integer, but the continuous coordinate could be an arbitrary floating-point number. We always add a 0.5 offset to the integer when we transform back to make up this rounding."

We ask Bowen what he thinks convinced the area and program chairs that his paper was good enough for an oral this year, and he points toward its ***surprising results***.

"We're not the first to try to reduce annotation time for instance segmentation, but in the past, when people tried to use weak supervision with less cost than the mask annotation, ***they always traded accuracy for annotation time,"*** he points out.

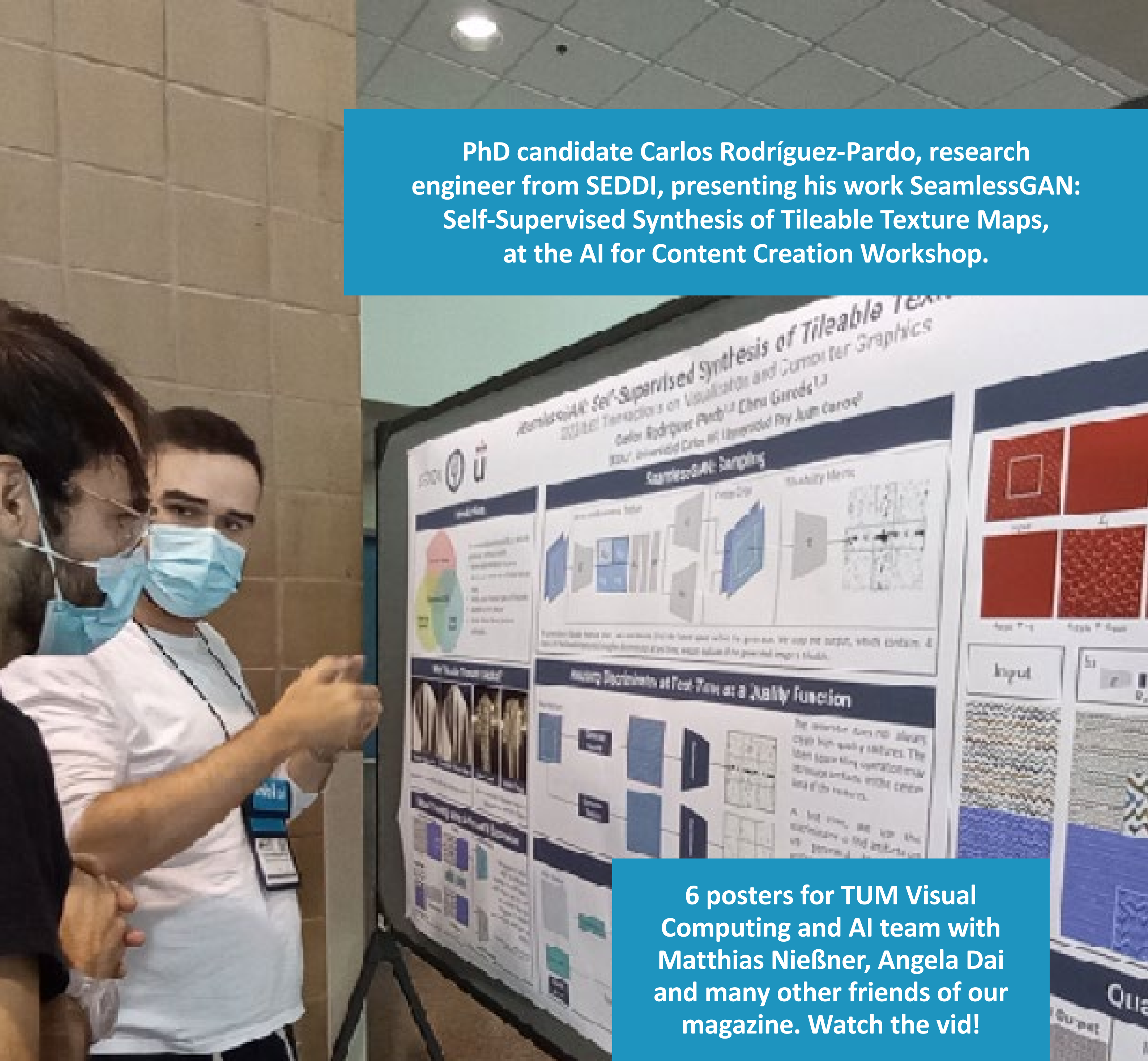
"Even the best methods can only achieve around 80% of the fully supervised method, whereas we get as close as 98%."

Even with a figure so close to 100%, Bowen says that if people aim for accuracy over reduced annotation time, they will still opt for mask annotation. Therefore, in terms of the next steps, he is curious to know if they could achieve 100% with much cheaper annotation.

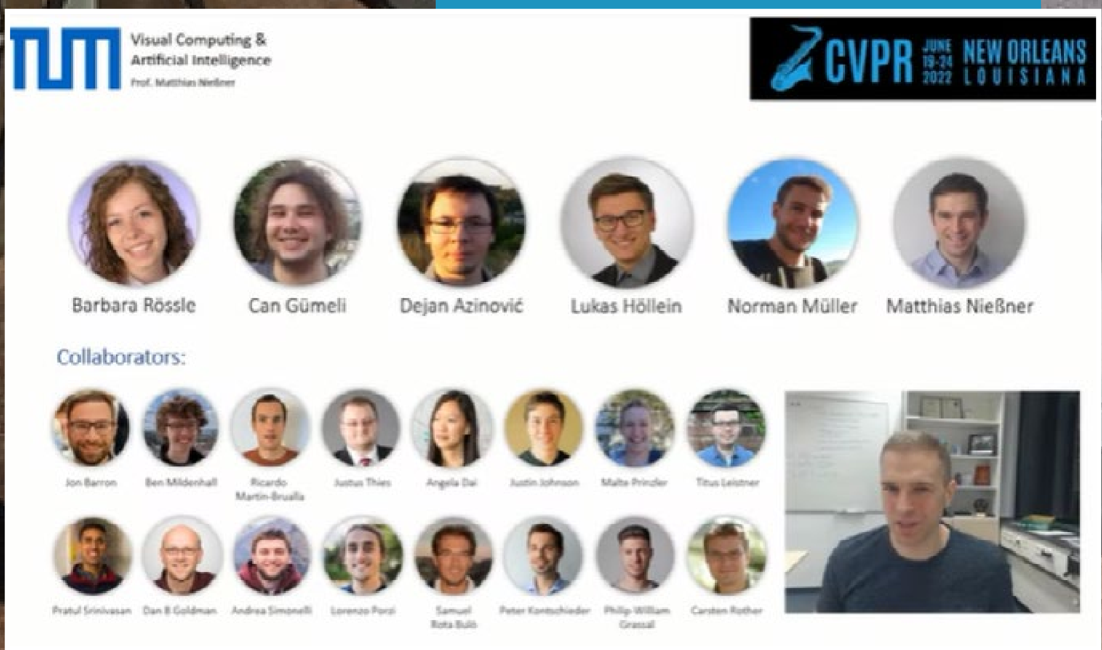
Bowen completed this work whilst interning at **Facebook AI Research** (now Meta), and he is keen to thank his mentor **Alexander Kirillov**, the last author of this work, for all his support:

"Alex is an incredible mentor. He always gives me insightful feedback, from discussing high-level ideas to digging into low-level technical or implementation details. This paper wouldn't be as successful without his support!"

PhD candidate Carlos Rodríguez-Pardo, research engineer from SEDDI, presenting his work SeamlessGAN: Self-Supervised Synthesis of Tileable Texture Maps, at the AI for Content Creation Workshop.



6 posters for TUM Visual Computing and AI team with Matthias Nießner, Angela Dai and many other friends of our magazine. Watch the vid!



INCREMENTAL LEARNING IN SEMANTIC SEGMENTATION FROM IMAGE LABELS

Fabio Cermelli (left) and Dario Fontanel (right) are third-year PhD students at Politecnico di Torino in Italy under the supervision of Barbara Caputo.

They speak to us ahead of their poster presentation.

This work aims to **incrementally extend a semantic segmentation model with new classes over time without using expensive per pixel annotation**. Instead, the model relies on image-level labels, which are cheap and can be found easily online. The goal is to exploit the model to extract the boundaries of the new classes by itself and then learn the new classes with this knowledge and the new annotations.

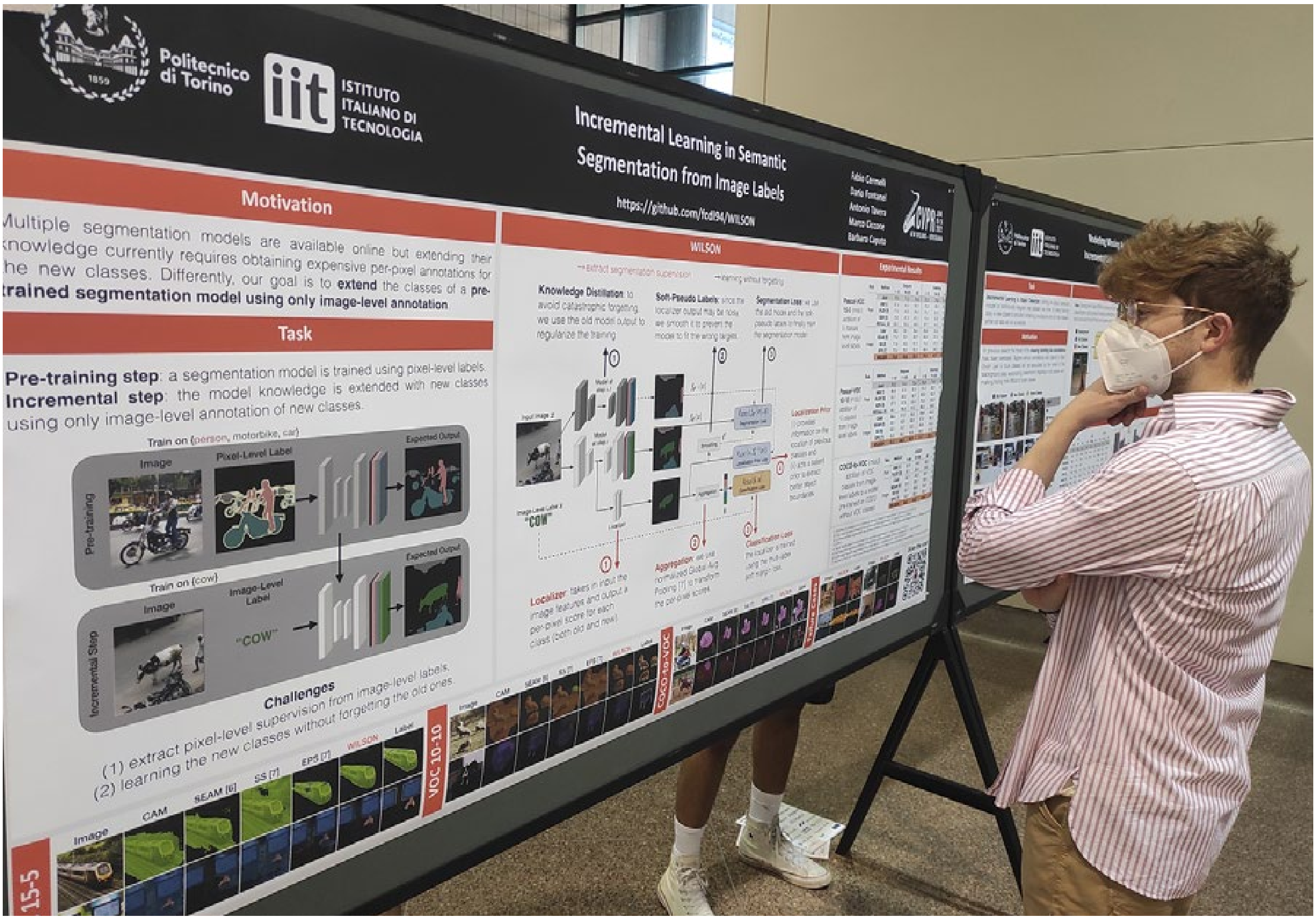
"We started with two things in mind," Dario reveals.

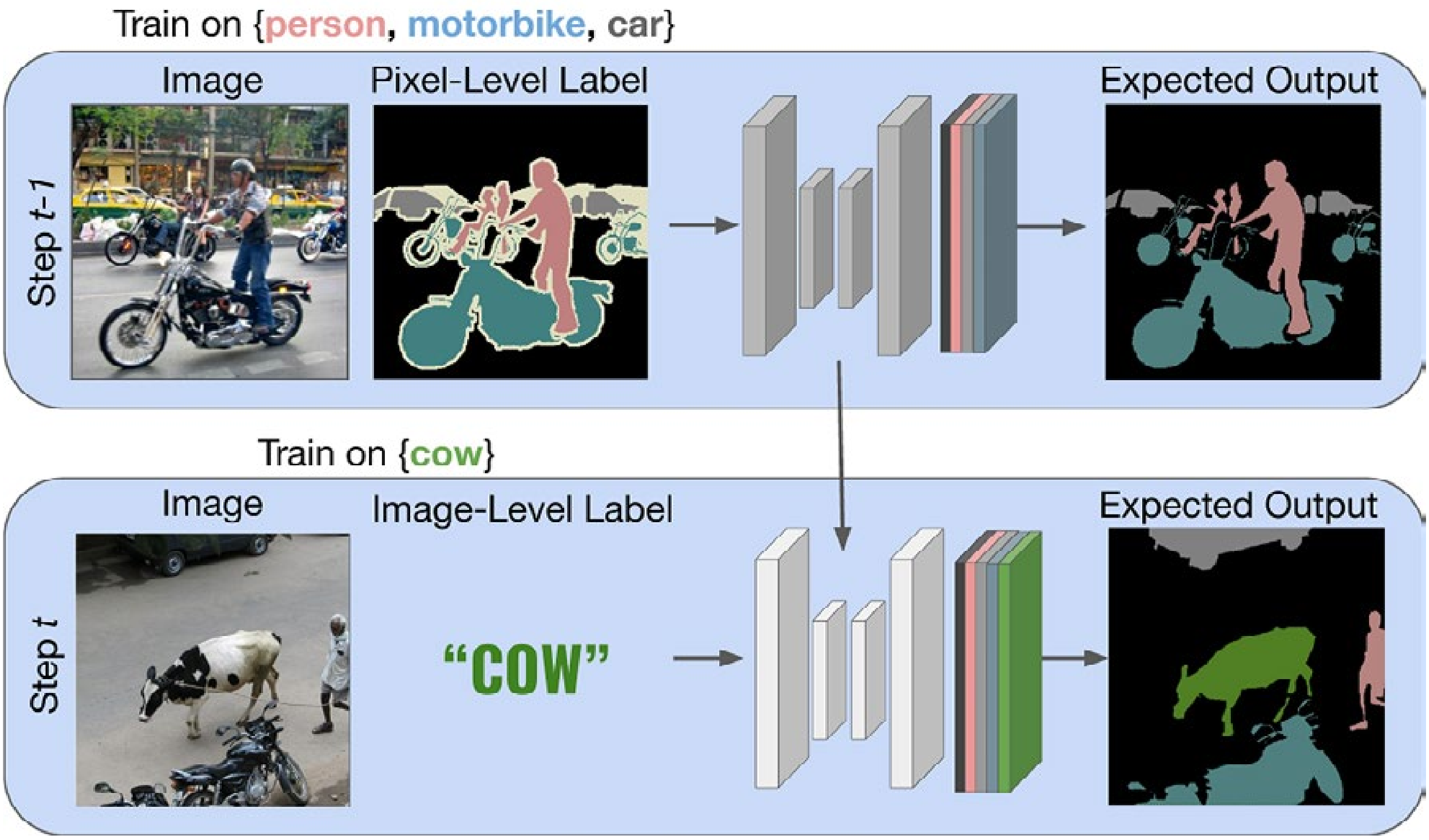
*"We wanted **something that could be used in the real world** because acquiring annotation, which is usually very expensive, is a challenge we all face. We also wanted to create **a model that could incrementally extend its internal knowledge** over time."*

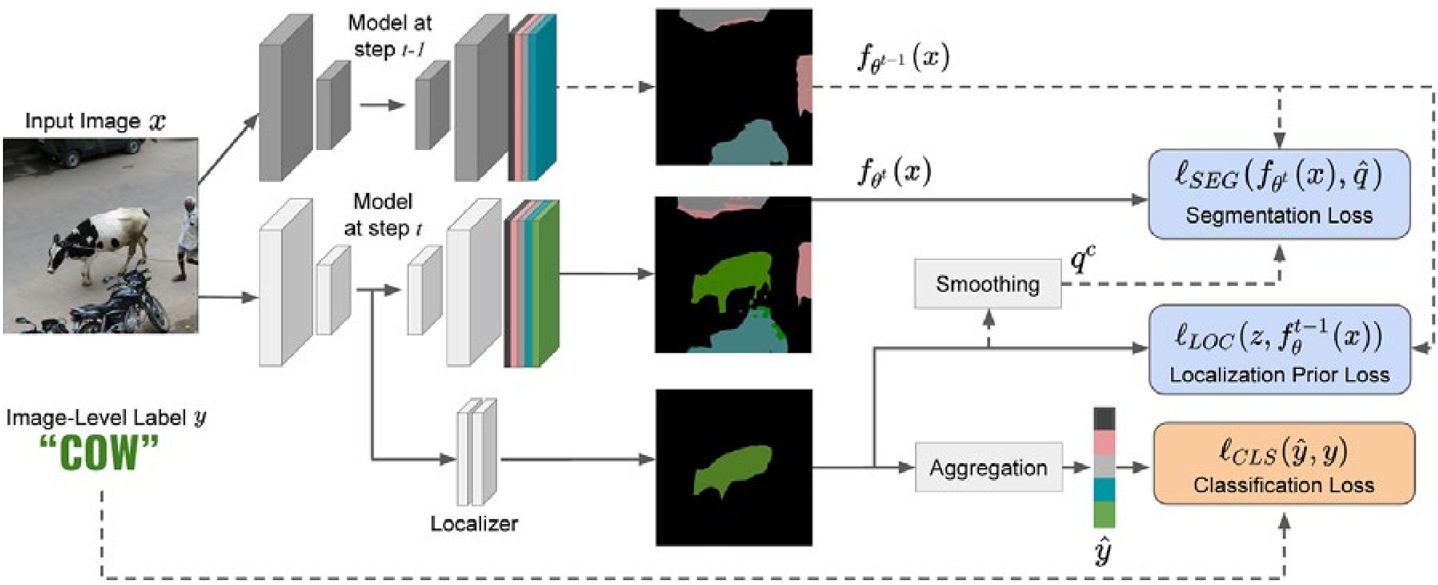
Fabio tells us that finding a jumping-off point for developing the model was the most challenging part:

"We tried lots of different things in the beginning. You won't read about those in the paper because they're failed experiments! The tough part was finding something that might work, even if not that well, but just as a starting point to develop our method."

The team mixed two well-known problems: **incremental or continual learning and weakly supervised segmentation**. Continual learning is a fundamental idea in machine learning, where you want to extend your model over time. It has been investigated in image classification and recently in semantic segmentation. Weakly supervised semantic segmentation is also


Politecnico di Torino
iit ISTITUTO ITALIANO DI TECNOLOGIA
Incremental Learning in Semantic Segmentation from Image Labels
https://github.com/fcdl94/WILSON
Motivation
Multiple segmentation models are available online but extending their knowledge currently requires obtaining expensive per-pixel annotations for the new classes. Differently, our goal is to extend the classes of a pre-trained segmentation model using only image-level annotation.
Task
Pre-training step: a segmentation model is trained using pixel-level labels.
Incremental step: the model knowledge is extended with new classes using only image-level annotation of new classes.
Challenges
(1) extract pixel-level supervision from image-level labels.
(2) learning the new classes without forgetting the old ones.
WILSON
VOC 10-10
15-5
COCO-to-VOC


Train on {person, motorbike, car}
Image
Pixel-Level Label
Expected Output
Step t-1
Train on {cow}
Image
Image-Level Label
Expected Output
Step t
"COW"



an old task, but this is the first time both tasks have been combined. The team hopes it will become a new field.

To train incrementally over time, the model uses the knowledge of previous timestamps to maintain the knowledge of the classes that it has learned earlier to avoid forgetting them. The team uses techniques like **segmentation loss**, which allow the model to predict a label for each single pixel, and in this specific scenario, there is a common solution that averages the model's features to be trained at the image-level label.

"We use a technique called CAM, or Class Activation Maps*, and make an average of the scores on the images to train them on the image-level labels,"* Fabio explains.

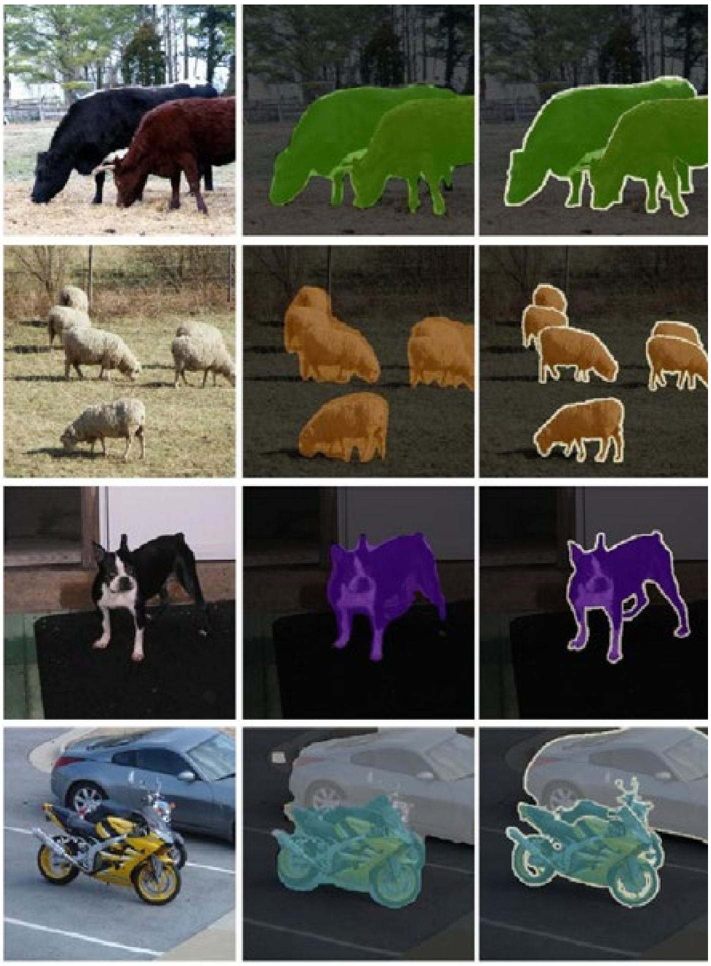
"We use this at the image level to extract a localization for the new classes, which helps us to train a segmentation model."

Dario adds: *"We've achieved results* ***comparable with a fully supervised semantic segmentation model****, even with our low annotation costs. This model has the potential to be applied to many real-world applications, and I think that it's a very valuable contribution."*

Fabio and Dario are missing their co-author **Antonio Tavera**, who cannot be here to present today. Fabio recalls a special eureka moment they all had together, and without which we may not be speaking:

"We had been in the lab all day, and it was a Friday night before the conference deadline, and we were in a restaurant with Antonio, still talking about our model. That's when we had the idea of the final model that got us these last performances. I remember telling Dario we can try that tomorrow, and it was a very last-minute tweak!"

The novelty of this work and its application in the real world are just two things that likely helped it to be chosen from so many



Pascal-VOC (10-10)

COCO-to-VOC

other papers to be presented at CVPR this year.

"We've proposed ***a simple yet effective solution to this problem that has achieved the highest performance compared to the current state of the art,*** *and that could be extended in many different ways," Dario points out.*

Do they have any advice for others in a growing field with such strong competition?

"It's crucial to collaborate with other people and exchange ideas with your teammates and others in the same field," Dario responds.

"That allows the community to grow and develop new ideas."

Fabio agrees: *"I encourage everyone to continue bringing brave ideas to the community and trying new paths. Something new, challenging, and applicable in the real world."*

We point out that he may be creating next year's competition. He smiles:

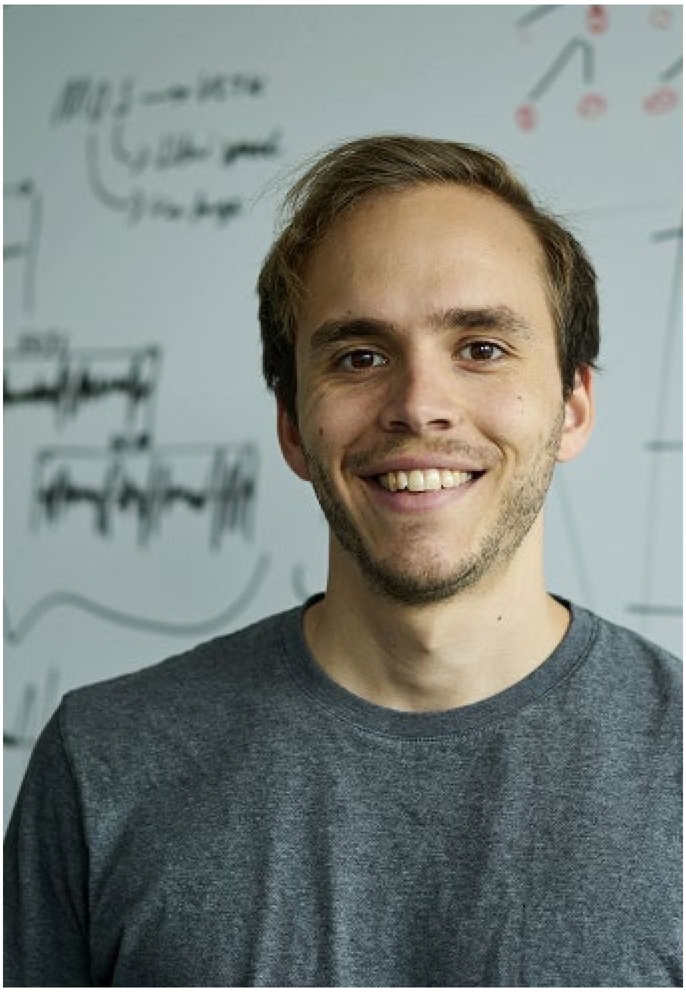
"Exactly. We hope so!"

BEST OF CVPR

B-COS NETWORKS: ALIGNMENT IS ALL WE NEED FOR INTERPRETABILITY

Moritz Böhle is a fourth-year PhD student at the Max Planck Institute for Informatics, under the supervision of Bernt Schiele.

He speaks to us ahead of his poster, which explores a novel direction for improving the interpretability of deep neural networks.



Deep neural networks are powerful and work well in classification and many other tasks, but **extracting faithful explanations for the model's decisions** is typically much more challenging.

These explanations are needed to increase trust in machine learning-based systems and to learn from the systems because we might not know the important feature of some datasets. This is especially true for medical data. If systems are interpretable by design and highly accurate in their classification, then there is the chance to understand better which features are indicative of, for example, a particular disease.

"The most common approach is ***to start with a pre-trained network****," Moritz tells us.*

"You're agnostic to how it was trained. You know it's a network for classification, for example, and then people try to explain it after the fact in a post hoc fashion. If you find a good explanation method that works for all networks, you maintain the classification accuracy and can explain any network."

The problem with this approach is that you do not know what the network used as a feature for classification, so you will not know whether these explanations can be

trusted and if they are faithfully explaining the model.

*"In contrast to common approaches, **we've designed and trained a deep neural network that is already optimized for interpretability**,"* Moritz explains.

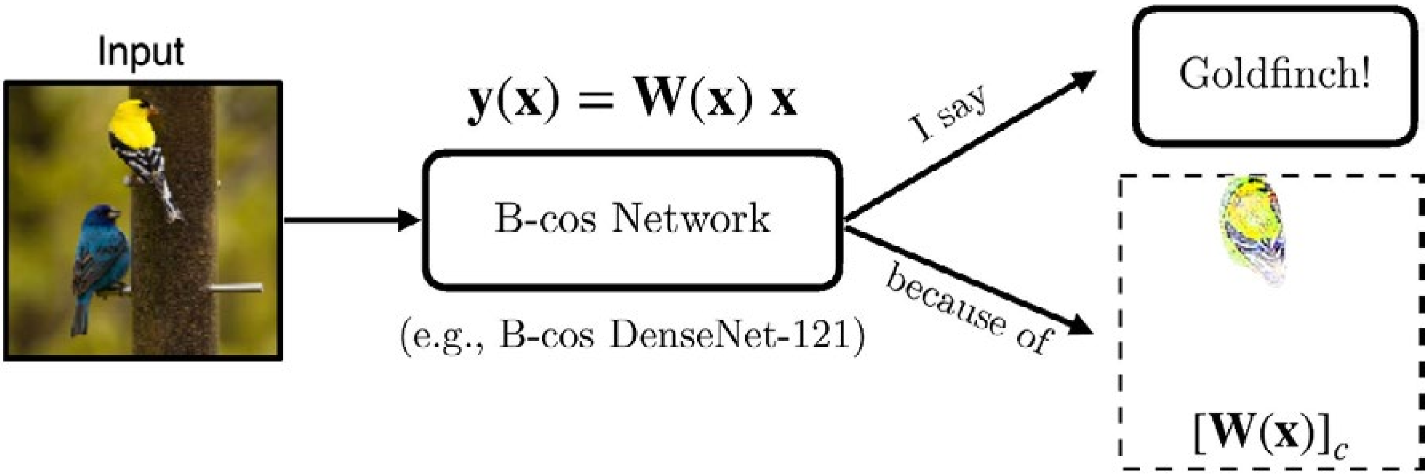
*"It's **inherently interpretable** so that you can easily obtain an explanation for the model decision. We did that in the context of image classification. We trained on the **ImageNet dataset** and managed to optimize the network in terms of architecture and the training paradigm so that **it will align** its **matrix with important features in the input**."*

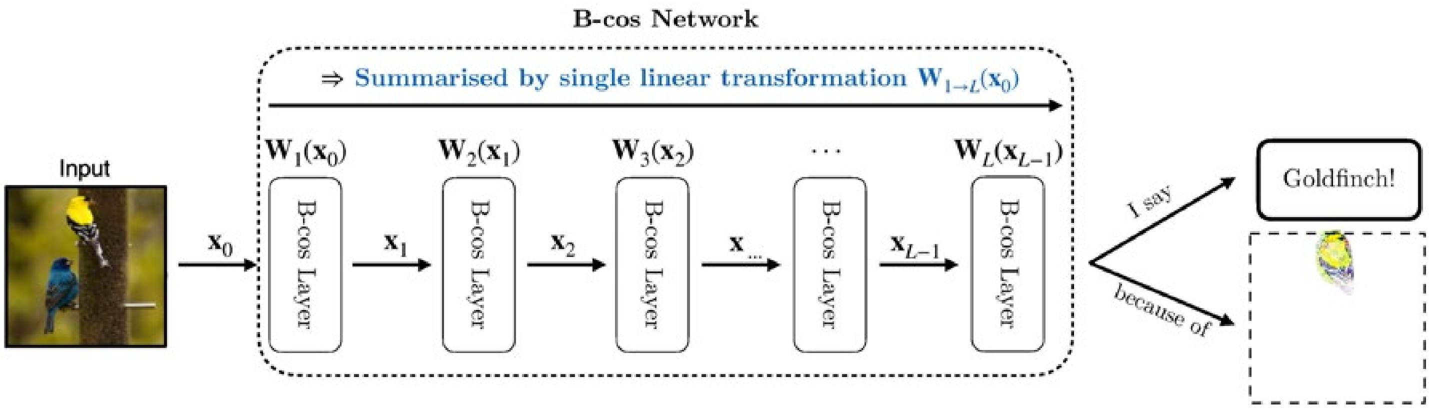
This framework can take existing architectures and make them inherently more interpretable while maintaining the original classification accuracy for the most part. **All the code is publicly available**, and experiments can be reproduced.

In terms of future development, Moritz says they are already working on a follow-up to this work, showing that **this approach can be integrated into vision transformers** to make them inherently interpretable. We suggest this could be a candidate for ECCV later this year.

"Let's see where we can get it!" he responds. *"In the future, I want to explore if we can use what we learned during this process to make natural language processing, audio processing, or other tasks more interpretable."*

Moritz presented this work at the **XAI4CV: Explainable Artificial Intelligence for Computer Vision Workshop** on Monday, a new workshop organized by **Meta AI**, demonstrating the high interest in this area now. It proved very popular.





"I didn't manage to go to many invited talks because there was such a lot of interest in my poster!" he laughs.

"I got very positive feedback, which encourages me."

Why does he think his work has proved so popular, particularly picked out from the competition for a coveted CVPR poster slot?

"I think it's because typically, there has been this trade-off between accuracy and interpretability," he answers.

*"We've been getting better and better in classification performance, but we've been stuck when it comes to understanding exactly how the models perform their tasks. We show in our paper that **if we design the models the right way, we can maintain accuracy but build inherently interpretable models**."*

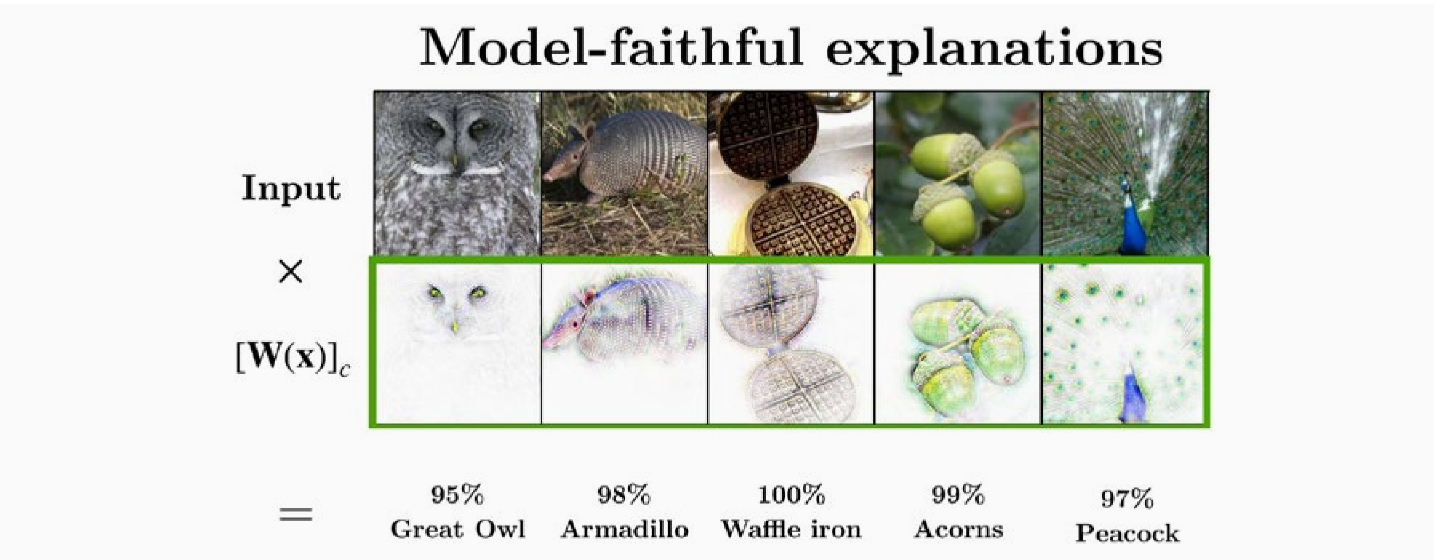
Finally, we tell Moritz that **Bernt Schiele**, his supervisor, has told us to give a look at this work.

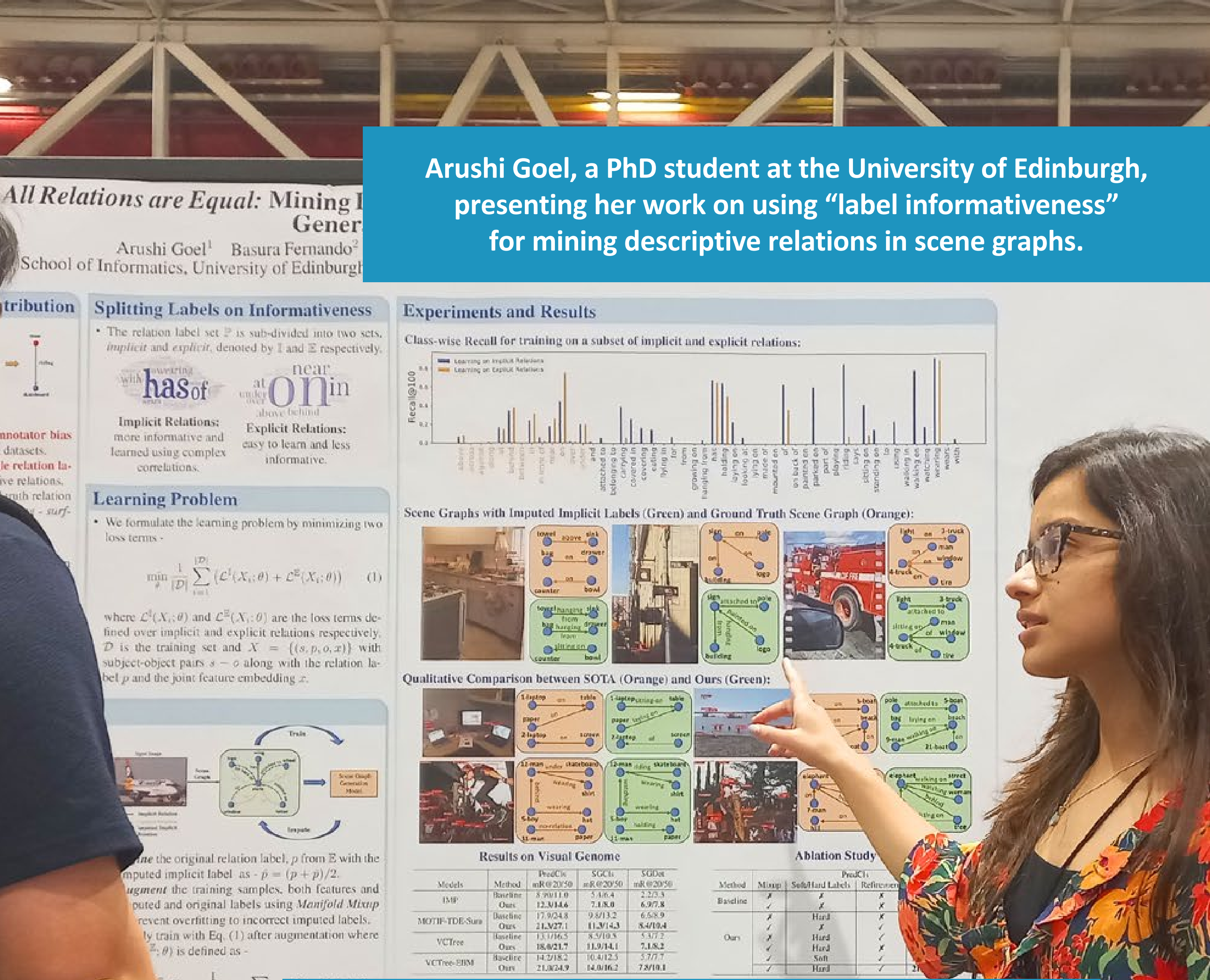
"This makes me so happy that he's supportive of this paper and took the chance to recommend it for CVPR Daily!" he smiles.

"Working with Bernt is special because of the close supervision, discussions, and feedback. It helps me refine my argumentation. He's also very fast! Both in terms of speaking, as many people will know, and in terms of understanding and connecting the dots."

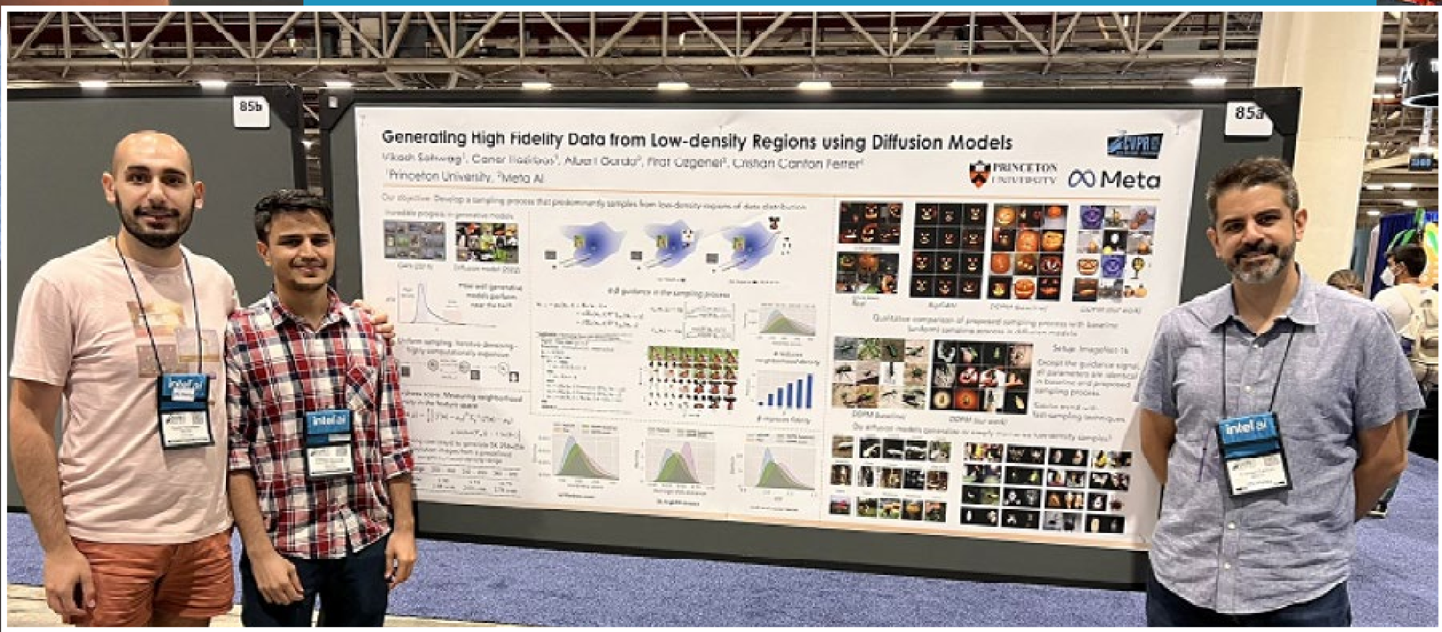
Is it true he can make the best out of the people he works with?

"Yes, that fits him very well. He knows where the value lies."





Arushi Goel, a PhD student at the University of Edinburgh, presenting her work on using "label informativeness" for mining descriptive relations in scene graphs.



Supervisors Caner Hazirbas (left) and Cristian Canton (right) find this work by Vikash Sehwag particularly relevant to generate rare class examples using diffusion models.

“I think what I do now is nothing compared to when I was training and going to university at the same time!”

Kakani Katija is a Principal Engineer at the Monterey Bay Aquarium Research Institute.

Read 100 FASCINATING interviews with Women in Computer Vision


MBARI

Kakani, can you tell us about the Monterey Center?

The MBARI, the Monterey Bay Aquarium Research Institute, is a research institute that focuses on developing technologies as well as methods for studying proxies or animals in the deep sea, originally founded by David Packard in the late 80s, maybe late 90's. It's about 250 people, ranging from engineers to scientists to marine operations people and communication.

MBARI is located in Moss Landing, California, which is between Monterey and Santa Cruz. Our sister institution is the Monterey Bay Aquarium, which I think more people have heard about or know about.

You're not originally from Monterey. Can you tell us where you come from?

I'm originally born in Hawaii, raised in Portland, Oregon, and then went off to undergrad at the University of Washington, Graduate School at Caltech, and then post-doc in New England, and then back in California for my current position.

Let's start by you changing fields.

Yes, I have changed fields. My previous field was aerospace engineering. It wasn't until graduate school that I realized that there were plenty of opportunities to develop technology to study biology, particularly biology in its natural environment. I got pulled into the ocean, literally and figuratively. Now I've been a member of the ocean sciences community for almost a decade, maybe a little bit more. We use imaging in novel ways to understand biological systems. So that's essentially the transition that I've gone through.

What do you want to discover in this new field?

All sorts of things! For me, my particular interest is in bio-inspired design. The idea is that if we can study animals, we can study them very well, and we can perhaps back out the kind of mechanisms that animals use to survive. Those mechanisms could potentially be used for other technologies that we probably haven't even dreamed up yet. What I focus on is trying to develop technologies that allow us to ask those kinds of questions of animals in their natural environment, and particularly in difficult access places like the deep sea.

Does this include understanding photography in the sea or it's a completely different thing?

No, it's part of that as well. Underwater imaging is so different from terrestrial imaging. The conditions are very different. There are a lot of things to learn from that process.

Maybe I should introduce you to Derya Akkaynak. She does great work in this area.

Oh, I know Derya!


MBARI

MBARI

OK, I don’t need to introduce you. What would you like to achieve in this field?

this field?

That’s a good question. A couple. The reason why I’m here at CVPR this year is because we’ve been building out a label data set called FathomNet to help with fine grain categorization, to try to automate processing visual data with the long-term goal of being able to detect when we’ve come across animals that are known or unknown to science. So to me, having a data pipeline in place where we can know right away if there’s a new animal completely undescribed that needs further study. Being able to do that whole process would be fantastic. For the bio-inspired design stuff, what I would love to be able to do is kind of multi-view, three-dimensional full reconstructions of animals that are time-varying so that you can take that reconstruction and put it in a computational fluid dynamic solver and be able to say something about the movement and physics for application to design.

What chance do you give yourself to achieve this?

With a lot of collaboration, probably pretty high. The challenge is a lot of the animals that we study aren’t the classical opaque bodies. These are more transparent, translucent animals. There are really no comparable models out there to do that kind of work.

What do you miss the most, from aeronautics? Do you regret anything?

I mean, if I can regret not being an astronaut, but I feel like I don’t have very much control over that process. I think what I miss the most is being able to do very straightforward repeatable measurements or studies of physics, where once you start delving into biology and organism systems, there’s almost no replicability. It is so difficult. Once biology is involved, it's far more complicated. And also natural environments are areas where you're trying to do studies or measurements with zero control over the condition. Set me up in a laboratory any day!

You have a resume of 27 pages, and you're not that old yet. What motivates you?

I feel old now...

You are half my age, and I barely fill one page.

I don't know if that's true.

What makes you run so fast and so far?

I mean, really, it's a testament to the wonderful people I've been able to collaborate with. There's no way if you're doing this kind of multidisciplinary work that I do that you can do it on your own. I've just been very fortunate to be able to work with these people.

Can you tell me the best teaching you have received from a teacher?

Yes. The best advice I ever received from someone was to make decisions about opportunities that lead to more opportunities. I think that's very clear in some of the stuff that I've done to move forward.

I do that all the time. I'll always ask questions that lead to more questions.

I'm sure! *[long laugh]* A professional...

Without learning that, I knew that already! *[we laugh again]* I suggest that our readers also take advantage of this advice, because it's a very precious one, independent of the field you work in.

Exactly. It works for everything.

Do you want to say the name of the teacher who taught you that?

Bob Breidenthal, who is a professor in aerospace engineering at the University of Washington.

Certainly a very impressive person.

Yes, very impressive.

I'm glad you met him because both of us are wiser now because of him.

I know. Right?

What is the greatest lesson that you learned from your students?

Well, there's a bit of a caveat there, because technically I don't have any students. It's a nonprofit research institute, so we don't have any academic standing.

But, I do mentor a lot of young people, and I think for me, I've learned a lot from them.

One of the things that I've learned from them is how to become a better mentor because everybody is different. They all have different approaches. They reach out in different ways, and it's recognizing what makes these people tick.

Do you make mistakes sometimes?

Oh, always. But then also learning from them to avoid…

… too much collateral damage?

Yes, it's true.

You've succeeded in many impressive ways, including sports! Maybe you want to share something about that.

It's something I don't talk very much about. It's something that I attribute to a lot of what I do now. I was a competitive figure skater on the US International figure skating team for several years. I think what I do now is nothing compared to when I was training and going to university at the same time. It gives me a really great perspective on time management.

Did you make the Olympics?

I was an alternate for the Olympic team, so I did not make it.

I am also a skater, but of a very different kind, an amateur rollerblader. There is almost no bone that I haven't broken yet! So I envy you because you certainly did not break a lot.

Not too many, but towards the end, I had a lot of injuries.

Tell me something that you learned in sports that helps you daily, whether it be discipline, effort, companionship, or anything else.

I would say discipline. I would say goal setting. The fact that as an athlete, you can set a big goal like going to the Olympics, and it would take years, decades even, to get to that point. It was being able to break down the problem or the challenge into bite-size pieces so that you can eventually

attain that goal. Then being able to apply that to professional things or to personal things and continuing that process later on in life, I think, has really served me well.

How different would you be as a scientist if you had stayed on one path?

That's a tough one. I don't know. But I think the path that I have been on is richer for all of those interactions and experiences.

Do you work for the community?

I work for the community. A couple of things. Obviously, I'm heavily involved with mentoring young and new, particularly diverse students, either in ocean sciences or engineering. More recently, through this FathomNet project, what we're trying to do is create mechanisms for enthusiasts so people who are not classically trained in biology and marine biology can actually contribute to the data analysis pipelines, the science that we're trying to conduct in the deep ocean. We're building community networks of taxonomists, enthusiasts, programmers, and ecologists to try and help work on this problem to create a global network for ocean life discovery. That's what we're trying to do now. That's the big goal.

And you'll achieve it with the group effort?

Only with the group effort!

What would you have done if you didn't become a scientist?

That's hard because I feel like science is in almost everything, like architecture.

Can you imagine doing anything else?

I guess there's science everywhere.

I'm not a scientist! I work in marketing. I barely know the three laws of Newton...

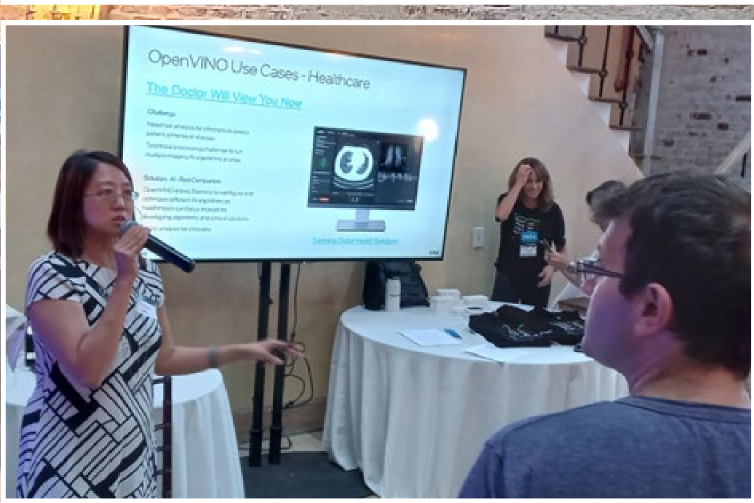
That's an interesting point. What would I do? I actually love journalism. That's my favorite!

Read more than 100 interviews with Women in Computer Vision!

Intel AI Happy Hour

Intel hosted a successful workshop and afterparty that enjoyed a very positive response. Paula Ramos asked us to thank all those who participated: *“At Intel we wanted to share our OpenVINO toolkit with developers, as it has been made by developers for developers to help simplify and accelerate AI app development.”*

At the Intel event, awesome Ilke Demir (left) presented: Real-time FakeCatcher - Deepfake Detection with Intel AI

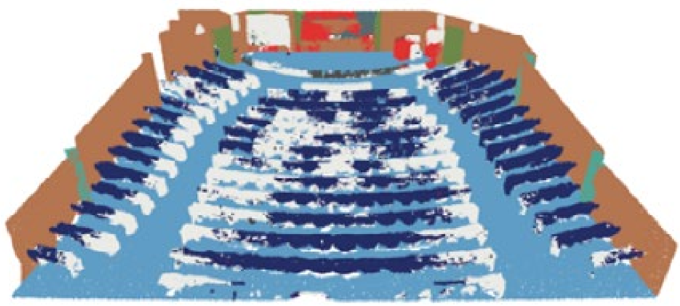


LEARNING MULTI-VIEW AGGREGATION IN THE WILD FOR LARGE-SCALE 3D SEMANTIC SEGMENTATION

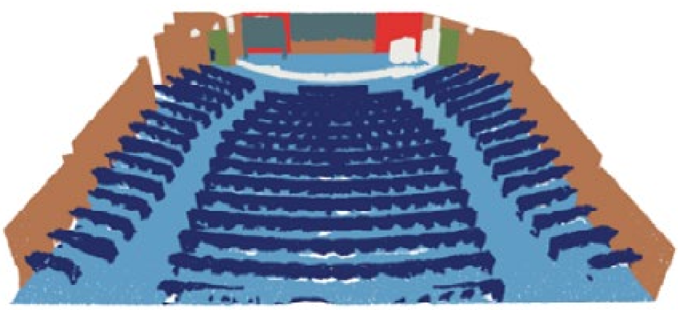
Damien Robert is a PhD student at IGN, the French national mapping agency, and Gustave Eiffel University. ENGIE also funds his PhD.

As the first author of his first paper on 3D semantic segmentation, he speaks to us ahead of his oral presentation.

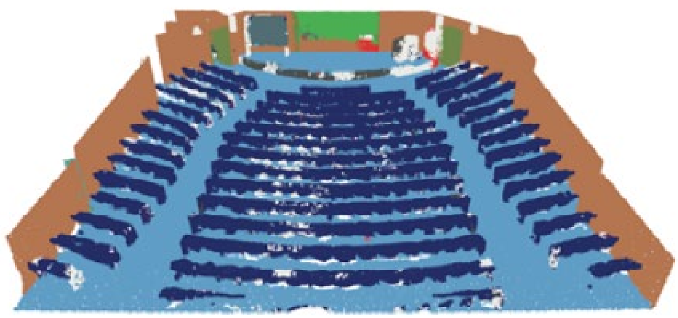
Colorized point cloud



3D backbone



Ground truth



Ours

Point clouds and images are the typical data produced by **3D acquisition systems**. This work proposes a simple set-up to merge the two without requiring meshing or depth sensors, but simply **raw point clouds, images, and the corresponding camera poses**.

When you have multiple images of a 3D scene, each object may be seen in a variable number of images. Some images may be close-up, some far away, while some may be occluded or see things from slanted angles. This situation is called the **multi-view problem**.

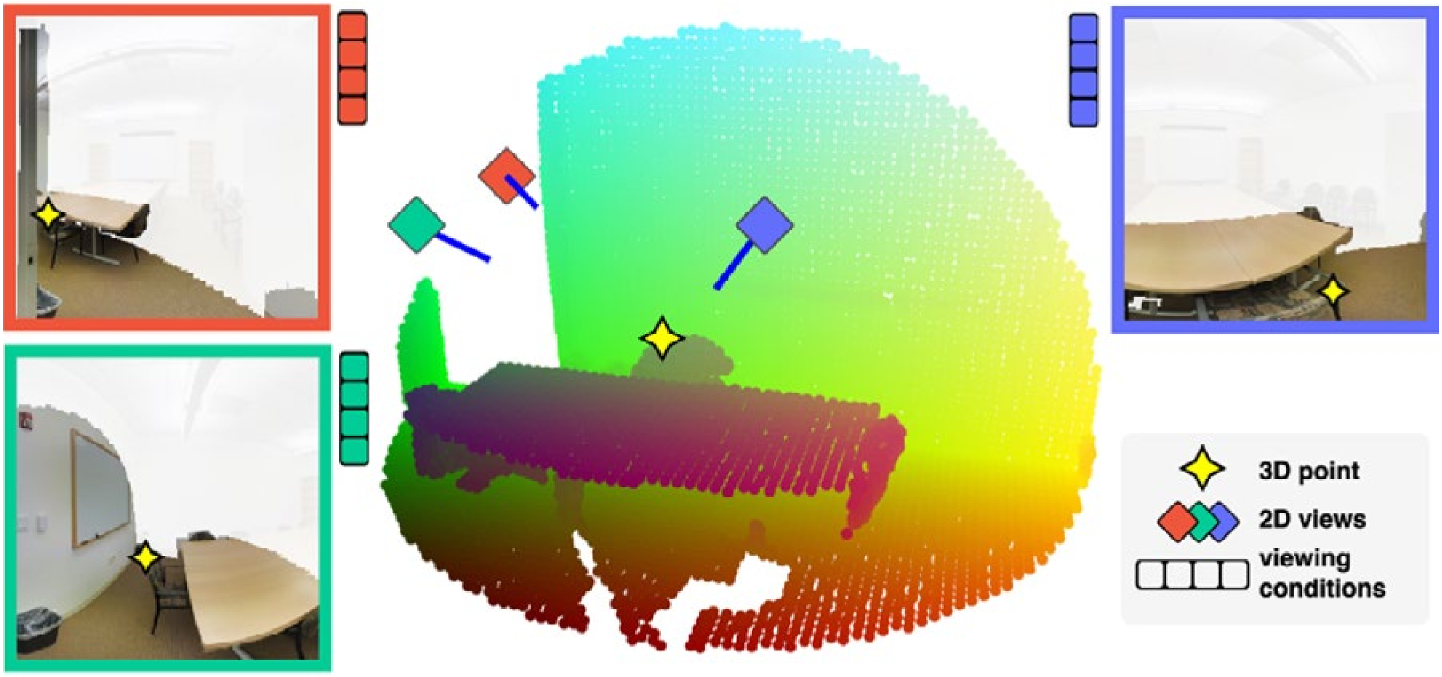
"Acquisition systems often produce 3D point clouds and images, and we want to be able to combine the information from each modality because they are complementary," he explains.

"The point clouds describe the scene's geometry, and the images describe other things like the texture and the context. We want to use information from both. We're not the first ones to do it, but we're the first to learn multi-view aggregation for large scenes."

The team wanted to work at a large scale, with **scans of cities and buildings** rather than a small point cloud of an isolated chair, for example, where you only have pictures of that object. However, manipulating the information that connects the point cloud and images at a large scale is difficult.

"I had to spend a lot of time coding to efficiently manipulate the link between images and point clouds and never break this connection through the whole pipeline," Damien tells us.

"You must ensure that each 3D point is properly connected to the corresponding pixels. It's very tricky."



As some objects can be seen by one image, some can be seen by several images, and some are not seen by anything, the team had to find a way to deal with this fluctuating information. To solve this, they used an a**ttention mechanism**.

"We wanted the network to focus its attention across various inputs to combine their information," he explains.

*"We evaluated our method on two large-scale data sets – one called **S3DIS (Stanford 3D Indoor Spaces Dataset)**, which has been used a lot as a benchmark in 3D semantic segmentation, and one called **KITTI-360**, which is more outdoors – and obtained state-of-the-art results on both."*

The team demonstrated much better results when extracting features from images than just colorizing the point clouds and hoping a 3D model would understand what the colors mean. Also, they proved that using a simple aggregation like taking the maximum or the average of what the images have to say about an object is not as powerful as trying to learn this attention scheme.

What is Damien most proud of about this work?

"**The way it's coded**," he responds.

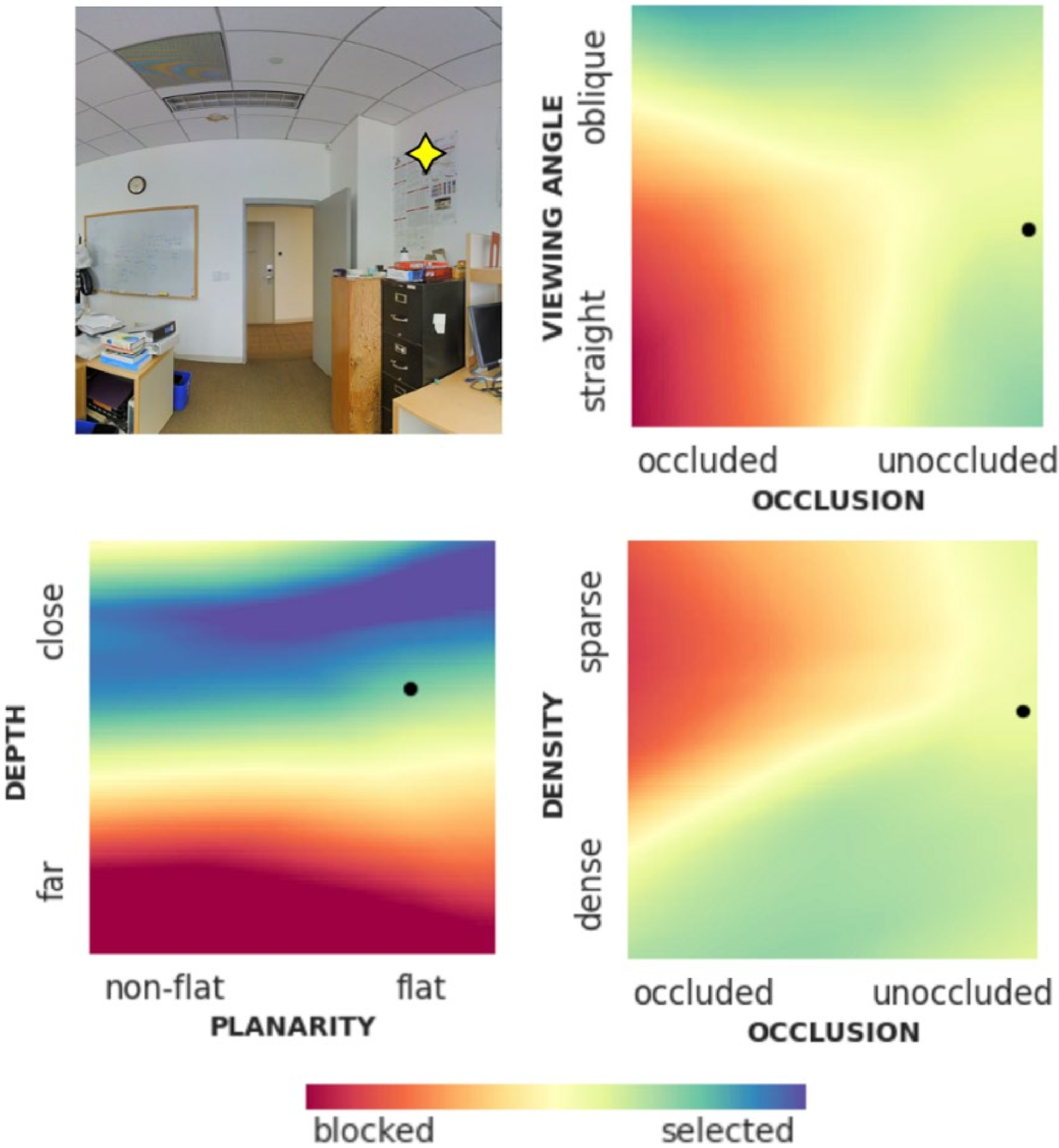
"It's not something we pushed forward in the paper because it's more of an engineering problem, but it required a great deal of effort, so I'm proud of what I've been able to build. It's a hidden element, but it's important so that you can do something in a reasonable time and in an efficient way."

Damien tells us the team is already discussing its next steps. They are looking at extending a paper that co-author **Loic Landrieu** worked on called **Superpoint Graphs**, designed to learn on point clouds at a large scale.

"We'd like to keep pushing in this direction," he affirms.

“We have a few ideas for extending this work to panoptic segmentation while scaling to even larger scenes with standard hardware.”

Damien says that despite putting much effort into making the paper understandable, people sometimes struggle with the notion of mapping he describes. He advises them to look at their visualizations, in the paper and on [GitHub](), to learn more and to better understand the link between points and images.



Mathieu Salzmann is a Senior Researcher at EPFL in Lausanne, Switzerland, and an artificial intelligence engineer at ClearSpace, a start-up on a mission to make space sustainable.

Mathieu, you teach as well as do research.

Yes, I do research in the field of machine learning for computer vision, so for image analysis and visual understanding. I teach machine learning courses at the Bachelor level and for the Master in Digital Humanities program.

What do you do at ClearSpace?

ClearSpace aims to develop satellites that will go into space and capture debris in orbit around the Earth. There's a tremendous amount of debris in orbit around the Earth. It can be non-operational satellites or actual debris resulting from collisions, and as there's more and more of it, the chance of collisions gets higher and higher. We need to clean up the space around the Earth so that space operations remain doable.

Do you need machine learning and artificial intelligence for that, or is it more about aerospace engineering?

We need both. Of course, the start-up's core is in aerospace engineering and robotics because we need to design a robotics satellite that can capture debris. But to be able to grab the debris, we need computer vision to analyze images and understand where the debris is and its position relative to the capture satellite so that we can grab it.

I understand that someone cleaning the street is paid for by the local town, but who is paying for cleaning debris in space?

Currently, the funding we have is largely from the European Space Agency (ESA). We have a mission funded by the ESA that will take off in 2025/2026. Essentially, the European states decided to fund part of the program via the ESA. More and more, there are regulations, not official regulations so far, but the space agencies are strongly encouraging companies that are going to launch satellites, like the constellation companies, to have a way of cleaning up the mess they will make.

Are the countries that pollute the same ones that pay for the cleaning up, or is there some dissonance between the two?

No, I think this is reasonably well aligned. Of course, ESA is one of the big space agencies, so they pollute, and now they want to find a solution. But there are also initiatives like this in the US with NASA. I think it's pretty equal.

I just read that they found debris on Mars from a previous mission.

Yeah, there is debris everywhere. We're now focusing on the Earth's orbit but could go further in the long run.

So, you're not looking for Laika yet?

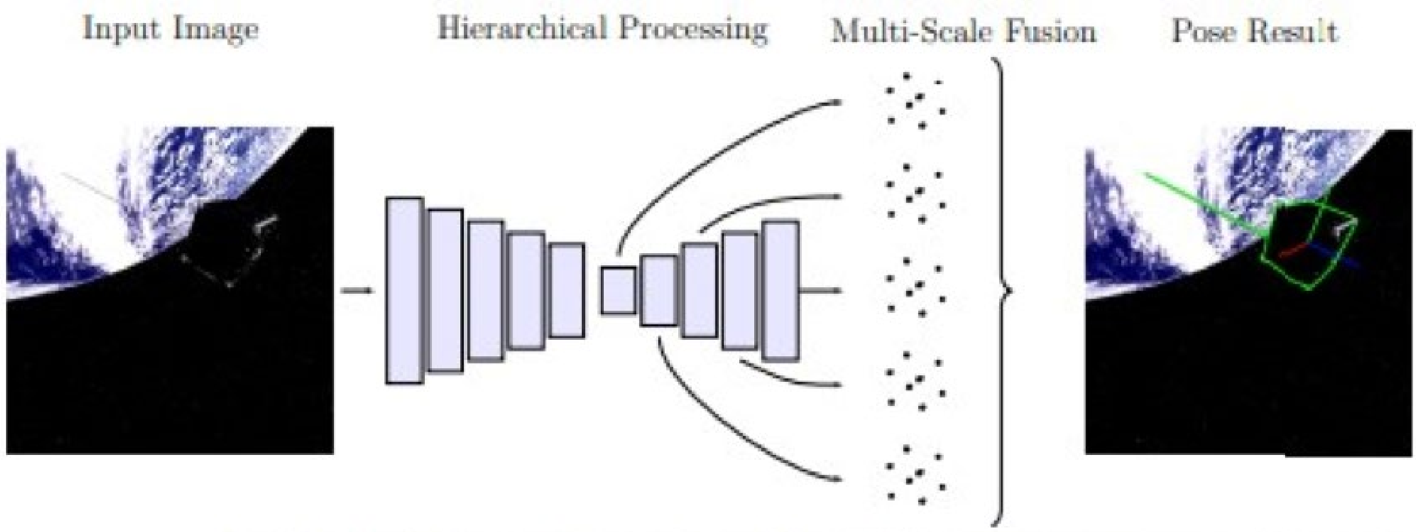
No, we're not going for Laika! *[he laughs]*

Where is computer vision most important in your activity?

Because we can't communicate in real-time with the satellites, we need to understand where the debris is, its relative position, and its orientation with respect to the capture system in an autonomous fashion. Computer vision comes into play here. We'll have cameras on board. From the cameras, we need to analyze the 6D pose automatically, so the rotation and translation of the debris with respect to the chaser.

How many are you?

For the computer vision part, we're five.



Our single-stage approach. We use an encoder-decoder architecture to progressively downsample the image and then to re-expand it. At each level of the decoder, we establish 3D-to-2D correspondences. Finally, we use a RANSAC-based PnP strategy [21] to infer a single reliable pose from these sets of correspondences.

Can you tell us more about the research side of your work?

As I said, there's a computer vision part at ClearSpace, but what we're doing at ClearSpace is quite applied. We need things to work three years from now, and even that's late. A whole world of research must be done to go beyond what we're focusing on now at ClearSpace, which is one specific debris. We know the shape of the debris, so we have everything at hand to capture it, but we need to generalize to completely arbitrary debris that can result from a collision and have any shape.

There's much work on how to generalize not only to new debris but to imaging conditions. We don't have images from space of the debris that we're trying to capture, so we need to be able to train our deep networks or whatever machine learning model we have on synthetic images, but they need to work in space.

The other aspect is that we don't have access to large clusters of GPUs in space. We have FPGAs and small CPUs. We need networks and machine learning models that work in real-time on minimal computer vision. There's the whole aspect of network compression, pruning, and quantization to make things work in real-time.

When did you graduate with your PhD?

I graduated in early 2009.

What has happened since then?

Many things! First, I went to do a postdoc at UC Berkeley with Trevor Darrell. I was there for a year and a half, working on 3D shape and non-rigid reconstruction. Then I moved to TTI in Chicago. I continued working on a similar topic. A bit of 3D human pose estimation as well. Then I moved to Australia. I had a senior researcher position there. That's where I started broadening my scope, working on Riemannian geometry for computer vision, visual recognition, segmentation, and many different problems.

Is that where you took the famous photograph of yourself with the cowboy hat?

Exactly! *[laughs]* It's made of kangaroo leather!

I was in Australia for three and a half years. My first daughter was born in Australia. My wife is Swiss, so we decided to move back to Switzerland to have the family around. That's when I had the opportunity to join CVLab with a permanent position at EPFL.

Tell us something about EPFL that we don't know.

The faculty that I'm in, the computer science and communication systems faculty, has an interesting program for PhD students. We have a way of recruiting PhD students such that they can get a fellowship for the first year. The labs don't need to fund them for the first year, which gives them the flexibility to try different labs. They can do a project in the lab and decide to change to something different for the second semester. That gives them a bit more freedom to choose in the end.

What are the main advantages for a young PhD student coming to EPFL?

It opens up many opportunities. One advantage that I mentioned before is the fellowship program. We have a strong computer vision presence at EPFL. Pascal Fua, Sabine Süsstrunk, Amir Zamir, a few other groups in some other faculties that do computer vision, and myself. A lot of machine learning as well. EPFL is a well-known school worldwide, meaning people get opportunities for faculty positions afterward, but also in companies. We don't have many big companies in Lausanne, but in Switzerland, particularly in Zurich, there are all the big ones.

Are Zurich and Lausanne connected enough to say I studied in Lausanne and I go to work in Zurich?

Yes, definitely. We've started building closer and closer connections with the people at ETH. We're organizing joint faculty meetings and retreats, and we're considering what we can do for summer schools, so we're trying to expand this a bit.

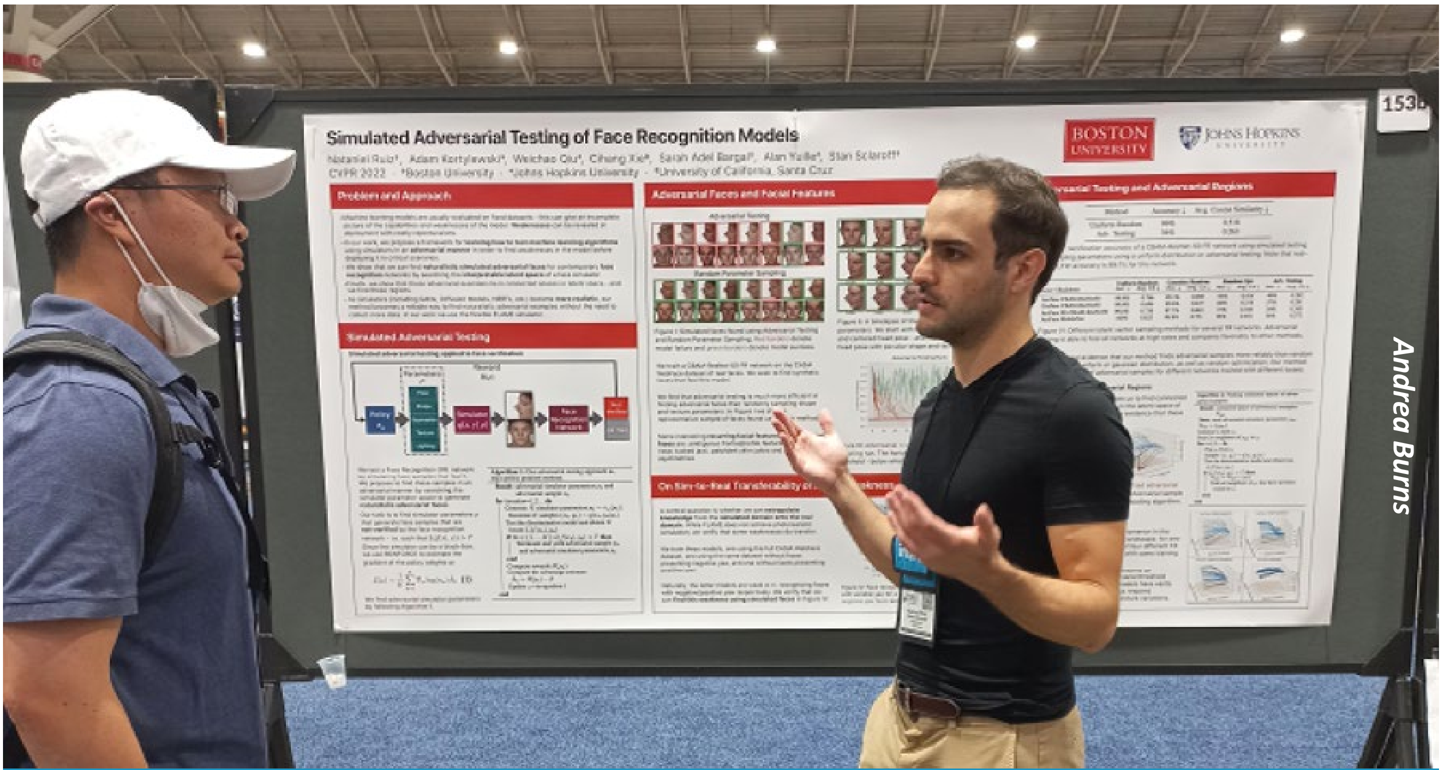
Is there anything that another university does that you are envious of?

That's an interesting question. Not really. I think we have a great PhD program. Of course, some schools are even better known than we are, but I think we're excellent in the model we follow.

Let's talk about CVPR. We are here in person for the first time in three years. Do you have a word for the community now that we're back together?

In the past two years, even though there's been a great effort to hold virtual events, I think they've been challenging for the community. They don't offer students the possibility to communicate their work and meet senior faculty. It's fantastic to be back here because it allows us to get back together, talk, exchange ideas, and show our work in a much more involved manner than we did in the past couple of years.

May we say to all those who stayed at home this year: please, next time, come over! It's much more fun!



Nataniel Ruiz, senior PhD student from Boston University, presenting his work on testing networks by generating simulated adversarial faces - a way to find interpretable weaknesses of models.

Cristian Canton
@cristiancanton

Cars at #CVPR22 this year are spectacular! Kudos to @Tesla @zoox @argoai @Cruise


HAPPINESS IS...
HOLDING TICKETS
THAT SAY
VANCOUVER
See you at CVPR 2023 in Vancouver.
Come Over. In-Person. It will be fun!

COMPUTER VISION EVENTS

MIDL 2022
Zürich, Switzerland
July 6-8

WBIR2022
Munich, Germany
10-12 July

Int. Conf. on Machine Learning & Data Mining
New York City, NY
16-21 July

ICML 2022
Machine Learning
Baltimore, MD
July 17-23

MIUA Medical Image Understanding and Analysis
Cambridge, UK
July 27-29

Summer School on Imaging for Medical Applications
Oradea, Romania
September 5-9

3DV
Prague, Czechia
September 12-16

TCT 2022
Boston, MA
September 16-19

MICCAI
MEET US THERE
Singapore
September 18-22

TechEx Europe
Amsterdam, The Netherlands
September 20-21

SUBSCRIBE!

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

FREE SUBSCRIPTION

(click here, its free)

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

Fill the Subscription Form
it takes less than 1 minute!

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

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

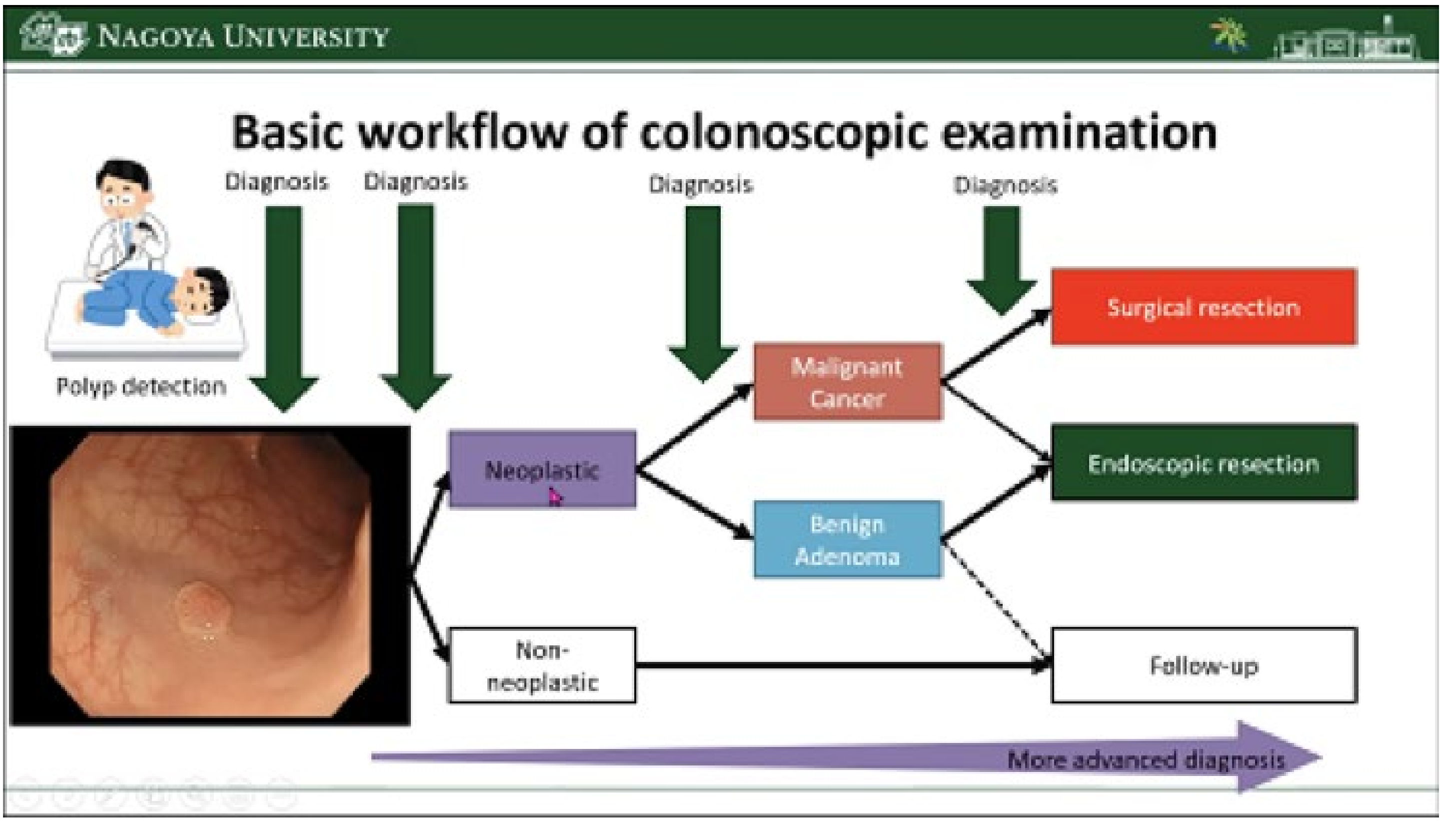
MEDICAL IMAGING NEWS

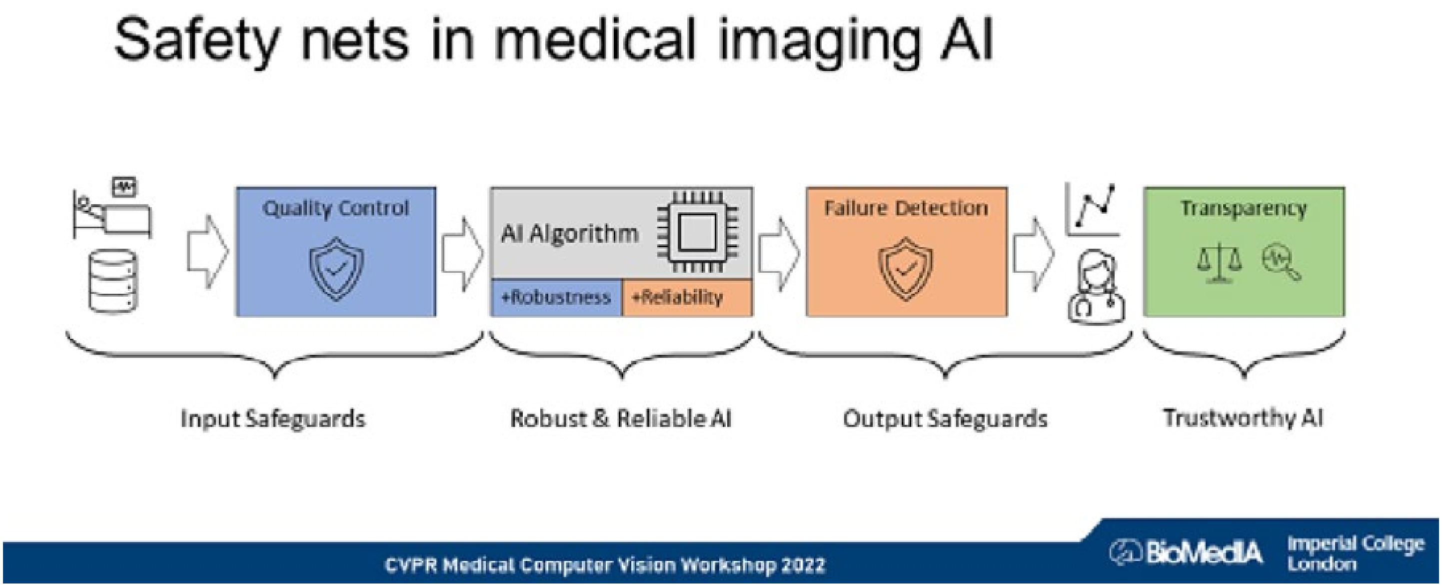
JULY 2022

The 9th edition of the CVPR Medical Computer Vision was held on Sunday in a hybrid format. The workshop is hosted by Mathias Unberath (Johns Hopkins University), Yuyin Zhou (University of California, Santa Cruz), Nicolas Padoy (University of Strasbourg, France), Tal Arbel (McGill University, Canada), Qi Dou (The Chinese University of Hong Kong), Vasileios Belagiannis (Otto von Guericke University Magdeburg).

The CVPR MCV workshop provides a unique forum for researchers and developers in academia, industry and healthcare to present, discuss and learn about cutting-edge advances in machine learning and computer vision for medical image analysis and computer-assisted interventions. The workshop aims to bring together stakeholders interested in leveraging medical imaging data, machine learning and computer vision algorithms to build the next generation of tools and products to advance image-based healthcare.

by Yuyin Zhou and Mathias Unberath





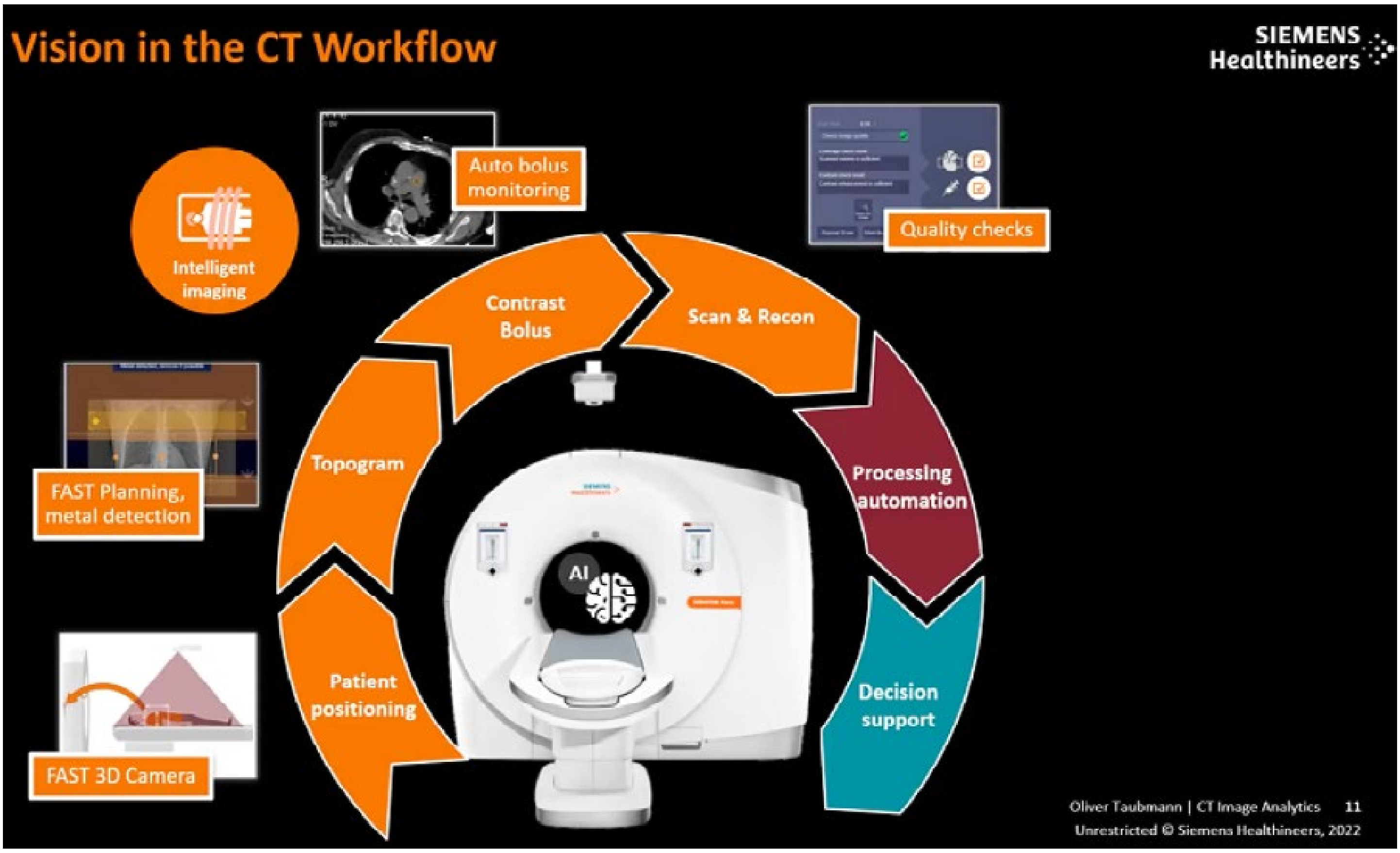
The 11 keynote presentations covered cutting-edge research topics from leading researchers from academia and industry, including Q&A sessions with both live and virtual audiences. A total of 60+ live attendees, as well as 50+ remote attendees, joined the event, engaging in intensive discussion regarding the technological advances, the latest progress, challenges, and future promises. The detailed schedule, highlights, and recordings can be found [here].

Kensaku Mori from **Nagoya University** kicked off the morning session by introducing **AI-based endoscopic procedure**. Specifically, he introduced multiple advanced AI techniques for several critical tasks to improve performing endoscopic procedures including blood vessel recognition on laparoscopic video, left gastric artery segmentation, depth estimation and shape recovery, and classification of laparoscopic images, among other tooics. Prof. Mori also shared valuable and unique insights on the current challenges as well as future trends.

Ender Konukoglu, from **ETH Zurich**, gave an overview of robust and trustworthy AI for medical imaging. He revealed a few key traits of trustworthy AI for medical imaging: 1) **robust** (adaptable systems robust to changes in input characteristics consistency), 2) **interpretable**, 3) **Self-aware** (identifying cases where confident predictions are not possible, 4) **Sensitive** (useful at realistic operating points that are applicable to clinical use cases).

Ben Glocker, from **Imperial College London**, introduced **"safety nets in medical imaging AI"**. Safety nets provide a layer of protection against AI failures. In medical imaging AI, we need to make sure that the use of AI is safe and that any predictions made by an algorithm are trustworthy. Especially, Ben discussed various safeguards, including automatic quality control, failure detection, and stress testing, and also delved into robustness and reliability in the context of mismatched data between the method development and clinical deployment. Further, the use of causal reasoning and how it can help identify potential biases concluded his talk.

In the first afternoon session, **Xiaoxiao Li** from the **University of British Columbia (UBC)** introduced different **advanced federated learning (FL) methods** to use more and diverse medical data under



privacy regulations, since AI models trained from small data sets are usually not accurate and generalizable. This new training paradigm enables multiple medical institutions to train a model collaboratively without data sharing. In this unique learning regime, Xiaoxiao's team has investigated novel optimization and learning schemes to tackle data heterogeneity, reduce dependency on data labeling, and adapt FL to different applications.

Ismail Ben Ayed from **École de Technologie Supérieur (ETS)**, introduced how to **fully leverage unlabeled data to enhance model generalization in a breadth of real scenarios and applications**. Especially, he talked about few-shot learning, unsupervised domain adaptation ad test-time adaptation as a few representative methods. He further introduced a series of latest works that use structure-driven / knowledge-driven / invariance / multi-modal priors to improve the above methods which achieve outstanding performances by utilizing unlabeled data.

Pablo Arbelaez from **Universidad de los Andes**, gave a wonderful presentation on **how to build fully autonomous surgical system to facilitate surgeons**. Especially, he talked about the importance of combining both long-term (surgical phases and steps) and short-term (instrument type, localization and atomic actions in frames) understanding for improving holistic surgical scene understanding. The PSI-AVA dataset offers a great opportunity for future researchers to delve deeper into this direction.

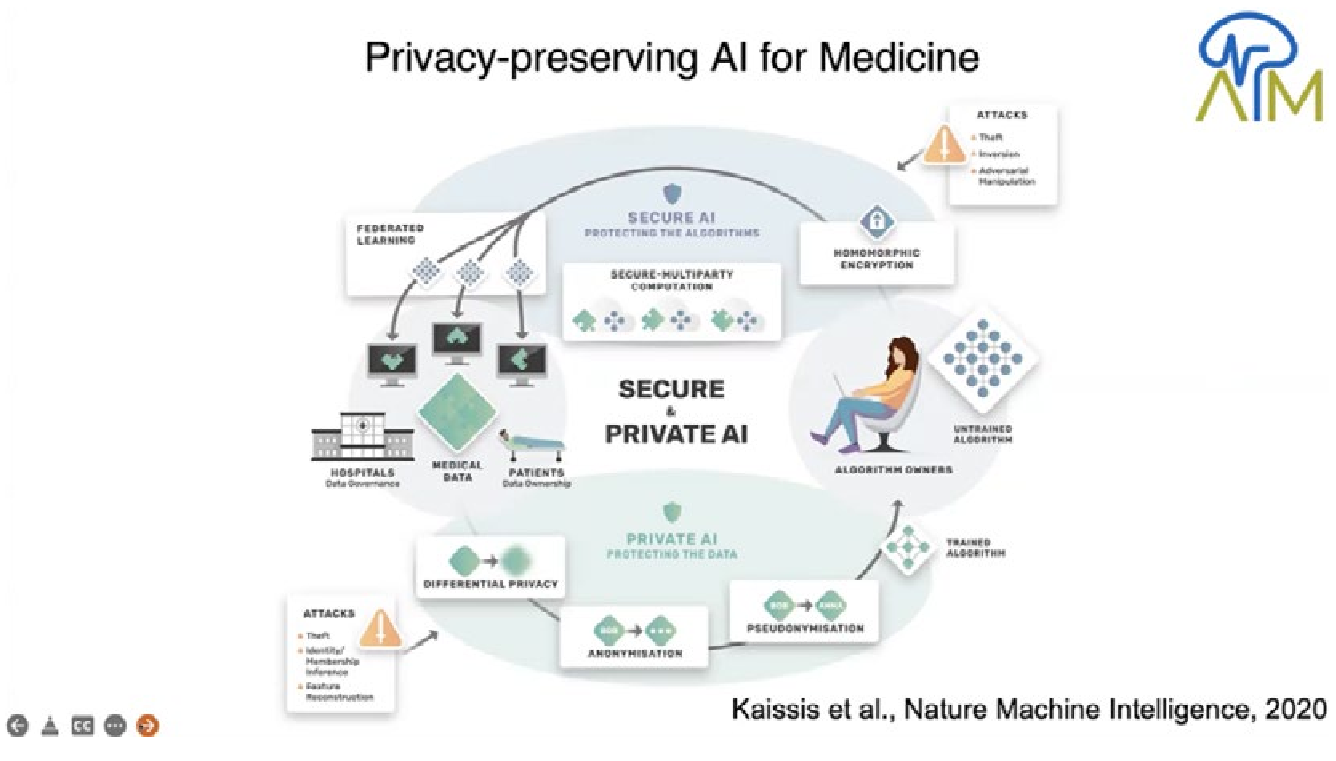
Mert Sabuncu from **Cornell Tech** is the last featured speaker for this session. His talk on **"Neural Encoding Models"** revealed great insights into the correlation between attention and eye gaze, which explains why visual attention plays a vital role in neural encoding. Further, Mert introduced his latest work on **individualizing encoding models** which suggest that neural encoding models need not be population-based, but can also be subject-specific. This finding can be quite useful for many clinical applications.

In the final session, **Oliver Taubmann** from **Siemens Healthineers** delivered the first talk on **“Vision in the CT Workflow”** from the industrial perspective, where he emphasized the importance of deploying explainable models such as counterfactual-based decision models for increasing the meaningfulness, intuitiveness, and enhancing imaging quality, which is crucial for real-world CT workflow.

Ulas Bagci from the **Northwestern University**, introduced how to build trustworthy AI for Imaging-based Diagnoses from the following three perspectives: 1) **algorithmic robustness**, 2) **interpretable ML methods** and 3) **expert-in-the-loop systems**. Especially, the combination of different aspects could offer the opportunity to build the most trustworthy medical AI system.

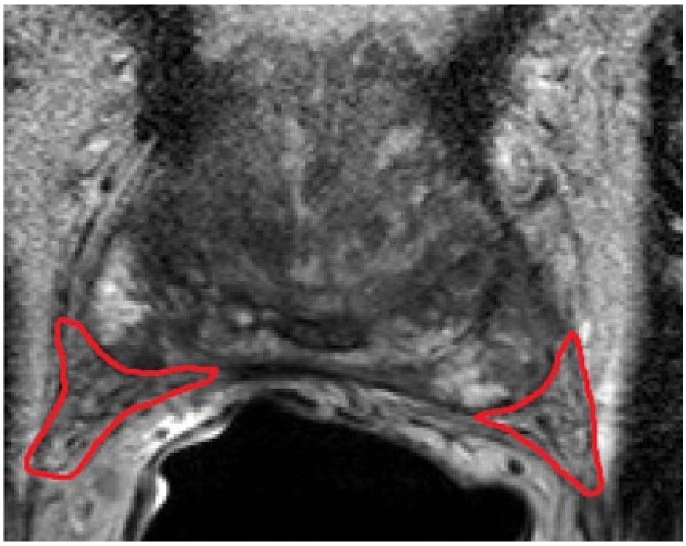
Jie Ying Wu from **Vanderbilt University** presented **“Modeling robotic surgery”** in person. With the increased adoption of surgical robots and their ability to record movements and events, Jie Ying introduced how to reconstruct entire surgeries. This, combined with recent advances in machine learning offers an unprecedented opportunity to improve surgical outcomes by modeling the interactions between each part.

Finally, Daniel Rückert from **TU Munich** gave a wonderful talk on **“Learning clinically useful information from medical images”**. In this talk, Daniel emphasized that current AI solutions are often brittle due to inadequate training data, and there are well-justified privacy concerns. Further, he elaborated on his latest works of building privacy-preserving AI by using federated learning to reduce the need of sharing data, and by **providing end-to-end privacy guarantees using cryptography and differential privacy methods**.



AI ASSISTED TISSUE SPARING IN UROLOGY

Tissue sparing is a common practice during surgeries. This approach aims to remove as little as possible of the surrounding tissue during a procedure. Studies have shown that tissue sparing procedures have fewer complications and faster recovery time.



In urology, **tissue sparing** is relevant for many surgical procedures. In **prostatectomy**, whether partial or radical, it is important to remove cancer cells without leaving any behind. However, intensive tissue resection may damage the neurovascular bundle (NVB), resulting in erectile dysfunction or incontinence. By sparing tissue during prostatectomy, the risk for NVB damage is reduced.

Female cystectomy is usually accompanied by **removal of surrounding organs**. This is performed to reduce risk of malignancy in the future. However, **preservation of gynecological organs** can assist in structural support of the remaining organs, as well as improve quality of life for the patient. This can be done by careful preoperative planning and accurate execution.

Tissue sparing can be achieved using the following techniques:

1. minimal incisions for surgical access

2. extreme precision for tumor removal

3. steady handling of surgical tools.

Below we will discuss how implementation of artificial intelligence (AI) and computer vision (CV) can improve these techniques and assist in tissue sparing.

(1) When attempting to minimize the entry port in a laparoscopic or open procedure, there are several challenges. The field of view (FOV) and the tool's operating range will be limited. When the FOV is limited, it is difficult to navigate the surgical scenery and recognize anatomical landmarks. Deep learning networks can be trained to detect specific landmarks and highlight them on the screen, assisting in recognition during the procedure. Additionally, utilizing the pre-op CT or MRI scan, a detailed procedure plan can be devised by segmenting and modeling the relevant anatomical structures, and the ideal incision position can be determined. This plan can be registered with intra-op imaging to verify incision location. Also, a full surgical plan can be devised using pre-op imaging, and this plan can be introduced into the surgeon's view using image registration. Developing such modules is done using a combination of classic computer vision algorithms and deep learning.

(2) Removing a tumor, radical or partial removal of an organ, requires precise dissection of the tissue. The tumor needs to be removed fully, with minimal damage to surrounding healthy tissue, vasculature, and nerves. Such precision can be enhanced by fusing pre-op images from CT or MRI with the tumor segmentation, with the intra-op video, ultrasound or X-ray images. The surgeon can view in real-time the surgical plan and perform accordingly. The pre-op segmentation can be easily obtained using neural-networks trained to segment tumors. The registration process can be achieved using deep learning techniques and classic computer vision, and further improved using existing or wearable landmarks designated for image registration.

(3) The above-described solutions assist in visualizing the tumor throughout the procedure. However, accuracy still depends on the tool-handling capabilities of the surgeon. To neutralize this effect, robots can be introduced. **Robotic assisted surgeries (RAS)** are becoming increasingly common. As expected, the robot's "hand" is more stable than the human hand, and can perform these surgeries with increased accuracy and precision. To improve robotic accuracy, distinct key points within the anatomy can be selected, and the tools' position relative to them can be calculated. This information provides real-time notifications of proximity to the tissue and warns against undesired resections. By using **tool tracking algorithms** - a well-established method which segments the tool in the field-of-view and utilizes prior knowledge of camera and tool characteristics to accurately position the tool in space - tool tracking is achieved. The robot can use this information to accurately maneuver the tool while avoiding unnecessary incisions. This can also be done using electro-magnetic (EM) tracking - a designated EM sensor is attached to

the tool and using an external EM field the tool's position is recorded continuously. Further developments in this field may also register the robot's coordinate system with the patient's, providing more accurate positioning relative to the anatomy. Now the surgeon can approach the surgical procedure with **high accuracy and stability**, ensuring minimal surrounding damage, and sparing healthy tissue.

Combining these methodologies can aid in tissue sparing during urologic procedures. A smaller port in an ideal position can be achieved without compromising the anatomical understanding of the surgical scene, the surgical target can be accurately viewed in real-time, and introduction of RAS will increase surgical accuracy, ultimately leading to less damage to healthy tissue. These solutions are challenging to implement, and advanced knowledge in AI and computer vision is essential for developing them. **RSIP Vision** has vast experience in developing computer vision solutions. Contact us for a speedy development process and faster time-to-market.

Vision Transformers in Medical Computer Vision

IOANNIS VALASAKIS, KING'S COLLEGE LONDON

@WIZOFE

Hello again! I hope that this month you enjoy the coverage of the most important conference in Computer Vision, if so, Ralph has you covered! Meanwhile, don't forget medical imaging. In this month, I will present you a paper from Laith Alzubaidi et al, from Queensland University in Brisbane, Australia who kindly provided permission to use parts of the paper.

As a reminder, it is always important to read the original paper and get your own appreciation of what is included there. I hope this month's article will make this journey easier!

Introduction

Deep learning (DL) methods, as you probably already know, allow computational models composed of multiple processing layers to learn representations of data with multiple levels of abstraction. DL methods have been extensively used and achieved improved results compared to previous research in many fields, such as visual object recognition, speech recognition and object detection. It is known that the performance of DL depends on a huge amount of data for representation learning.

For medical images, such as ophthalmology or pathology slides, the available data are often not sufficient to obtain a good model by using **ImageNet dataset**. Moreover, medical image analysis struggles from a lack of sufficient data for training DL models. Transfer learning helps to obtain accuracy for multimedia tasks. It was recently proven to help pre-trained models with medical grey-scale images, but as medical imaging tasks are becoming more prevalent, transfer learning is also becoming more difficult.

This will improve the performance of machine learning algorithms when used on limited-data clinical problems, such as that of diseases diagnosis and prediction. In this paper you'll discover a proposed new model for addressing the previous shortcomings of medical image classification. The model is called **MedNet** and is two-fold: a **Gray-MedNet** and a **Color-MedNet**. Both can be trained using a publicly available 3M medical image dataset of each version coming from multiple sources: MRI, CT, X-Ray, and PET, among many others.

MedNet and perspective

The lack of training data is a problem with modern deep CNNs. The common solution in state-of-the-art (SoTA) is to use **transfer learning**. Transfer learning relies on the knowledge of a specific task and uses it to address both the target task and other correlated tasks with the use of previous knowledge. Transfer learning is also a form of single-task learning.

Transfer learning is an extremely useful tool that can help in the classification of images in a timely manner. This is because it is a large amount of labelled data which is extremely difficult to collect. Transfer learning also has high accuracy in this task.

Deep CNNs are effective for image recognition and classification tasks, but usually require large datasets to prevent over-fitting and generalize properly. Using small datasets or a deep CNN with fewer layers would not be highly effective due to under or over-fitting issues. Then, models with fewer layers are less accurate because they are unable to benefit from the increased hierarchical features of large datasets. In the field of medical image analysis, the lack of available data is a problem, but collecting the data is extremely costly.

Recently many researchers have demonstrated that the use of translation likelihood in deep learning models is effective and efficient. However, to solve complicated issues in deep learning models, one must have a large amount of data. Knowledge about the image is crucial for these models. Furthermore, because of the high cost and low accuracy of human-provided annotations, these models often depend on the experts for accurate annotation. To produce more accurate and cost-effective models, this becomes unrealistic.

Most researchers currently use data augmentation, a technique which creates new data. Although data augmentation enhances the data, current CNN models still have issues with overfitting. Transfer learning is a technique in which DCNN models are trained on a targeted dataset, but there is still a fundamental challenge, which is the difference between the source and target datasets. In general, DCNN models trained on ImageNet, i.e., made up of mostly natural images, are used to enhance medical image classification tasks. Additionally, ImageNet is different than other common medical imaging datasets that has a negative impact on the performance of the result. It is believed that different domain Transfer learning does not significantly affect the performance on medical imaging tasks like lightweight models trained from scratch perform well in comparison to those trained using ImageNet. There is a clear lack of a pre-trained model on medical images to help in learning, representation, and generalization. This problem will be solved with the two versions of MedNet introduced in the next section.

Architecture and testing

The MedNet DCNN model is not just one model but an amalgamation of features, features, features.

Acknowledging that networks with smaller input images have lower per-image computation, the authors introduce a new layer which scales images.

The text is about how to train a machine learning algorithm and how perform a good implementation. It is about how to train a model and how to use them to make decisions.

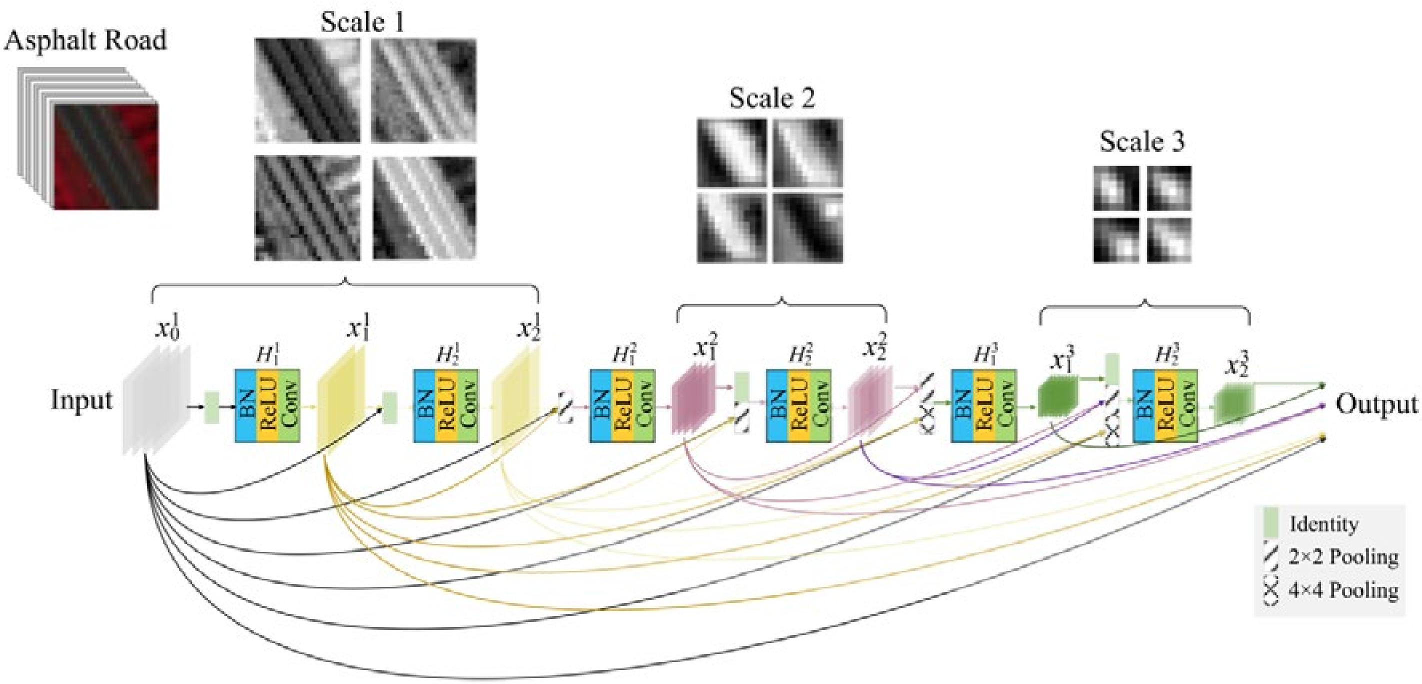
The idea of the specific introduction is to show how training deep learning models can be used for semantic representation. It shows without residual connections can fail to learn features correctly and how to remedy the issue. But let’s now analyse the specific elements of the proposed deep learning network.

Batch Normalization (BN) is used to expedite the training. It is a method used to make training of artificial neural networks faster and more stable through normalization of the layers' inputs by re-centering and re-scaling. A rectified linear unit does not squeeze the input value, which helps minimize the effect of the vanishing gradient problem. Dropout is a strategy often used with supervised learning algorithms. Specifically, it is a regularization technique for reducing overfitting in artificial neural networks by preventing complex co-adaptations on training data.

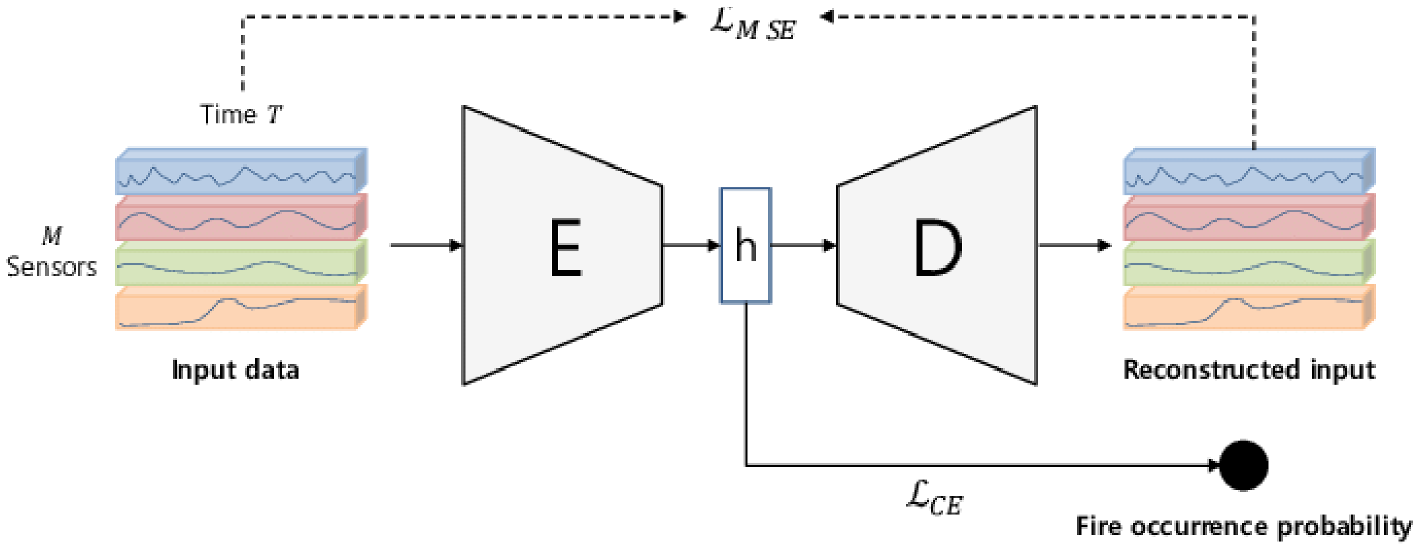
Global average pooling makes an extreme dimensionality reduction by transforming the entire size to one dimension, helping to reduce the effect of over-fitting.

The model starts with a traditional convolutional layer followed by a BN layer and **ReLU layer**. The previous layer is followed by eight blocks of parallel convolutional layers that have four distinct filter sizes (1x1, 3x3, 5x5, and 7x7). The output of these four layers is concatenated and integrated into the following block to create the final model. The convolutional layers in MedNet are followed by BN and ReLU layers.

Just a reminder here for different applications, such as remote sensing there has been propose a fully connected dense connectivity pattern layer shown below



Back to the MedNet though. There are twelve connections between the blocks of the convolutional layers. These connections maintain the ability of the model to maintain different levels of features for the purpose of achieving a better representation of them. Both parallel convolutions and the connections are extremely important for gradient propagation



as the error can back-propagate from multiple paths. Finally, two fully connected layers are adopted with one dropout layer between them. SoftMax is employed to finalize the output. In total, MedNet consists of 44 convolutional layers.

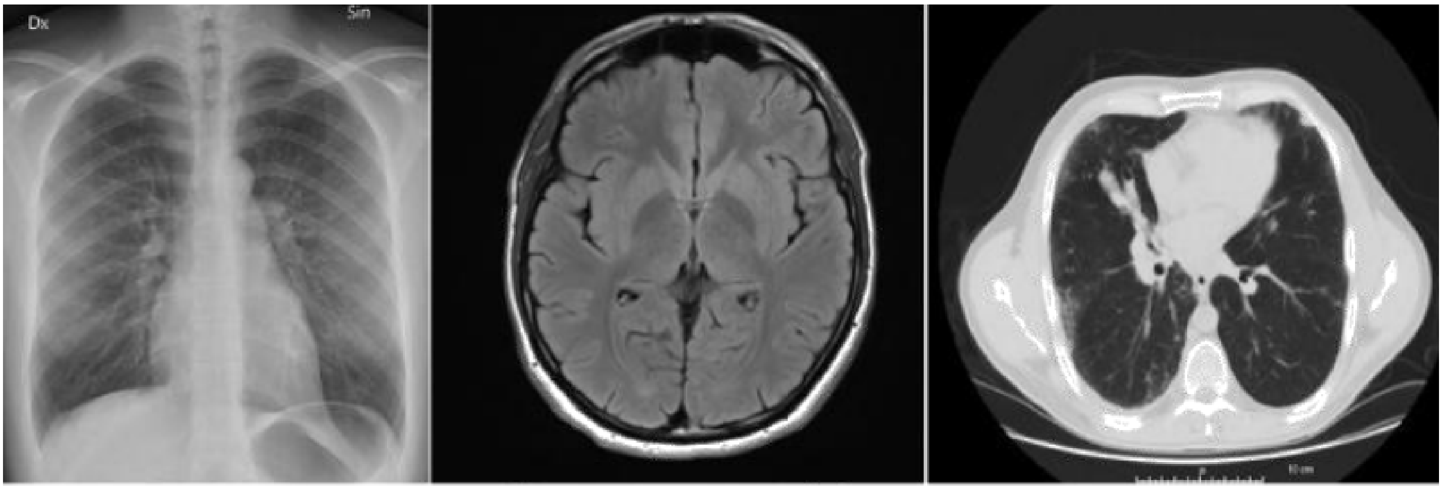
Previous networks are as important, as it can be seen in the figure below on a very high-level description (not including the individual layers).

The proposed is a novel technique for translation using transfer learning to overcome the issues of translation from a pre-trained model of ImageNet. The technique is used in medical imaging applications and helps to address the lack of training data. The approach is to perform training on a model called Gray-MedNet with publicly available data in different medical fields, including CT, MRI, PET, histology, and so on. It is also based on training a model called Color-MedNet with publicly available data in different medical fields, including CT, MRI, PET, histology, and so on.

These datasets include: CT images (abdomen, bladder, brain, chest, kidney, cervix, breast), MRI (neuroimaging, cardiovascular, liver, functional, oncology, phase contrast), PET (cardiology, infected tissues, neuroimaging, oncology, musculoskeletal, pharmacokinetics), histology (epithelium, endothelium, mesenchyme, blood cells, neurons, germ cells, placenta), X-Rays (radiography, mammography, fluoroscopy, contrast radiography, arthrography, discography, dexa Scan), Ultrasound (breast, doppler, abdominal, transabdominal, cranial, spleen, transrectal), ophthalmic fundus images, corneal topographic maps, skin cancer images, and many more.

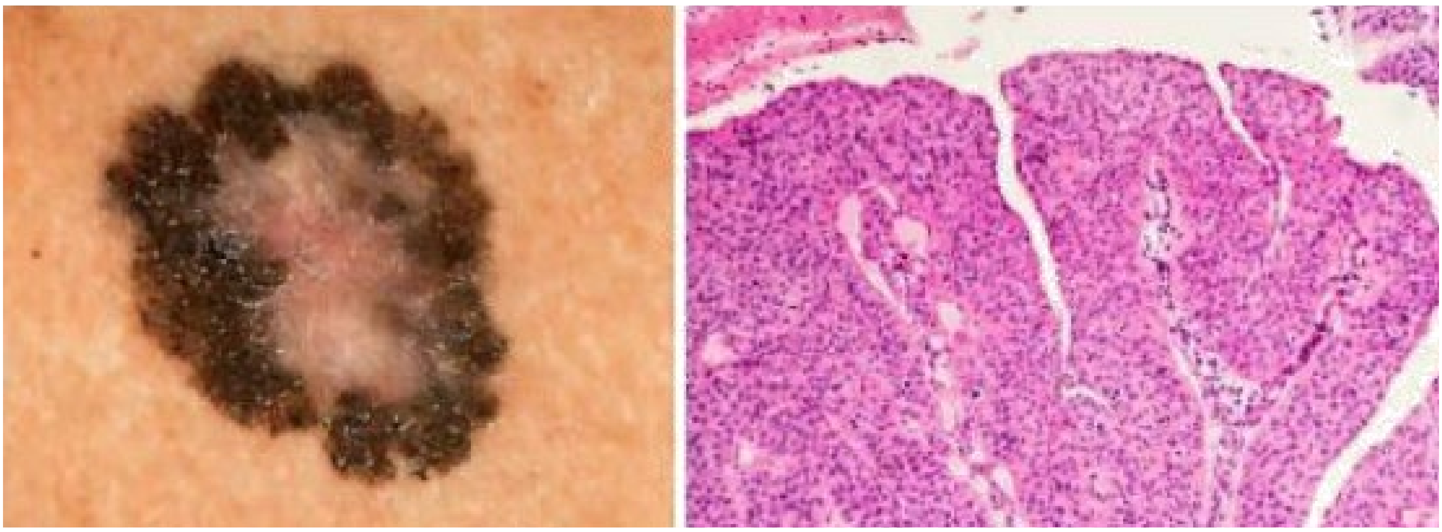
The first step is to take data-augmentation techniques and use them to address the imbalance issue, followed using generative adversarial networks to train the models. This is done through the collection of data and filtering of information, followed.

First, MedNet is trained and validated on five small target datasets to check the effectiveness of MedNet. One important aspect, considering open science, which those days it becomes more and more prominent is that MedNet and its source code will be available in the research community to be specifically trained for a particular application.



The following is the description of a single experiment on two different medical image applications.

Two different challenging medical imaging scenarios were used to study how transfer learning from medical images benefits supervised learning. In both cases, they found that transfer learning can improve the learning stage of the model and generalize better in a more suitable manner. Transfer learning proved more beneficial than the pre-trained models of ImageNet. They went on to implement different experiments for other medical applications such as the MedNet idea.



Medical-imaging applications fail when moving from ImageNet to clinical tests due to the following problems: a small training set and generalization issues, and over-fitting. By using MedNet, they can easily notice those prominent effects and improve performance.

Overall, MedNet as described and analysed earlier has the following advantages: a high performance, a small training set and generalization issues, and prevention of over-fitting. A question for all of you: are there applications where MedNet will be useful to you, and if so, which are they? Let us know!

Next month

Thank you for reading the article this month about a medical deep learning layered network. We hope that you enjoyed it and don't hesitate to send our way any corrections, suggestions, or ideas for next month's issue!

Take care and as always have a great time and always be curious! ☺

INTRODUCTION TO DOCKER

By Marica Muffoletto (twitter)

Hello everyone, welcome back to Computer Vision News. After a full month dedicated to CVPR, we are coming back with a handful of articles, including a very special one to me. Why special you may ask? Because it's one that I've been wanting to write for a while and a topic I find extremely useful. So, let's start ☺

Today we are going to talk a bit about Docker, a software framework that helps with packaging and distributing our applications, hence it facilitates the transition between development and production stages, or just the process of sharing our code with other fellow researchers or collaborators. And who does not believe this is as critical as ever in this decade? In the Medical Imaging community, for example, Docker is already used for the MICCAI conference challenges to allow users to upload their code and let it be tested without incurring into hardware issues on the reviewers' end. This is a typical case study that we are going to focus on in this article.

Let's start with the installation on Linux terminal. Assuming no previous versions of Docker have been installed on the machine and access to the Internet is granted, we can install directly from the Docker's package repository.

```
sudo apt-get update
sudo apt-get install ca-certificates curl gnupg lsb-release
sudo mkdir -p /etc/apt/keyrings

sudo curl -fsSL https://download.docker.com/linux/ubuntu/gpg |

sudo gpg --dearmor -o /etc/apt/keyrings/docker.gpg

echo  "deb [arch=$(dpkg --print-architecture) signed-by=/etc/apt/keyrings/docker.gpg] https://download.docker.com/linux/ubuntu \
  $(lsb_release -cs) stable" | sudo tee /etc/apt/sources.list.d/docker.list > /dev/null

sudo apt-get install docker-ce docker-ce-cli containerd.io docker-compose-plugin

apt-cache madison docker-ce
```

The last command will list the available versions in the repository. Alternatively, installing from a deb package for Ubuntu is also possible. For Windows and Mac installation, you can check the original manual from Docker, but you might need to install Docker Desktop instead, which offers a GUI and a virtual machine inside which Docker is running.

To install a specific version, run the command below substituting the desired string for docker-ce and docker-ce-cli. I am installing `5:20.10.16~3-0~ubuntu-bionic`.

```
sudo apt-get install docker-ce=5:20.10.16~3-0~ubuntu-bionic docker-ce-cli=5:20.10.16~3-0~ubuntu-bionic containerd.io docker-compose-plugin
```

To verify that docker is installed correctly, we can try the following command: `sudo docker run hello-world`

Now that we are positive that the installation worked, we can get familiar with the basics of Docker. Docker allows users to package an application with all its dependencies into a standardized unit. Seems like a dream, right? A background service (called Docker Daemon) is running on the host to build, run and distribute Containers. These are a packaged ensemble of code and dependencies that “activate” when a Docker Image - a saved blueprint of an application - is run on the Docker Engine. Inside a container, the application, which was previously uploaded through an image, can be safely run. Images can be pushed and pulled locally, in a way that Git users are probably already very familiar with.

Finally, the Docker Client is the command line tool that allows the user to interact with the Daemon.

Now let’s try to use it! We are going to package an application made of a python script, a test folder containing .png images, and a .h5 file.

The goal is to use a pretrained classification model to predict the type of retinal disease in fundus images.

The inference model we are looking to share with other uses is written in Tensorflow keras and looks like this:

```
from keras.preprocessing import image
from tensorflow import keras
from PIL import Image
import numpy as np
import matplotlib.pyplot as plt
import os

#Function to pre-process new images
def data_resizing_process(dirs,img_width,img_height):
    datas = keras.utils.load_img(dirs, target_size = (img_width, img_height))
    res = keras.utils.img_to_array(datas)
    res = np.expand_dims(res, axis = 0)
    res /= 255
    return res

#Inference function
def infer():

    img_width=300
    img_height=450
```

```
    TESTING_DIR = os.path.join(os.environ["BASE_DIR"],'Test')
    print(TESTING_DIR)
    MODEL_DIR = os.environ["BASE_DIR"]
    print(MODEL_DIR)
    OUTPUT_DIR = os.path.join(os.environ["BASE_DIR"], 'Output')
    # List of classification labels
    classes=['DR','MH','DN','TSLN','ODC']
    list_of_images = os.listdir(TESTING_DIR)
    # Import pre-trained model
    model = keras.models.load_model(MODEL_DIR+'/multi_CNN.h5')

    # Run model through test images and save result
    for img_filename in list_of_images:
           img = data_resizing_process(os.path.join(TESTING_DIR,img_
filename),img_width,img_height)
        result = model.predict(img)
        print(np.argmax(result))
        plt.imshow(Image.open(os.path.join(TESTING_DIR,img_filename)))
        plt.title(classes[np.argmax(result)])
        plt.savefig(os.path.join(OUTPUT_DIR,img_filename.split('/')[-1]))
        print(os.path.join(OUTPUT_DIR,img_filename.split('/')[-1]))
        #plt.show()

if __name__ == '__main__':

    infer()
```

The next step to package our application is to create a Dockerfile, a simple text file which contains indications on what commands need to be executed in the docker environment for building an image. An important thing to note about Dockerfiles is that, besides running usual terminal commands such as *mkdir* or *pip install,* they can call other images (with the keyword FROM) which already contain sets of libraries. Moreover, they usually include commands to copy the necessary files into the new container.

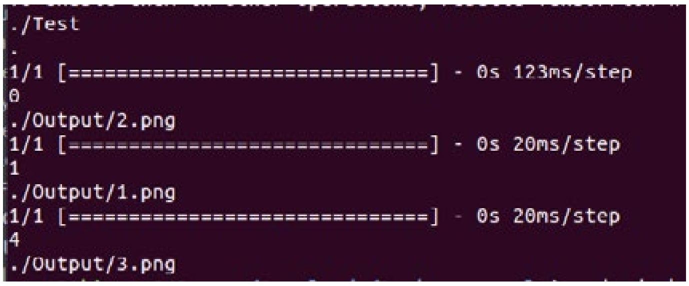
```
FROM python:latest

RUN pip3 install Keras
RUN pip3 install numpy
RUN pip3 install Pillow
RUN pip3 install matplotlib
RUN pip3 install tensorflow

RUN mkdir ./Output
ENV BASE_DIR=.

COPY infer.py ./infer.py
COPY multi_CNN.h5 ./multi_CNN.h5
COPY Test ./Test
```

We can now run the commands below to 1) build an image which contains our inference model, 2) run the container to produce and save the results. If everything has worked correctly, we should see something like the following message on our terminal window:




```
docker build -t docker-dl-example -f Dockerfile .
docker run docker-dl-example python3 infer.py
```

After signing up to **Docker Hub**, it's also possible to create a repository and upload an image. This is a registry where all available images are stored and can be pulled by other users. It's also the place where you can search for images to write in your Dockerfile.

Other basic commands to start with Docker:

`docker images` -> to list the images on your local machine

`docker ps -a` -> to list the containers running on your local machine

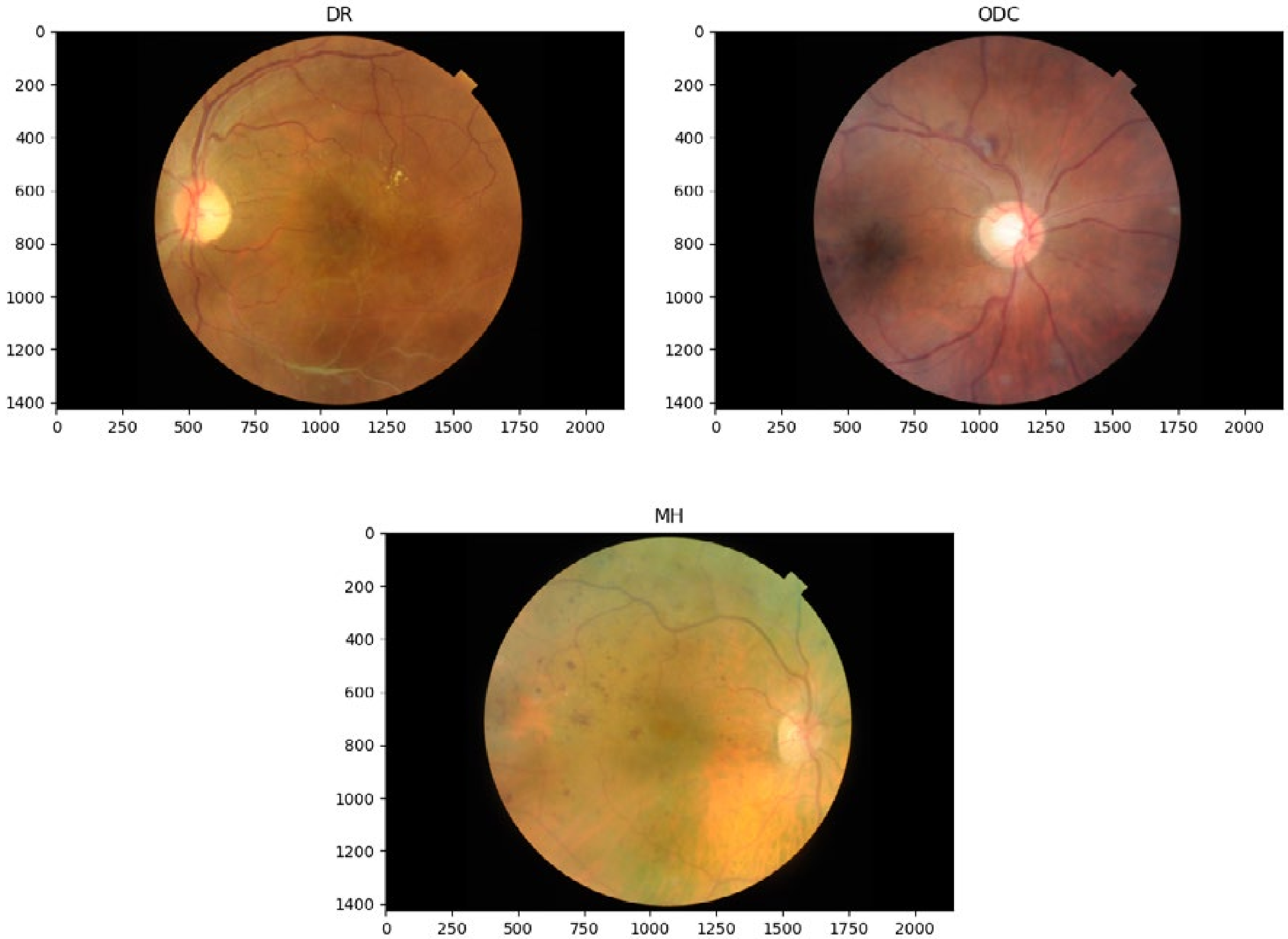
`docker rmi $ID` -> to remove an image (provided one knows the ID)

`docker rm $ID` -> to remove a container (provided one knows the ID)

After your image is shared, other people can pull it and run it with the following commands:

```
docker pull username/repository:tag

#to run
docker run username/repository:tag
```



In our case, to check the obtained predictions, we need to first copy the output folder from the docker container into a local directory:

`docker ps -a` -> to read the container ID for the image

`docker cp $CONTAINER ID:/Output ~/$local _ dir`

This article is a short introduction to Docker and shows the initial steps to set it up and use it for a medical imaging application. This tool shows great flexibility, from deploying web applications to deploying an app to sharing deep learning projects and let it run by many.

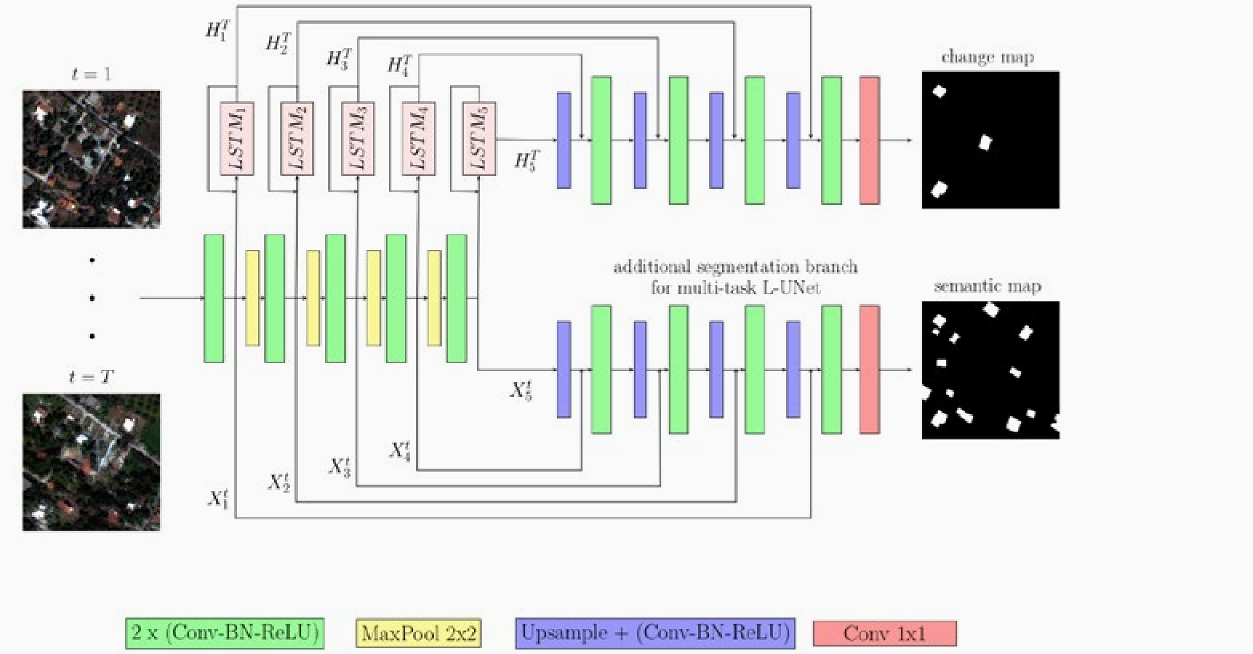
And, as in the writer’s opinion, after spending countless hours setting up new machines and conda environments, I would jump at any alternative to make this process less painful for developers and end users. Hence, I highly recommend looking into Docker. It might be your chance to reduce the risk of conflict between languages, libraries, or frameworks and run your models smoothly.

Maria Papadomanolaki recently completed her PhD at the Remote Sensing Lab at the National Technical University of Athens, collaborating also with the Mathématiques et Informatique pour la Complexité et les Systèmes (MICS) Lab in CentraleSupélec, Université Paris-Saclay. Her research focused mainly on the topics of image registration and urban change detection using satellite images with deep learning methods. Maria continues her research as a Postdoctoral Scholar in CentraleSupélec at the MICS Laboratory working on the medical imaging field following the methodological advancements on computational pathology. Congrats, Doctor Maria!

The unlimited access to earth observation data has resulted in the development of powerful algorithms able to survey the planet. Still, problems like **misregistration, intraclass variation** as well as **complex spatial and spectral distributions of ground objects** continue to hinder the formulation of robust and generic algorithms. To this end, our research focused on **creating an effective multi-task change detection algorithm based on sequential data (in our case time-series images)** that takes full advantage of the temporal relationship between them. At the same time, a multi-step deformable image registration method was developed, able to handle not only the global affine displacements but also the more complex local pixelwise deformations caused by the different viewpoints of the sensors. All our methods were examined proposing **novel data-driven deep learning methods on high and very high resolution satellite images**.

Giving some more details on the change detection topic, the proposed algorithm [Figure1] is a **deep multi-task learning framework** able to couple semantic segmentation and change detection using fully convolutional **Long Short-Term Memory (LSTM) networks**. In particular, we present a UNet-like architecture which models the temporal relationship of spatial feature representations using integrated fully convolutional LSTM blocks on top of every encoding level. In this way, the network is able to capture the temporal relationship of spatial feature vectors in all encoding levels without the need to downsample or flatten them, forming an end-to-end trainable framework. Moreover, we further enrich this architecture with an additional decoding branch that performs

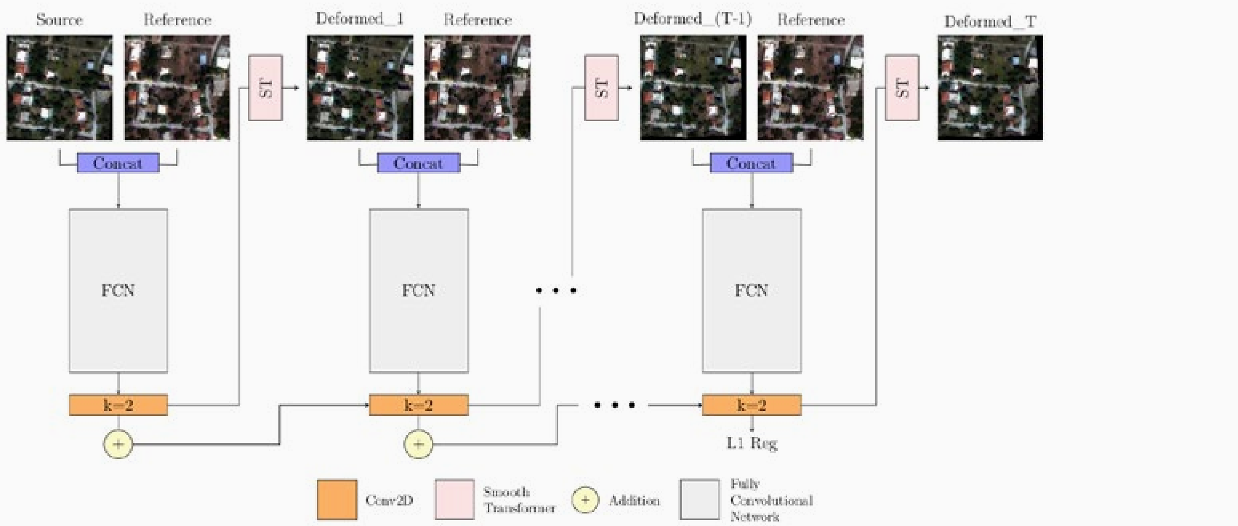
[Figure1]



semantic segmentation on the available semantic categories that are presented in the different input dates, forming a multi-task framework.

Concerning the image registration task, the developed method is a **multi-step deformable registration scheme based on the expression power of deep fully convolutional networks**, regressing directly the spatial gradients of the deformation and employing a 2D smooth transformer layer to efficiently warp one image to the other, in an end- to-end fashion. The displacements are calculated in an iterative way, utilizing different time steps to refine and regress them [Figure 2]. In addition, information about the images' edges is integrated during the training process, contributing to better image alignment and preservation of shapes. To handle properly the areas where there are no correspondences, a prior on the change regions is also added, guiding the model to relax the registration constraints in the areas of change. Our formulation can be integrated into **any kind of fully convolutional architecture, providing at the same time fast inference performances**.

[Figure2]



RSIP VISION
Visual Intelligence
for MedTech

dkfz.
GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

BAY VISION VIRTUAL
MEETUP

Machine Learning in Medical Imaging: Current Challenges

Despite its vast potential, the actual practice-changing clinical impact of machine learning in medical imaging has so far been rather modest. Why is that? The talk covers several major challenges that Klaus considers essential in unlocking the full potential of machine learning in medical imaging, and he presents current examples of ongoing research that address them.

GUEST SPEAKER

Klaus Maier-Hein

Professor at Heidelberg University. Managing Director of Data Science and Digital Oncology at the German Cancer Research Center (DKFZ).

HOSTED BY

Moshe Safran

CEO of RSIP Vision USA.

Defining & developing innovative Medical Visual Intelligence solutions, in partnership with MedTech industry leaders.

Missed the Meetup?

Don’t miss the video!

Be sure not to miss next time :-)

RSIP VISION
Visual Intelligence for MedTech

UCL

BAY VISION VIRTUAL
MEETUP

AI-Assisted Surgery for Next Generation Interventions

Recent trends in AI and surgical data science have shown promising technical advancements in imaging, surgical navigation and robotic intervention. This talk will highlight AI applications in various surgical procedures and where we stand in terms of their clinical translation as we head towards the next generation of surgical interventions.

JULY 27th 2022

9:00 am PT | 12:00 pm ET

GUEST SPEAKER

Sophia Bano

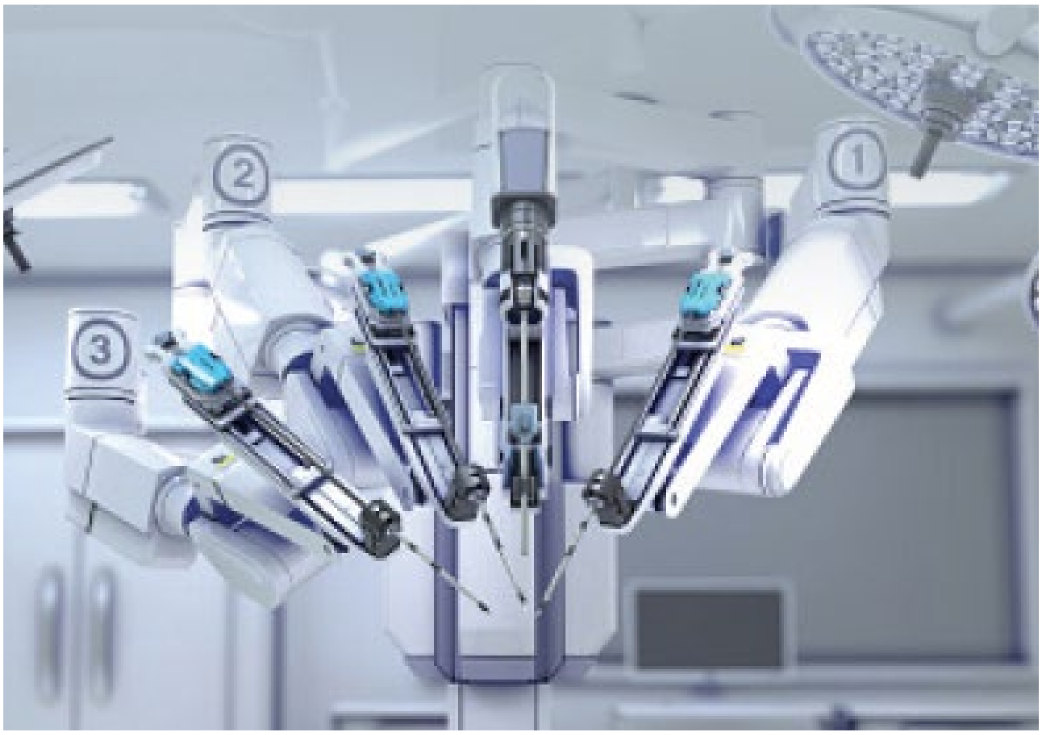
Senior Research Fellow at the Surgical Robot Vision Research Group, Wellcome / EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London.

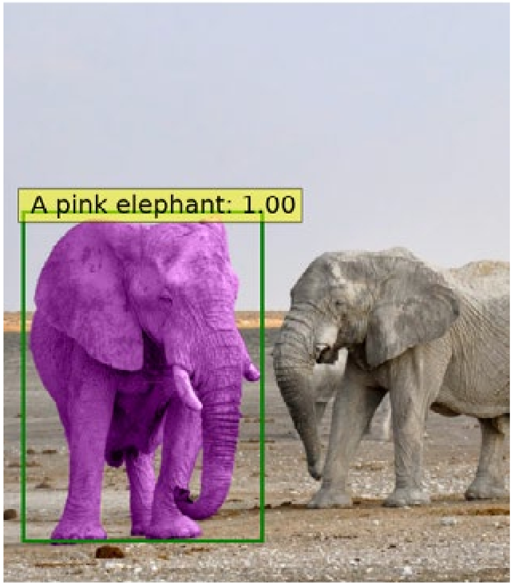
HOSTED BY

Moshe Safran

CEO of RSIP Vision USA.

Defining & developing innovative Medical Visual Intelligence solutions, in partnership with MedTech industry leaders.




A pink elephant: 1.00