OCTOBER 2022

Computer Vision News
& Medical Imaging News

The Magazine of the Algorithm Community

RSIP VISION
Visual Intelligence for MedTech

DILBERT

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

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editors:
Marica Muffoletto
Ioannis Valasakis

Publisher:
RSIP Vision

Contact us

Free subscription

Read Previous Magazines

Follow Us

Dear reader,

We're still buzzing from the runaway success of **MICCAI 2022 in Singapore** last week. A truly awesome event that reunited the community after three years apart and brought the conference to Southeast Asia for the very first time.

After two virtual meetings, **the enthusiasm for being back together was palpable**, and the standard of scientific research and technological innovation on show was out of this world. Luckily for you, we've rounded up some of the finest examples in our BEST OF MICCAI section this month. Don't miss it on page 20!

We're eagerly anticipating MICCAI 2023 and CVPR 2023, both to be held in Vancouver. Exciting plans are already well underway for both. There's so much to look forward to, and we hope to see even more of you at future events!

This time last year, we launched our new supplement, Medical Imaging News, and were blown away by the response. It has amassed a dedicated following, and we think you'll agree that it's seen 12 months of great content. Long may it continue!

This month in Computer Vision News, we continue our policy of reporting on celebrated papers, reviewing the **Best Paper from SIGGRAPH 2022**. Meet the impressive **Yael Vinker**, a researcher and artist who proposes a fascinating method for **converting images into sketches**.

Enjoy this issue, and don't keep us a secret – remember to tell all your friends and colleagues about us so they can subscribe for free too!.

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

THIS HAPPENED IN SINGAPORE


HAPPINESS IS...
HOLDING TICKETS
THAT SAY
VANCOUVER!
CVPR and MICCAI 2023 in Vancouver.
Come over. In-person. It will be fun!

CLIPASSO: SEMANTICALLY AWARE OBJECT SKETCHING

Yael Vinker is a PhD computer science student at Tel Aviv University, under the supervision of Daniel Cohen-Or and Arik Shamir, with a particular interest in the intersection of art and technology. Yael's work, proposing a new method for converting images into sketches, scooped a Best Paper award at SIGGRAPH 2022, and she is here to tell us all about it.

While many papers have explored **image-to-sketch translation**, Yael's work stands out as it recognizes the importance of abstraction and the concept of **multiple levels of abstraction**. Abstraction in art involves identifying the fundamental visual properties of an object or scene. At the heart of sketching, artists must decide on the essential features to include in simple and minimal line drawings.

"You can't offer a computational method for image-to-sketch translation and ignore abstraction and this main idea of different levels of abstraction," Yael tells us.

"If I ask you to draw a cat, you will draw it differently than me, an artist, or someone else. There is no single solution we can learn for sketches; rather, there are many correct solutions."

"The Bull" by Picasso - this is a good example of progressive abstraction, this also emphasizes that this task is not trivial, even for artists, it's hard to make the right choices such that the abstract object remains recognizable and also visually appealing.

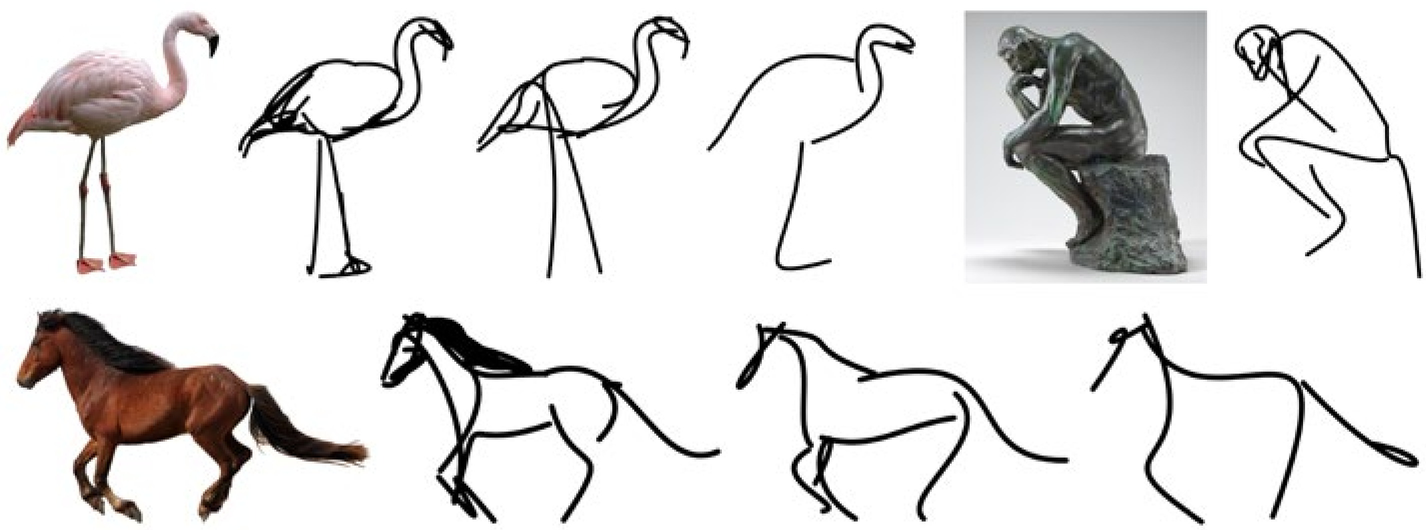
Another important concept in the paper is that the sketches are **semantically aware**. When asked to draw something, we first analyze and understand what we see. If we were drawing an image of a cat, we know the triangular shape of cats' ears and that they have whiskers. These are probably the first features we would draw, even if they do not necessarily relate to the image of the cat we are sketching. This work takes inspiration from this prior semantic knowledge and tries to mimic it computationally to make the images more human-like.

Abstracting visuals is a core concept in art and design, but performing this abstraction is a non-trivial task. Making the correct choices about what to emphasize in a piece depends on your context, goals, personal taste, and customers' taste. Therefore, performing this task in a visually pleasing way is highly challenging. A tool that computationally performs abstractions would be helpful for designers and artists alike.

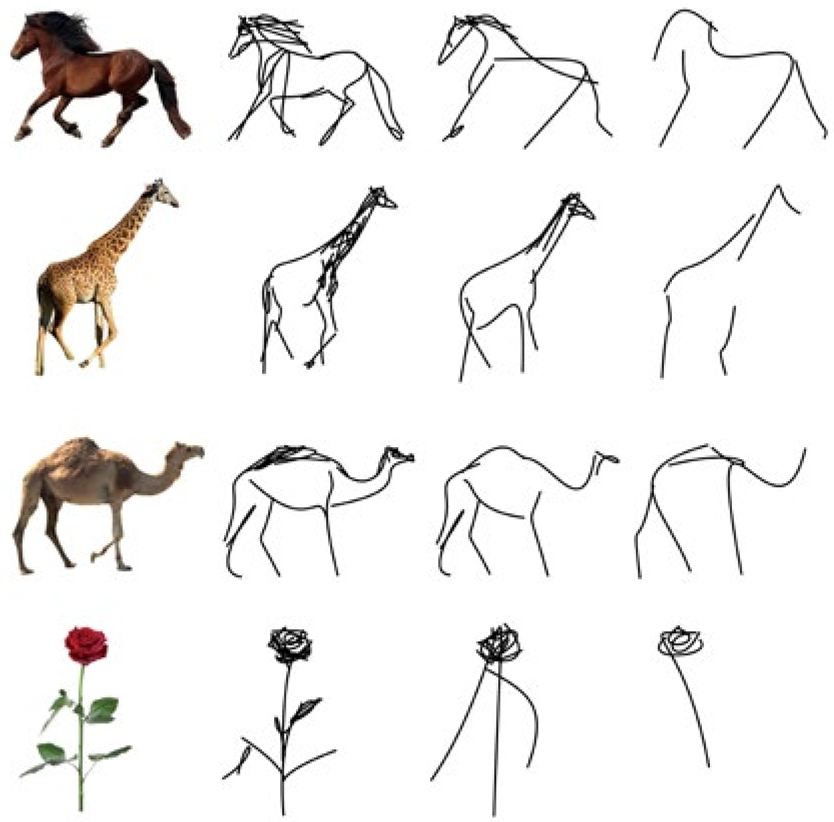
"I also think there is a less practical and more philosophical point, from a research perspective," Yael points out.

"These days, we have large new models combining text and images, which have been proven to be very strong. By defining challenging tasks like ours performing this progressive abstraction, I think we've shed light on the advancement of this field and how far it can go. If we can take it one step further to solve it with computational methods, it's really exciting research-wise."

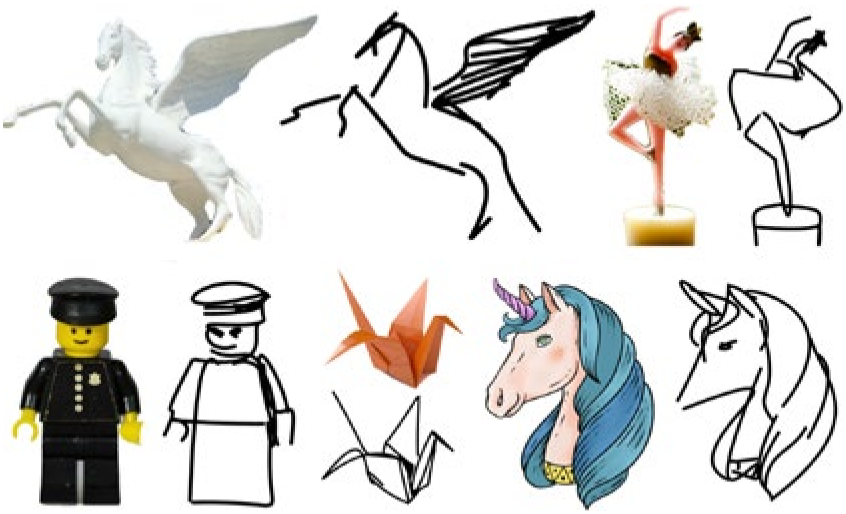
This work uses a popular model from OpenAI called CLIP. CLIP was originally trained to fit pairs of images and text. Being trained on a vast dataset of text-image pairs, CLIP is primarily semantic in nature and, until now, has been used mainly to generate images from text, not photos. However, this work is only interested in visuals; it does not use text at all.



Results of the method.



One image shows more examples of progressive abstraction, and the other one demonstrates the robustness of the method to unique categories that we don't usually see in common sketch datasets.



*"We want the model to **draw an input image and maintain a connection to that image**,"* Yael says.

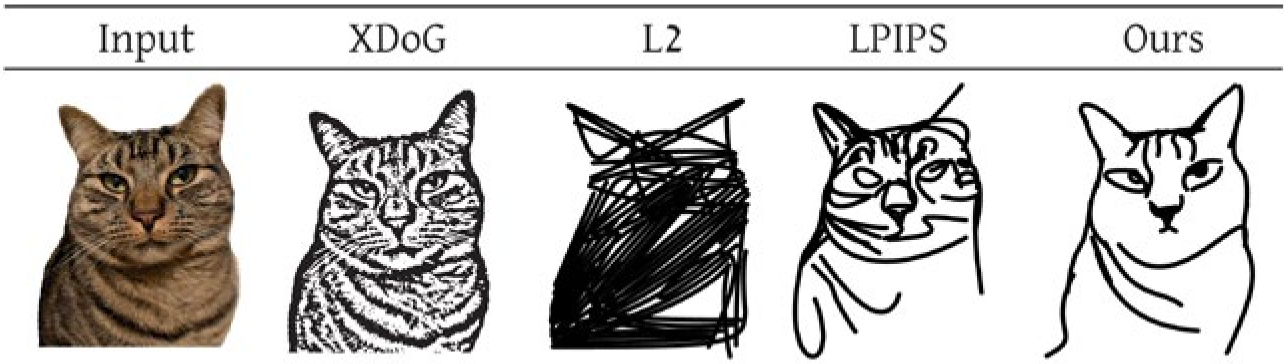
*"If I give it an image of my cat, and I want it to draw it, I don't want it to draw any cat; I want it to draw my cat. It must bear some relation to the input image, which is challenging to do with CLIP because it wasn't meant for this task. It tends to lose the geometric relation to the input image. **How do we combine the strong semantic features we get from CLIP while maintaining some relation to the structure of the input subject?**"*

This paper proposes two approaches to solve this issue. The first is inspired by a method called **LPIPS**. When comparing two images, you can use something simple like pixel-based loss or something more advanced called perceptual loss. In perceptual loss, instead of comparing pixels, you are comparing the activations of a pre-trained neural network. You are comparing something already learned about the images and the important features of each image.

"Inspired by that, we proposed to use CLIP as a perceptual loss," Yael continues.

"We use the intermediate-level activations of CLIP to define a perceptual loss between the input image and the output sketch. That way, we can use the huge power of CLIP to compare the two images."

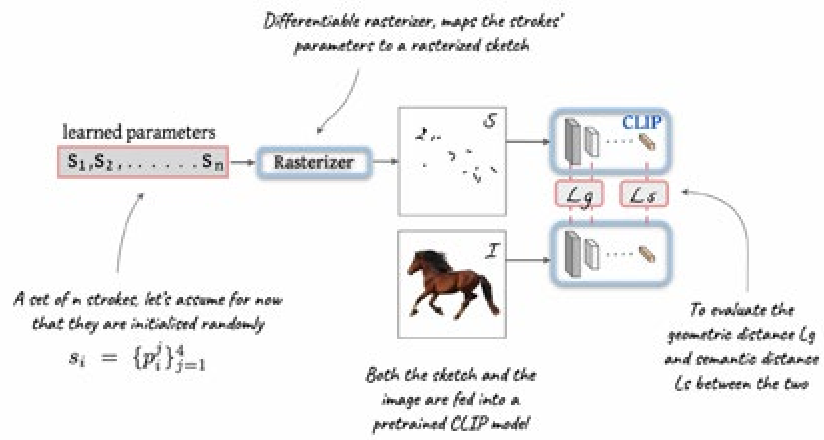
Secondly, the sketches consist of a limited number of strokes, and implementation is based on directly optimizing the strokes' parameters. This process is highly non-stable because it optimizes the parameters of the strokes directly rather than training a network. Initialization matters a great deal here.



An ablation of the proposed CLIP-based perceptual loss compared to L2, LPIPS, and just an edge map. This figure can help to understand the "semantically-our" part. When using the proposed CLIP-based loss, the semantic features of the cat are emphasized (such as the nose, eyes, and ears). In contrast, simple methods that are based only on pixel intensity such as XDoG or L2 do not capture the essence of the input image, as such operators do not "understand" the semantic concept behind the image (i.e. "a cat").

"We propose a method for initializing the strokes based on the salient regions of the input image," Yael explains.

"We use CLIP to analyze the input image and extract a heat map of the pixels, where we get a higher score for more important pixels. When drawing a cat, you want to focus on the eyes, the ears, the whiskers, and not necessarily the body. This approach gives the optimization process a better chance. If it starts from a better initialization based on the salient regions of the image, then we show that the strokes converge to a better solution. ***We also propose controlling the abstraction level by changing the number of strokes, which people haven't done before.****"*



Yael was one of five winners of a **Best Paper award at SIGGRAPH 2022** in August for this work. A huge achievement of which she should be very proud. We ask what she thinks convinced the judges.

"That's a hard question," she says modestly before taking a few moments to consider.

"I guess it was a combination of the idea and the implementation. The idea is quite original. We're aware of only two other works that tried to generate sketches with different levels of abstraction. We recognize the importance of this field and that it's not well explored. Also, the outcome is visually pleasing. That's why I think abstractions are core to art and design. People like to see it. It's fun to look at it, and it's beautiful. ***Our implementation, thanks to CLIP, led to high-quality results, and our method is simple and easy to understand. I think these are the reasons we won!****"*

Find more examples and videos on the project page, **here**.

Free and easy-to-use demos **here** and **here**.

INTRO TO NeRF

By Marica Muffoletto (twitter)

Dear readers, do you remember when the paper **NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis made its appearance** came out? A couple of years ago, Computer Vision News featured this paper and his first author in the ECCV Daily and in **the BEST OF ECCV 2020**. Since then, this type of generative model has become more and more popular.

NeRF is doing a reconstruction of scenes by using multiple images as input for a scene. This is referred to as ”novel view synthesis”.

Doesn‘t it look great?

In the following pages, we are going to use a repo made available by @brandontrabucco on GitHub which implements NeRFs on PyTorch (the origienal version of the code, released 2 years ago, was written in TensorFlow).

Let's start by importing some basic tools and packages.

```
# Download sample data used in the official tiny_nerf example
import os
import torch
import torch.optim as optim
import torch.utils.data as datatorch
import numpy as np
import matplotlib.pyplot as plt

# Setup GPU usage
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
```

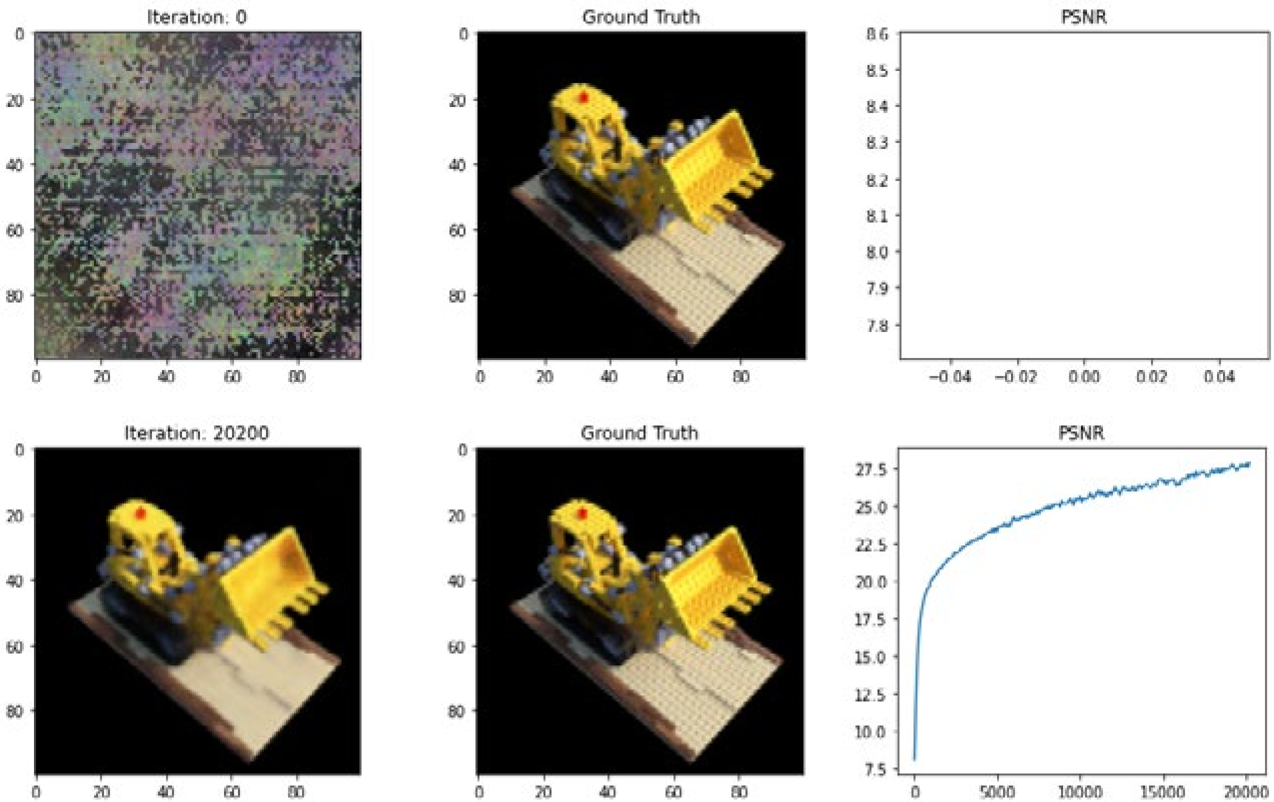
Now we can download an example dataset used to train NeRF. We used this link **https://people.eecs.berkeley.edu/~bmild/nerf/tiny_nerf_data.npz**, and moved the downloaded file to the current directory.

```
# Load input images, poses, and intrinsics
data = np.load("tiny_nerf_data.npz")

# Images
images = data["images"]
# Camera extrinsics (poses)
tform_cam2world = data["poses"]
tform_cam2world = torch.from_numpy(tform_cam2world).to(device)
# Focal length (intrinsics)
focal_length = data["focal"]
focal_length = torch.from_numpy(focal_length).to(device)

# Height and width of each image
height, width = images.shape[1:3]

# Near and far clipping thresholds for depth values.
near_thresh = 2.
far_thresh = 6.

# Hold one image out (for test).
testimg, testpose = images[101], tform_cam2world[101]
testimg = torch.from_numpy(np.expand_dims(testimg, axis=0)).to(devi-
ce)
testpose = testpose.expand([1,4,4])

# Map images to device
images = torch.from_numpy(images[:100, ..., :3]).to(device)

plt.imshow(testimg[0].detach().cpu().numpy())
plt.show()
```

Now we can install the repo through pip and import the model. The NeRF model, as described in the original paper, is just a multilayer perceptron that takes encoded 3D points and directions (x, d) as inputs and returns RGB values (c, σ). Quite a simple architecture that is helped by other components: a *positional encoder* to extract high frequency variations, a *differentiable volume renderer* to convert the RGBA output points in 3D (c, σ) to an imaage, a *stratified sampling*, and a *hierarchical volume sampling*. These elements are already integrated in this package, so we can just easily import the full model, after downloading **nerf-pytorch** through pip.

```
!pip install nerf-pytorch

from nerf.model import NeRF
from nerf.dataset import PixelRayDataset

# build the NeRF model with default parameters
model = NeRF(normalize_position=6.0).cuda()
nerf_optimizer = optim.Adam(model.parameters(), lr=0.0001)
poses = tform_cam2world

# create a dataset of pixels and corresponding rays for NeRF
dataset = PixelRayDataset(images.to(device),
                          poses.to(device), focal_length)
data_loader = datatorch.DataLoader(dataset, batch_size=1024, shuf-
fle=True)
train_features = next(iter(data_loader))
print(train_features.keys())

num_samples_per_ray = 64
randomly_sample = True
density_noise_std = 1.0

psnrs = []
iternums = []

iteration = 0
for epoch in range(100):
  for batch in iter(data_loader):

      # render a pixel for each ray using NeRF
      pixels = model.render_rays(
          batch['rays_o'],
          batch['rays_d'],
          near_thresh,
          far_thresh,
          num_samples_per_ray,
          randomly_sample=randomly_sample,
          density_noise_std=density_noise_std)

      nerf_optimizer.zero_grad()

      # mean squared error in pixels
      loss = ((pixels - batch['pixels']) ** 2).mean()
      loss.backward()
      nerf_optimizer.step()

      if iteration % 100 == 0:
          with torch.no_grad():
              test_render = model.render_image(
                  testpose[..., :3,  3],
                  testpose[..., :3, :3],
                  height,
                  width,
                  focal_length.item(),
                  2.0,
                  6.0,
                  64)
```

```
        psnr = -10.0 * torch.log(((
            test_render -
            testimg) ** 2).mean()) / 2.30258509299

        psnrs.append(psnr.cpu().detach().numpy())
        iternums.append(iteration)

    plt.figure(figsize=(15,4))

    plt.subplot(131)
    plt.title(f'Iteration: {iteration}')
    plt.imshow(test_render[0].detach().cpu().numpy())

    plt.subplot(132)
    plt.title('Ground Truth')
    plt.imshow(testimg[0].detach().cpu().numpy())

    plt.subplot(133)
    plt.plot(iternums, psnrs)
    plt.title('PSNR')

    plt.show()

iteration += 1
```



After some epochs results look pretty good already.

Different datasets are available online to play with NeRFs, and they include both synthetic and realistic data. NeRF can also be used to generate meshes, since they return a 3D shape explicitly as a continuous function. If you are interested in this application, don't hesitate to let us know and we will talk more about it in the following months!

Congrats, Doctor!

Oana Ignat completed her PhD in Computer Science at University of Michigan, where she was member of the **Language and Information Technologies Lab**, advised by Prof. Rada Mihalcea.

Her research is at the intersection of Natural Language Processing and Computer Vision, using social media videos and their transcripts, to develop datasets and models for human action understanding.

She is passionate about teaching and increasing diversity in Computer Science and AI, and has been a primary instructor for **Discover CS (EECS 198)**, a class primarily designed for underrepresented minorities in CS. Oana wants to be involved in the AI4Good movement, applying AI in problems that have potential to improve the world.

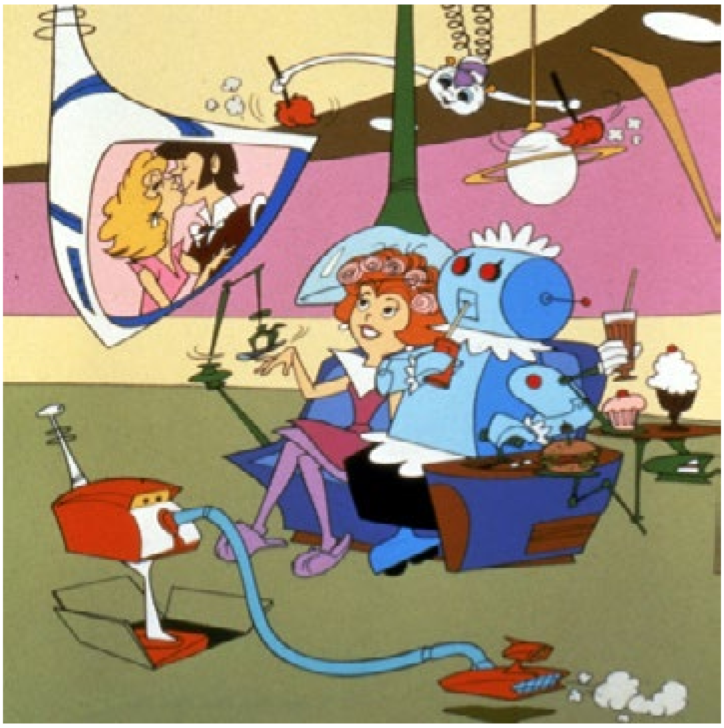


Fig 1. We explore how we can use multimodal (textual and visual) information from online narrated videos/ vlogs, to enable automatic models to learn about human daily actions.

Human action understanding is one of the most impactful and challenging tasks a computer system can do. Once a computer system learns how to interact with humans, it can assist us in our everyday life activities and significantly improve our quality of life.

Despite the attention it has received in fields such as Natural Language Processing and Computer Vision, and the significant strides towards accurate and robust action recognition and localization systems, human action understanding still remains an unsolved problem.

As a step towards endowing systems with a richer understanding of human actions in online videos, this thesis proposes new techniques that rely on the vision and language channels to address four important challenges: i) **human action visibility identification in online videos** [1], ii) **temporal human action localization in online videos** [2], iii) **human action reason identification in online videos** [3], and iv) **human action co-occurrence identification** [4].

We focus on the widely spread genre of **lifestyle vlogs**, which consist of videos of people performing actions while verbally describing them. We construct a dataset with crowdsourced manual annotations of visible actions, temporal action localization and action reason identification in online vlogs.

We propose a **multimodal unsupervised model** to automatically infer the reasons corresponding to an action presented in the video (Fig. 2), a simple yet effective method to **localize the narrated actions** based on their expected duration, and a **multimodal supervised classification model of action visibility in videos.** We also perform **ablations on how each modality contributes to solving the tasks** and compare the multimodal models' performance with the single-modalities models based on the visual content and vlog transcripts. Finally, we present an **extensive analysis of this data**, which allows for a better understanding of how the language and visual modalities interact throughout the videos and pave the road for rich avenues for future work.

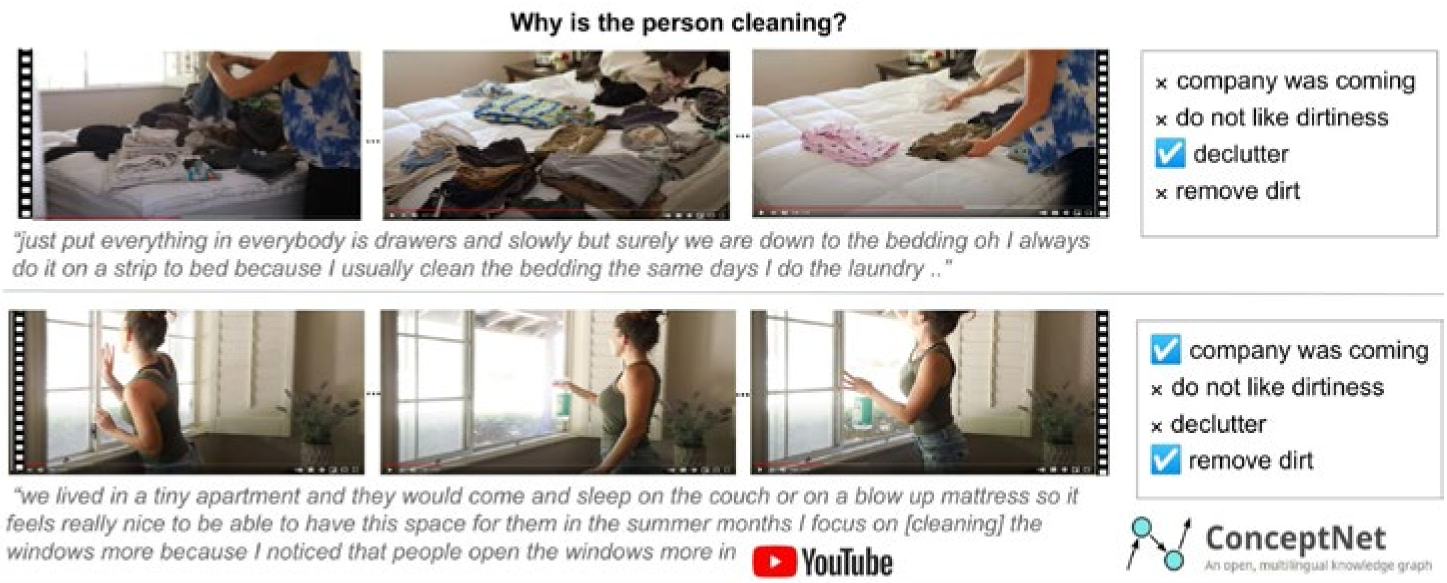


Figure 2. Overview of our task: *automatic identification of action reasons in online videos.* ***The reasons for*** *cleaning* ***change based on the visual and textual (video transcript) context. The videos are selected from*** **YouTube*****, and the actions together with their reasons are obtained from the*** **ConceptNet** ***knowledge graph which we supplement with crowdsourced reasons. The figure shows two examples from our*** **WhyAct** ***dataset.***

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

Lyft, Argo AI begin robotaxi rides in Austin

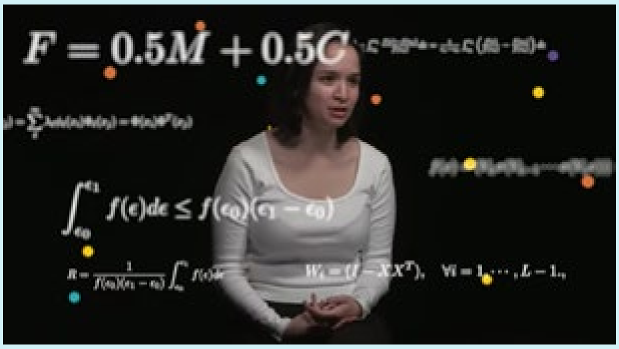
Almost there: after hitting Las Vegas and Miami, Lyft and Argo AI are deploying their robotaxi also in Austin, Texas. Does it mean that the era of autonomous cars has officially started? Not yet. These Ford-built vehicles are still hosting not one, but two safety drivers in the front seats. Rides are requested via the Lyft app and ride prizes will be the same as regular Lyft rides. The goal is to scale up to at least 1,000 autonomous vehicles on the Lyft network over the next five years. Of course, the company aims at offering fully autonomous rides in the future. When? When regulatory environment, safety data and community acceptance will permit it. **Read More**

Intel Launches Geti OpenVINO-Optimized Computer Vision Platform, Early-Access Developer Cloud

We already hinted at this 3 months ago on our annual **BEST OF CVPR report**: the great folks at **Intel**, still aiming at being considered **a software company** (with 20,000 software developers!), have launched Geti, a new computer vision platform, designed specifically for OpenVINO. Geti provides a single unified interface for data upload and annotation, model training, and retraining; while OpenVINO is Intel's Open Visual Inference and Neural-network Optimization toolkit. Why is this so important? Do you want to know more? Drop us a word and we'll invite the Intel folks to tell us more in one of our next issues. **Read More**

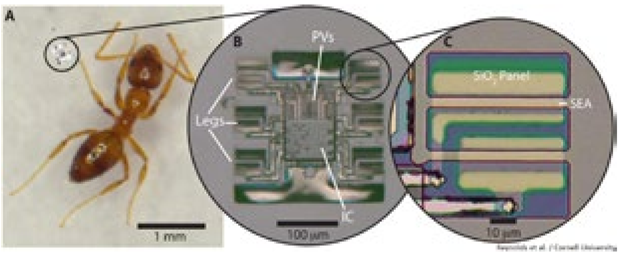
Understanding reality through algorithms

This astonishing young woman - **Fernanda de la Torre** is her name – was born in Mexico and immigrated very young to the USA. She is now a **Neuroscience PhD student at MIT** and she uses complex algorithms to investigate philosophical questions about **perception and reality**, *"to understand how the mind works, and how it is that we can all be in the same environment and feel very different things."* Fernanda is now engaged in what she describes as more philosophical questions about how one develops a sense of self as an independent being. **Watch the Inspiring Video!**



Make-A-Video: An AI system that generates videos from text

That's the announcement made by **Meta** of **Make-A-Video**, a new AI system that lets people turn text prompts into brief, high-quality video clips. With just a few words or lines of text, Make-A-Video can create videos full of colors, characters, and landscapes. The system can also create videos from images or take existing videos and create new ones that are similar. The system is not available yet for public use and – obviously – most existing demos are videos of animals – existing or imaginary. If you want to learn about this, we can invite the fine Meta folks on one of our next issues to tell us more. Read More

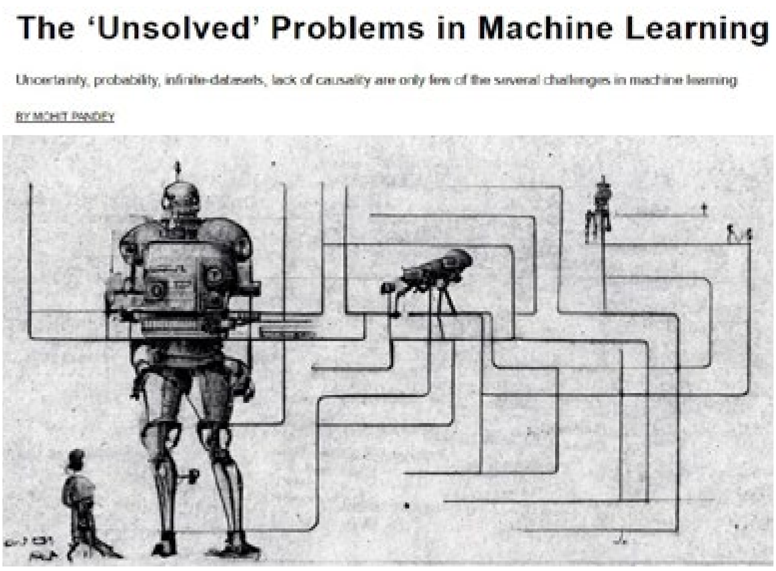


These Autonomous, Wireless Robots Could Dance on a Human Hair

Apparently these robots are really really tiny, but that's the direction we are heading into: they are called Antbots by their creators in the Cornell lab led by Michael Reynolds. As they say, they are about the size of an ant **to an ant**. They comprise three major systems: a photovoltaic cell to accept light as power, a tiny integrated circuit for controlling and directing that power and a set of hinged legs that it uses to scoot itself around. It is only a bit disappointing that the intelligence of these robots is directly limited by the scale of the electronics it is just not yet possible to fit enough on there to perform complex logic. Read More

The 'Unsolved' Problem in Machine Learning

There are various domains in the field of AI and machine learning that developers dive deep into and come up with small incremental improvements. However, challenges to further advancement in these fields persist, which make these problems yet unsolved. Stimulated by a very lively discussion on reddit.com/r/MachineLearning, these problems were reviewed and categorized. You will not be surprised to find among them issues like uncertainty prediction, overfitting, estimating causality instead of correlations and reproducibility. A very interesting Reddit post suggests: low-energy deep learning systems. Read More



COMPUTER VISION EVENTS

AI & Big Data Expo North America Santa Clara, CA 5-6 October	AI in Healthcare Summit Boston, MA 13-14 October	ICIP Bordeaux, France 16-19 October
ECCV 2022 MEET US THERE Tel Aviv, Israel 23-27 October	IROS 2022 Kyoto, Japan 23-27 October	MedTech Conference Boston, MA 24-26 October
SIPAIM Valparaiso, Chile 9-11 November	Innovation Summit San Francisco, CA 14-16 November	
GeoMedIA Workshop Amsterdam, the Netherlands 18 November		
MICAD University of Leicester, UK 20-21 November		

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!

SUBSCRIBE!

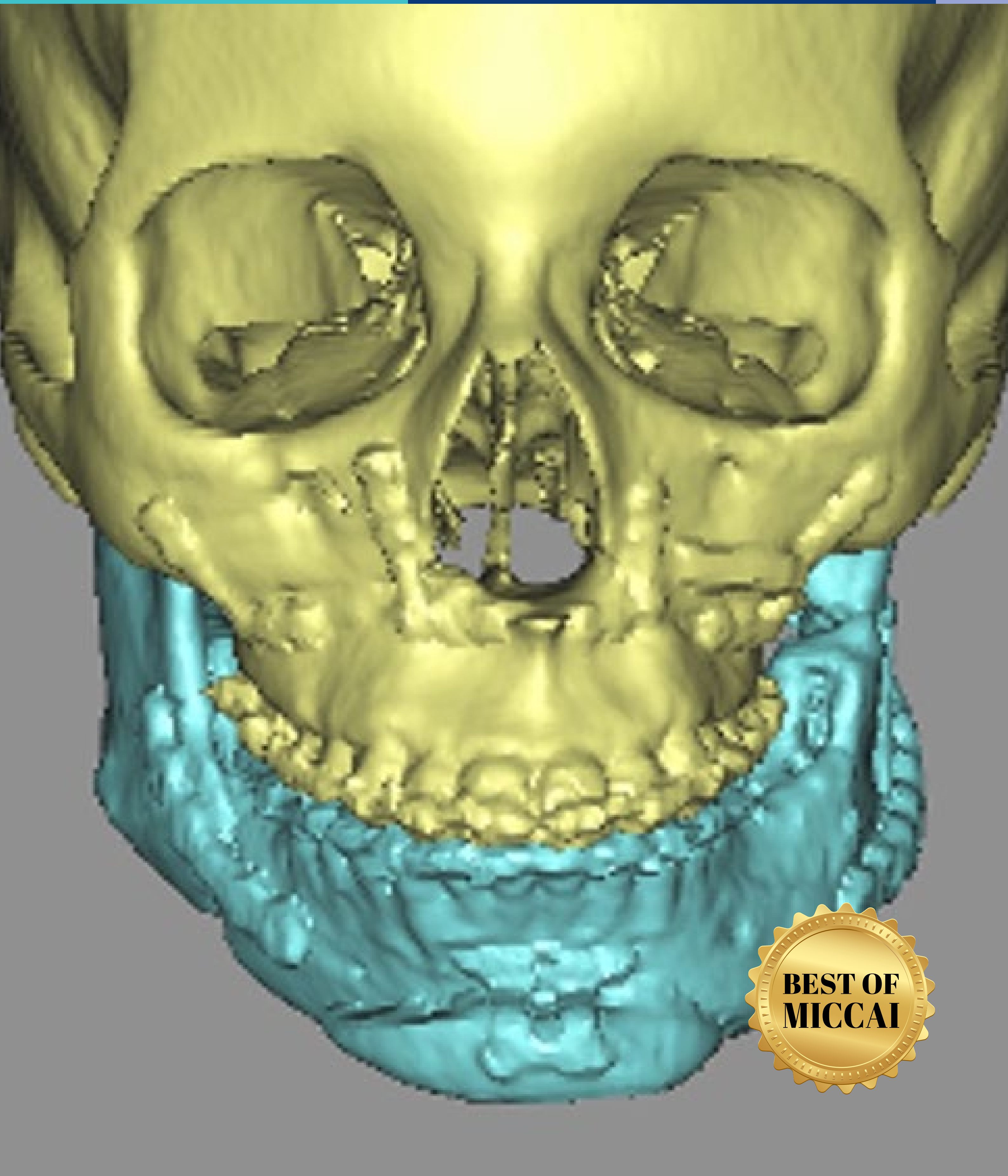
Join thousands of AI professionals who receive Computer Vision News as soon as we publish it. You can also visit our archive to find new and old issues as well.

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

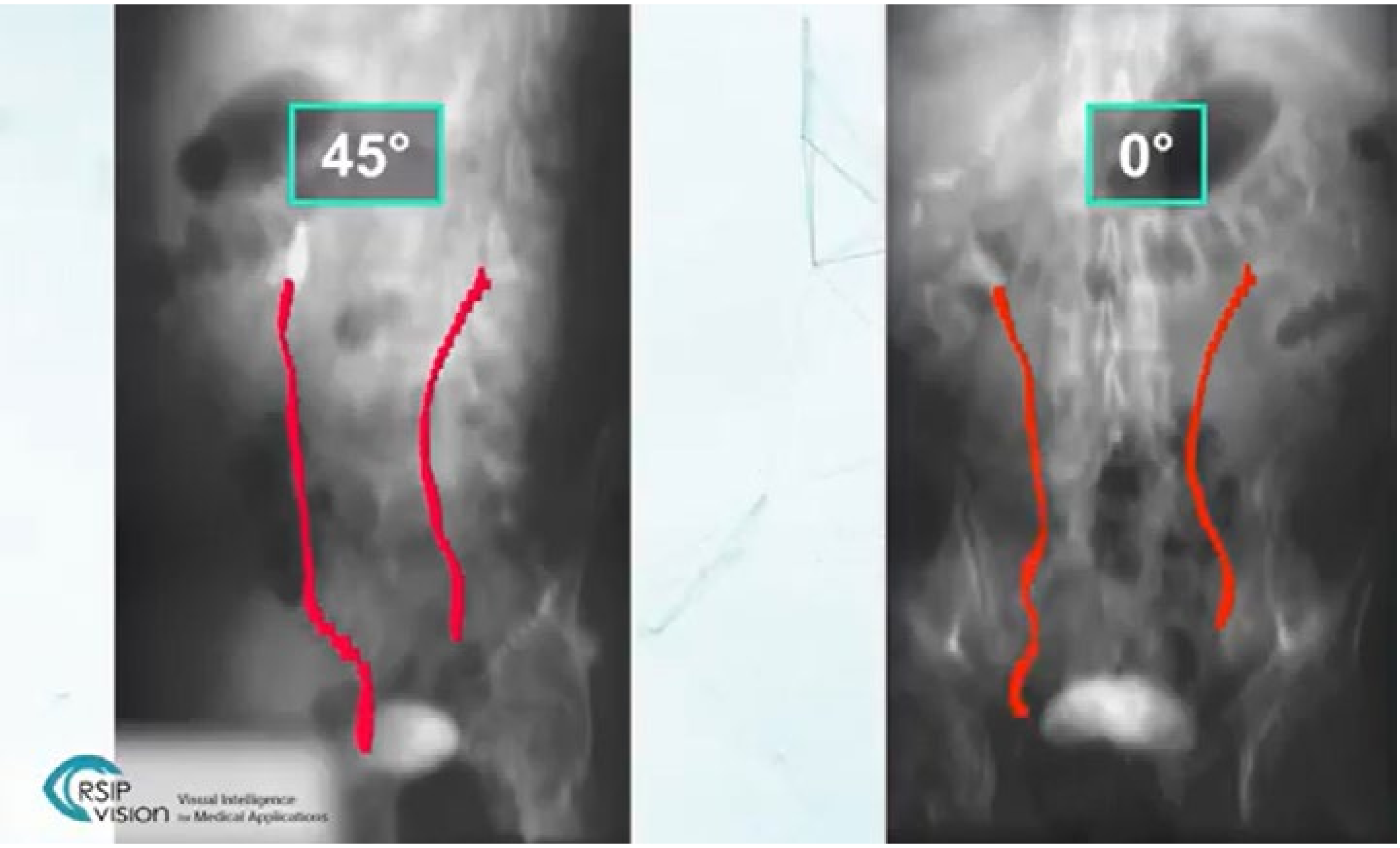
Due to the pandemic situation, most shows are considering going virtual or to be held at another date. Please check the latest information on their website before making any plans!

MEDICAL IMAGING NEWS

OCTOBER 2022



RSIP VISION PRESENTS NEW UROLOGICAL AI TOOL FOR 3D RECONSTRUCTION OF THE URETER DURING UROLOGICAL PROCEDURES



Ureteroplasty is a procedure for surgical reconstruction of the ureter: it is conducted when urine flow is disrupted and non-invasive treatments are ineffective. When ureteroplasty is deemed necessary, retrograde or intravenous pyelogram, contrast injection into the urinary tract followed by fluoroscopy, is performed to demonstrate the ureter's position, integrity and patency. These fluoroscopy images adequately portray the aforementioned features, however, as they lack the depth dimension, surgical planning and navigation lacks accuracy as well.

Recently, **RSIP Vision** developed an exceptional tool which utilizes different fluoroscopy projections, and reconstructs an accurate 3D model of the knee. This concept was recently further developed to accommodate ureter fluoroscopy.

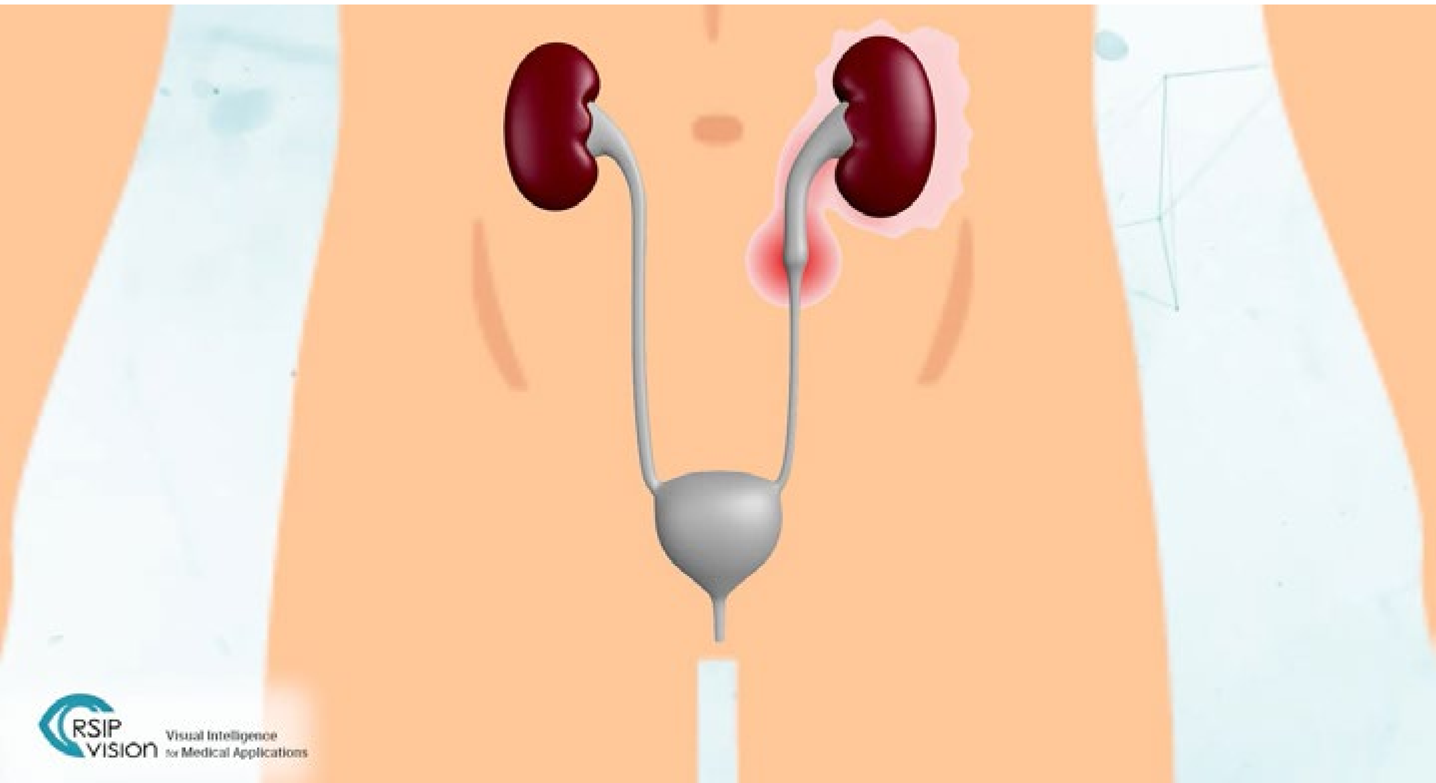
3D ureter reconstruction is conducted in two steps - both a custom-tailored algorithm combining deep-learning and classic computer vision techniques. Once at least two fluoroscopy images are acquired from different projections, the ureter is automatically segmented, and the images are calibrated. The calibration can be adjusted using a pre-calibrated jig within the field-of-view, or by extracting specific c-arm geometry. Segmentation is performed using trained neural networks dedicated for this task.

The second step consists of the actual reconstruction. The adjusted images and segmentations are passed to a deep neural network that produces the 3D reconstructed ureter. This model accurately portrays the ureter and can be used by the physician for procedure planning and general anatomical consideration. A 3D model of the ureter can be used as a navigational aid in other urological procedures too: stent placement, stone localization and retrieval, etc.

Throughout the procedure, this model is used as a reference to verify positioning and anatomical location. As the procedure progresses, additional fluoroscopy images are acquired to assess procedural progress, and a recent ureter model can be compared to the reference one for better progress tracking. Also, the availability of accurate 3D information during the procedure results in better surgical precision and higher confidence for the surgeon.

This technology can be applied to other procedures conducted with fluoroscopy imaging. As detailed above, we already developed a 2D-to-3D reconstruction solution for the knee. Additionally, this can be applied to bladder reconstruction, coronary artery 3D modeling, ERCP, etc. Each use-case requires a custom-tailored solution, and the baseline is leveraged to fit the new need.

RSIP Vision's long track record allows the development of this tool by combining advanced segmentation algorithms with 3D reconstruction techniques in an efficient manner, providing a real-time solution for 3D ureter reconstruction. Ureteroplasty can be conducted faster, accurately, and ultimately with less complications.




James Xia

DEEP LEARNING-BASED FACIAL APPEARANCE SIMULATION DRIVEN BY SURGICALLY PLANNED CRANIOMAXILLOFACIAL BONY MOVEMENT


Daeseung Kim


Pingkun Yan

Xi Fang

Houston Methodist Hospital and Professor of Oral and Maxillofacial Surgery.

Daeseung Kim (bottom left) works for James Xia as an Instructor at Houston Methodist Hospital Research Institute.

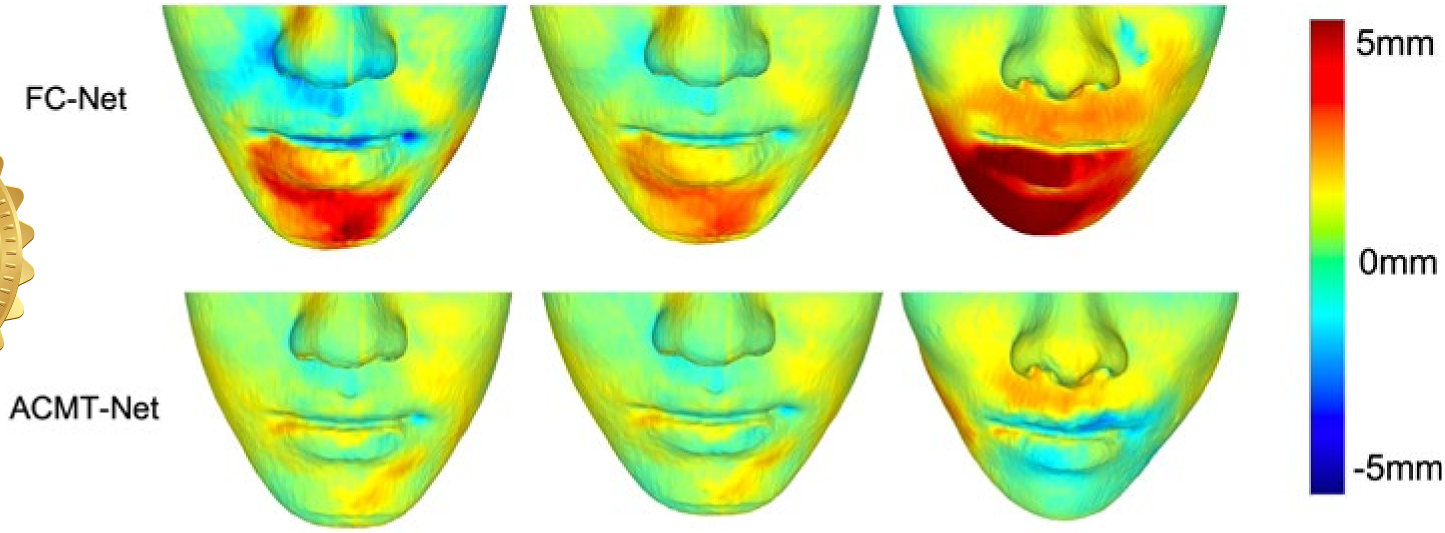
Pingkun Yan (top right) is an Associate Professor in the Department of Biomedical Engineering at Rensselaer Polytechnic Institute.

Xi Fang (bottom right) is a PhD student in Dr. Yan's lab. Their paper proposes a new deep-learning method for simulating facial appearance change following bony movement. Together, they spoke to us ahead of Xi's oral presentation.

Orthognathic surgery is a bony surgical procedure designed to correct jaw deformities by **aligning the upper and lower jaw**. The surgery can help people in several ways, including those with an underbite or overbite or those suffering from sleep apnea, where the upper and lower jaws are advanced to enlarge airspace.

Although the facial soft tissue is not directly operated on, changing the underlying bone structure will cause facial change. The challenge is how to model the facial tissue deformation efficiently and accurately when you move the bone.

"When you start moving the bones, you do not know how the facial soft tissue is going to react," James tells us. *"We can use our artistic imagination and say, if I advance the upper jaw a*



few millimeters this way, and I set back the lower jaw a few millimeters that way, your face is going to look like this, but ***there is no scientific way to predict how the face is going to react after the bony surgery."***

This project was initiated around a decade ago, and James recently worked with Daeseung on a paper exploring a novel approach using the **finite-element method (FEM)** to simulate facial changes. FEM is a numerical biomechanics-based method reported to be the most common and accurate for simulating or analyzing mechanical changes in body structure. However, its prediction accuracy in clinically critical regions, such as the lips, was below the acceptable range.

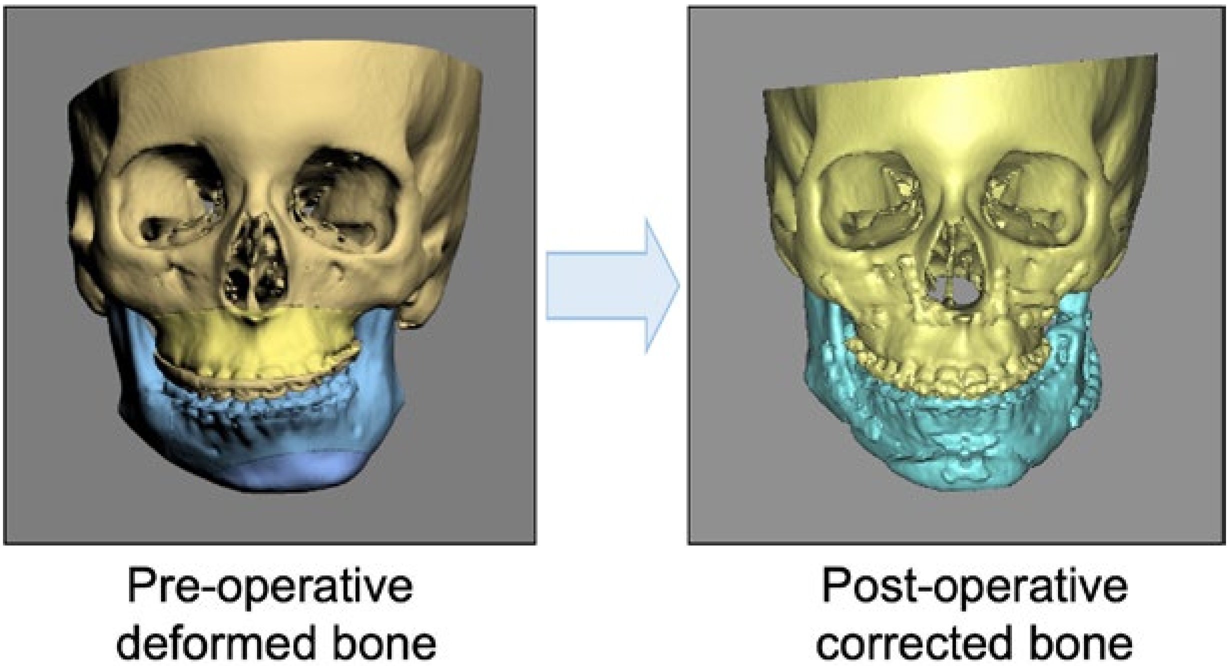
"We proposed to improve the accuracy along this clinically critical area with a method called ***incremental facial change simulation with realistic lip sliding effect,"*** Daeseung tells us. *"We focused on the sliding effect of the soft tissue on the bone. We needed to accurately simulate the interaction between the bony structure and soft tissue to predict the facial changes."*

Most previous works used a simple simulation condition, assuming bony structure and soft tissue were attached and moved together. The reality is that when bone moves, soft tissue freely slides over its surface, so their paper's method applied mucosa and lip sliding effects.

"Previous methods did not consider the opening of the lips," Daeseung points out. *"In those models, the upper and lower lips were connected, so you could not accurately simulate individual lip movements following bony movements. We separated the upper and lower lips in our model and and could accurately simulate their movements following each upper and lower jaw movement."*

Daeseung found that balancing the accuracy and complexity of the simulation condition was the most challenging aspect. If it was too complex, it used too many computational resources. The realistic lip sliding effect was the perfect balance between accuracy and efficiency. Facial changes were simulated little by little, which made it more natural, and it demonstrated a significant improvement over previous methods, improving prediction accuracy around the lips in both quantitative and qualitative measures.

Despite these positive results, FEM still presented some challenges, particularly



regarding **efficiency**. It would take some time to prepare the model, and surgeons with hectic schedules would not be able to wait in front of a computer for a model to be printed.

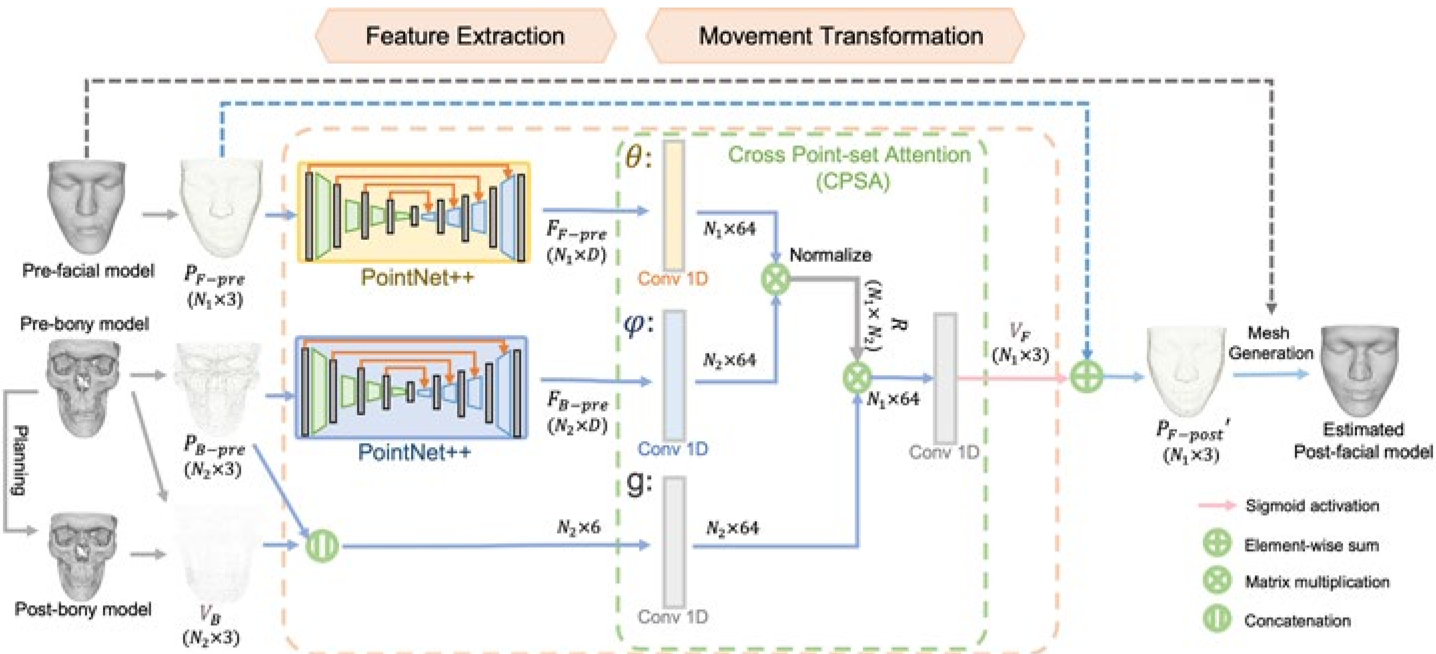
"Scientifically, it was sound, but in reality, it was not clinically practical," James says. *"That is why we started thinking about the deep learning method, which is where Dr. Yan and his group come in."*

Pingkun's group has been working on deep learning techniques, including direct image processing, image analysis, and image construction, for several years. The two groups have come together to work on this problem and have joint authorship of the paper. Daeseung and Xi are co-first authors, and James and Pingkun are co-corresponding authors. **It is a marriage of the clinical and the technica**l. *"We propose a deep learning-based approach,"* Pingkun tells us. *"FEM works because it can model the correspondence well between the bone and the facial tissue. Our method uses the attention mechanism to establish a correspondence, so that deep learning knows how the facial tissue moves correspondingly if a part of the bone moves.* ***We transform the movement of the bone to the soft tissue. That is a major technical innovation."***

Xi will be presenting the work today - his first oral presentation paper. He describes the group's method in more detail:

"First, we want to learn the spatial correspondence between bony and facial structures. We use ***PointNet++ networks*** *to extract the structural features from the bony and facial point set. We learn the local spatial features from the facial and bony models and then compute their similarities. Each facial point has its corresponding bony points, and all the corresponding points have a weight to contribute to that specific facial point.* ***There is a point-to-point correspondence matrix to transform the effect from the bone to the face."***

They use **MLP** to encode the bony movement into a local bony movement feature. Compared to previous methods, which encode the bony movements into a global vector that cannot be decoded locally, they encode it locally and



decode it locally, using the just-learned correspondence matrix to transform the local bony movement into the local facial change.

"Using the correspondence matrix, we can transform the corresponding bony movement into the corresponding facial change, and then decode it locally so that we can make the output the same dimension as the input bony movement," Xi explains. *"We simulate the facial change based on two key parts – the spatial correspondence between local bone and facial structures and the non-linear relationship between the corresponding bony movement and the facial change."*

The team performed a prediction accuracy evaluation using clinical data, comparing their proposed method with two other approaches, including state-of-the-art FEM with realistic lip sliding effect, against which it achieved comparable quantitative accuracy with significantly improved efficiency.

"We also asked clinicians to look at the results because numbers do not mean everything," Pingkun adds. *"We asked them to visually check whether the method worked well. We did a reader study, compared the results, and it showed very good performance."*

As we wrap up, James is keen to emphasize the importance of this project being a true collaboration between the medical and the technical side.

"Any new technology you present in the medical application base of conferences needs to have a true medical indication," he asserts. *"I have seen so many papers presented at MICCAI that have no real medical indication, meaning you are finding a problem where a problem never existed. We are solving actual medical problems that we encounter during our clinical practice. That is the goal."*

This chimes with us - **Nassir Navab** told us some time ago how important it is to him that his students are always in touch with clinicians and working on solutions driven by what the field needs. James agrees.

"Exactly. I know Nassir very well, and he is absolutely correct!"

GREEDY OPTIMIZATION OF ELECTRODE ARRANGEMENT FOR EPIRETINAL PROSTHESES

Ashley Bruce

Ashley Bruce is a recent graduate from UCSB and a Software Engineer at Veeva.

Michael Beyeler is an Assistant Professor in Computer Science and Psychological & Brain Sciences at the University of California, Santa Barbara.

Ashley and Michael spoke to us ahead of their poster presentation about their work exploring the optimal arrangement of electrodes in epiretinal prostheses.

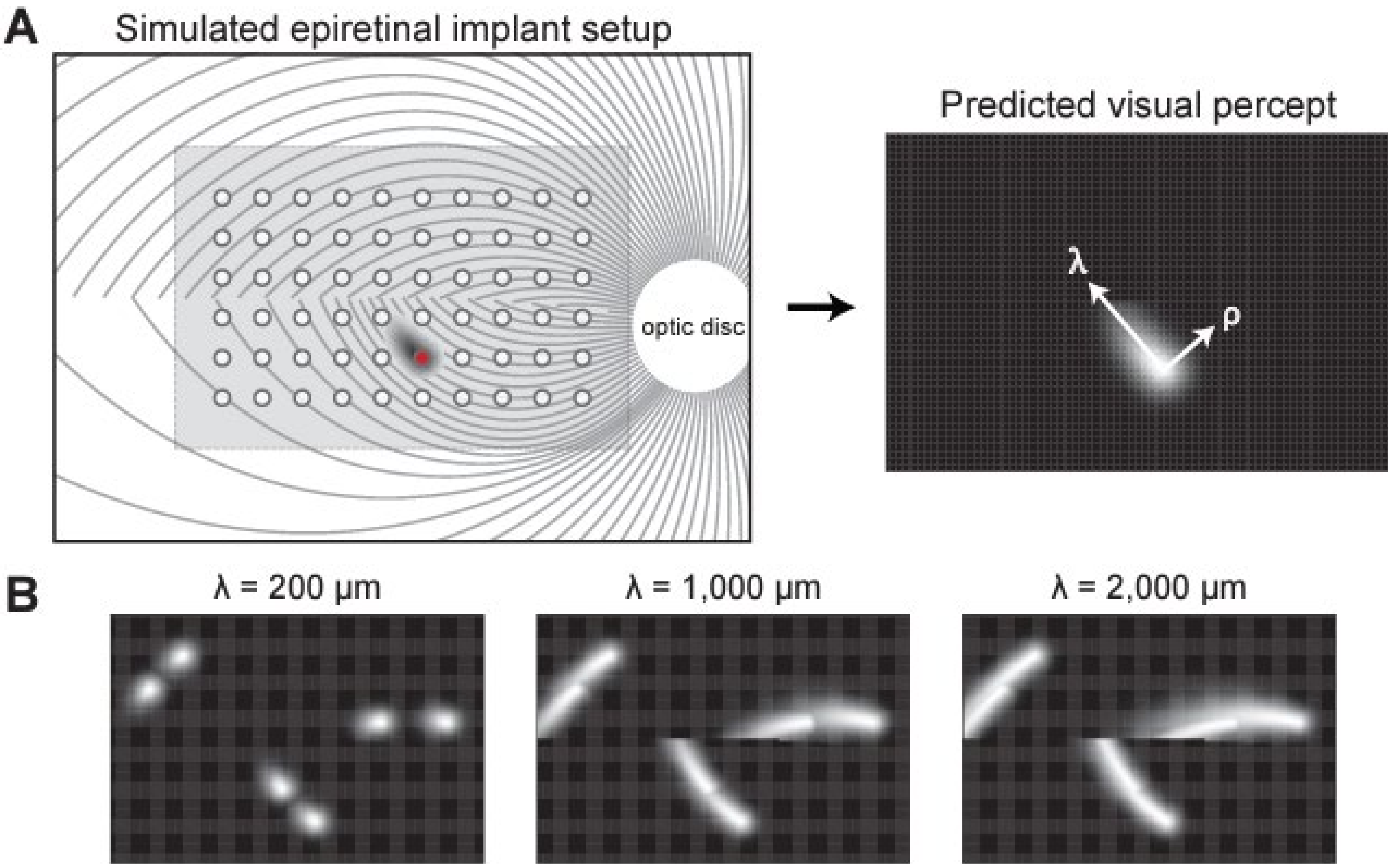
Several diseases could result in blindness, including some that slowly attack the retina. For the most part, these are hereditary diseases with no cure. **Retinal degeneration** leads to sight loss because retinal cells are the first step in the vision process. However, even when that first step is gone, it is still possible to hijack the pathway if everything else works.

"Epiretinal prostheses bypass the dead retinal cells by stimulating the next part of the pathway," Ashley explains. *"They stimulate the surviving cells and can produce these phosphenes or flashes of light and, quote-unquote, restore vision."*

There are already prostheses that use this pathway, but little research has gone into **optimizing the placement of electrodes on the prosthesis**. That is where this work comes in, proposing a better way to arrange electrodes on the implant to produce greater phosphene coverage.

"Current devices arrange their electrodes on a rectangular grid because it's compact and easy to fabricate," Michael tells us. "*Some people have looked at where to place the whole implant on the retina. Ashley was the first to ask, what if we moved every individual electrode around based on what we know about how these electrodes produce artificial vision?"*

However, moving every electrode presents the problem of combinatorial explosion. Even in current devices with only 60



electrodes, there are many possibilities for arranging them. It is not usually technically feasible to find a solution.

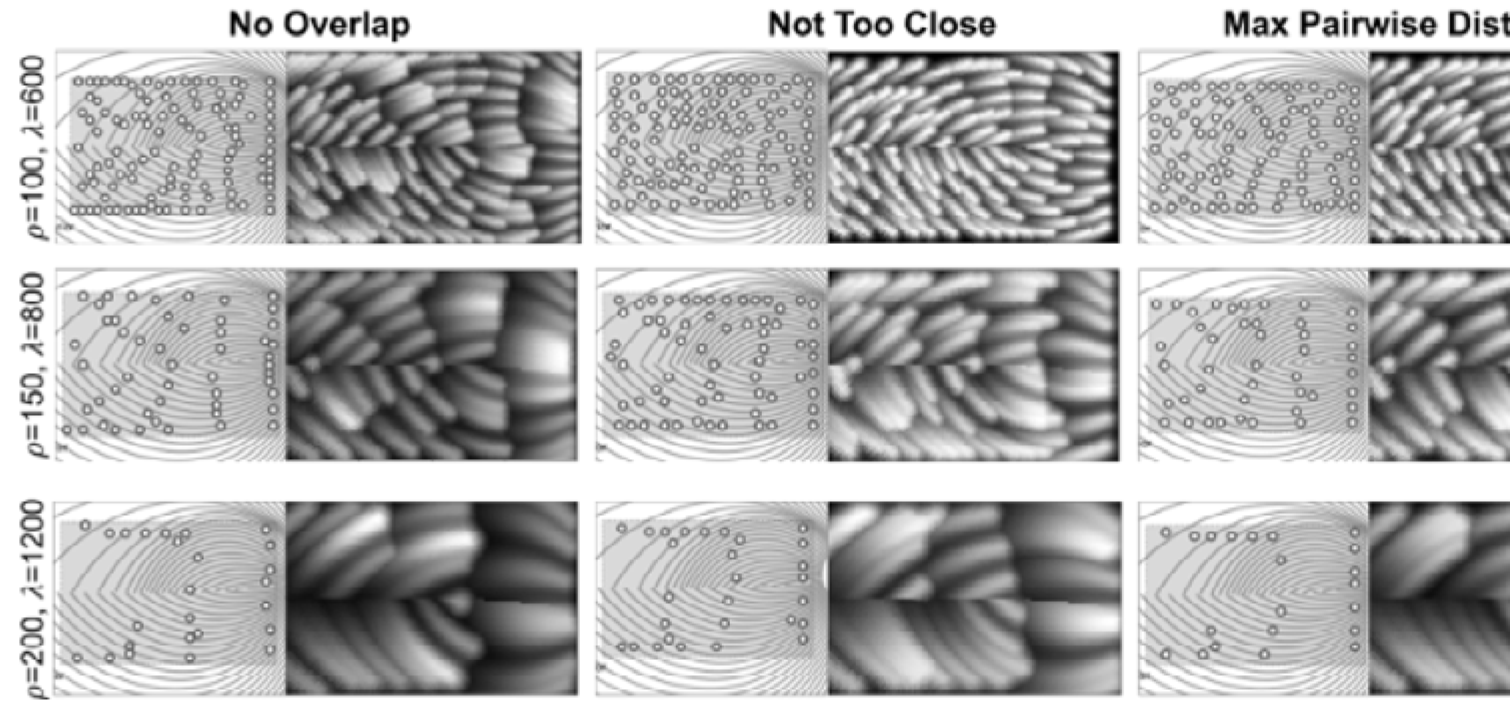
"Ashley approached this as a **greedy optimization problem,** *where one electrode is placed after another,"* Michael explains. *"We used a computational model of bionic vision to help predict what the vision would look like for a given electrode placement. By iterating over that, Ashley found a mathematically proven optimal solution."*

Greedy optimization is just that – a greedy approach to optimization. Each electrode is taken one at a time, its best placement is found, and it stays there. Then the same is done with the second electrode, taking a greedy approach to placing the electrodes in their optimal positions on the implant.

"We don't place the electrode if it doesn't improve the final result," Ashley adds. *"When you start getting higher ρ and λ values, or when you get too high in the number of electrodes, each electrode might not add as much to the next iteration. Therefore, we considered a small upper value to ensure we were still increasing our results; otherwise, that electrode was not helpful."*

Before settling on the greedy approach, Ashley and Michael went through several other options. They started exploring biological methods, including **particle swarm optimization**, which looked promising.

"We spent a lot of time on that, but there was no guarantee that the minimum we found was optimal, and it took way too long for something that wasn't even a fully optimal solution," Ashley recalls. *"When that didn't give us the results we wanted, it*



was rough, but it helped us re-evaluate and find a better solution."

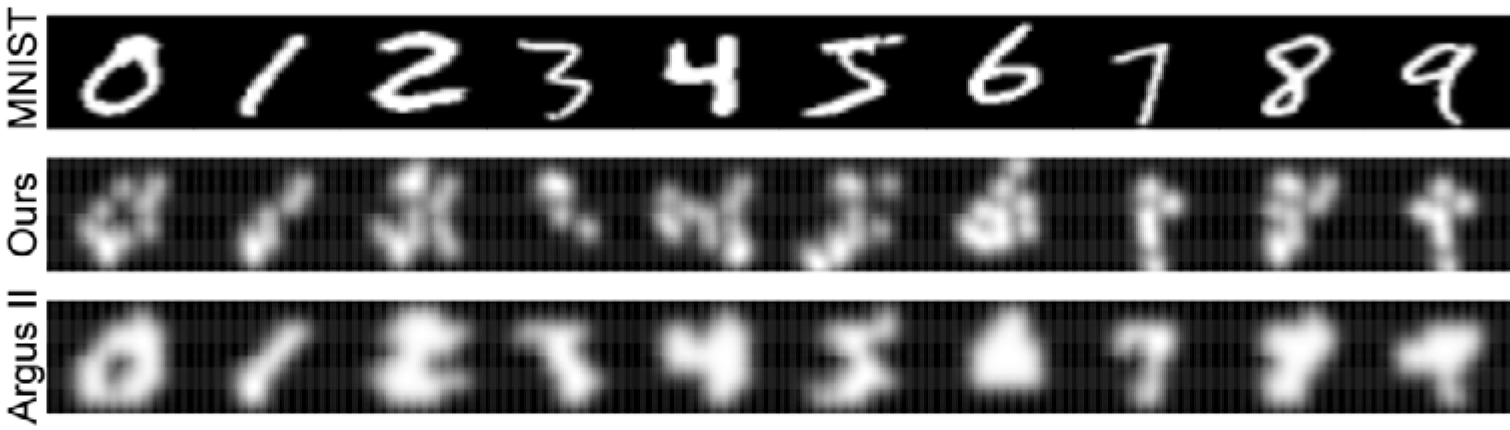
Outside of this project, Michael's lab works on how retinal and cortical implants produce artificial vision, using **insights from neuroscience and methods from computer vision and AI**. Meanwhile, Ashley just accepted a job at **Veeva**, working on software engineering for the life sciences.

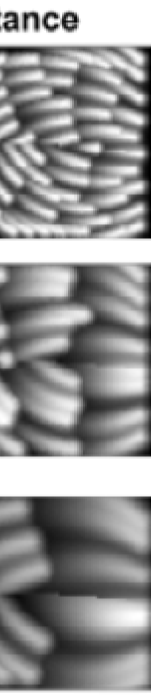
"It's really cool because I did my undergrad in biology and graduated in computer science," she says. *"When choosing a lab to work in, I was drawn to Michael's because it was a mix of that bio and computer science world. I'm happy to continue my passion for merging the life sciences with computer science and hopefully building a better future."*

Where does she see herself in five or ten years?

"It's hard to say because I used to think so far ahead, but now I like to figure out what I enjoy and continue doing that. If I'm no longer enjoying it, I'll make some changes, but currently, I like what I'm doing."

This conference paper is Ashley's first and

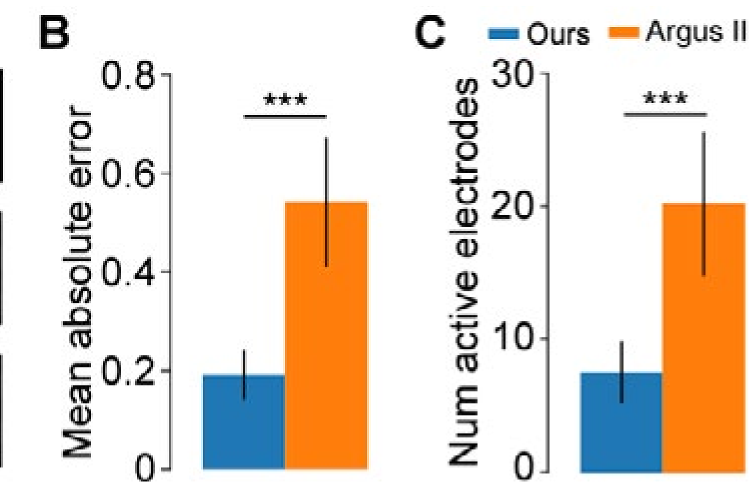




Michael Beyeler

has been a labor of love at times.

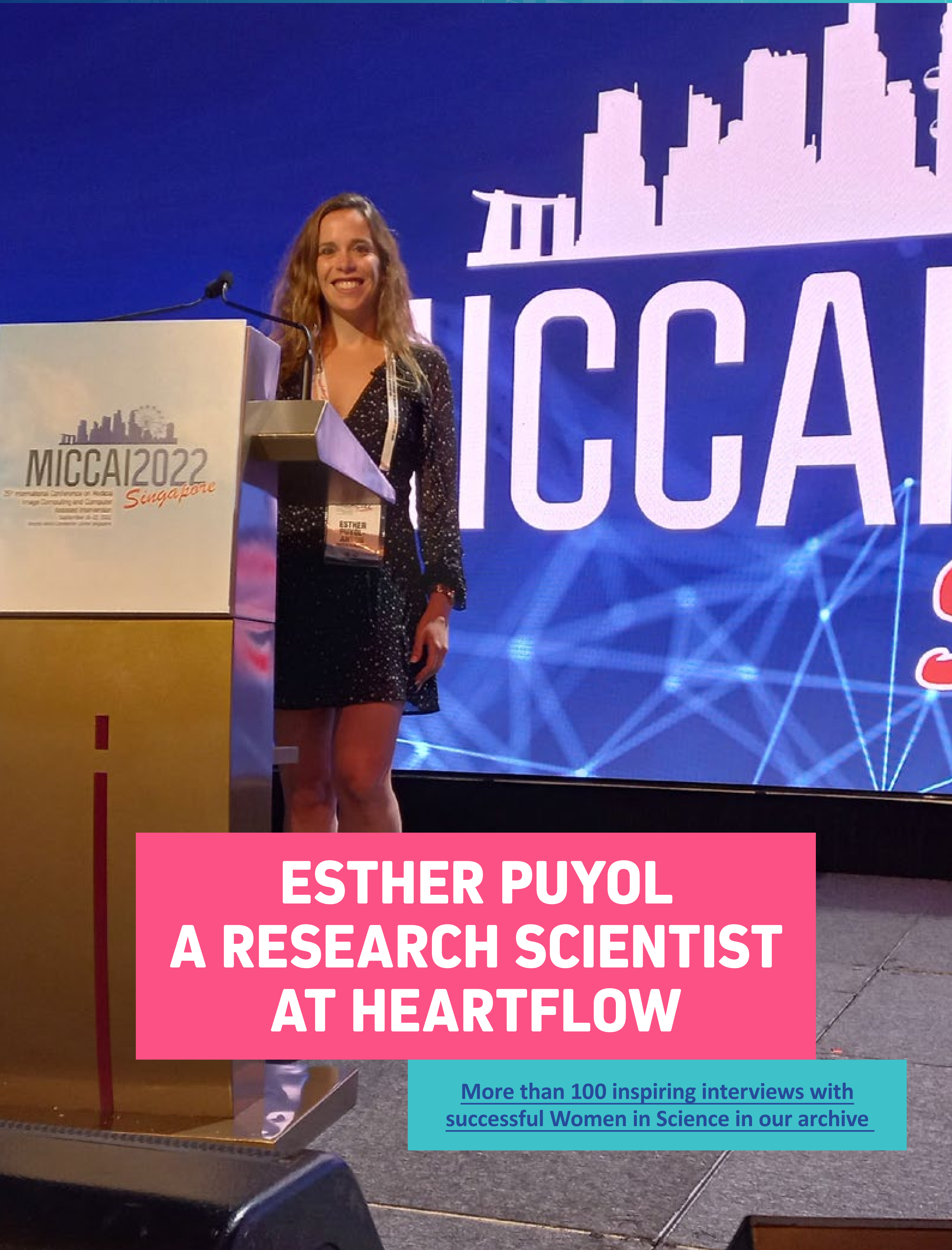
"Running these simulations took a while – I always had ten instances of Google Colab up at once to run every simulation with a bunch of different parameters," she tells us. *"I'd be sitting at my computer just going through all the tabs making sure that none of my Colabs had timed out. It could take more than eight hours to run them, and if I let them time out, I'd have to start over. If I ran them while I was asleep, I'd sometimes wake up in the middle of the night and have to get out of bed to check! It's funny looking back on it because I would wake up every morning and think, time to run my ten tabs of Colab before I start my day!"*



Ashley and Michael have been looking at the optimal arrangement of electrodes in epiretinal prostheses, but **their work could be generalized to other prostheses**. The field is relatively new, so there are many opportunities to explore.

There is lots of potential for this type of research to be used in the community to help patients, but it is still early days. Simulations do not always correlate precisely with real life, and patients have many different ρ and λ values that impact how the implant affects them, so more testing in the real world is needed first.

"We've shown there is merit in using computational models to design new implants and that the current electrode layout might not be the best." Michael points out. "In that *way, we're using insights from neuroscience to design better technology. We hope our work can help guide future prototyping of new devices."*



ESTHER PUYOL
A RESEARCH SCIENTIST AT HEARTFLOW

More than 100 inspiring interviews with successful Women in Science in our archive

Esther, what path brought you to HeartFlow?

I just started three months ago. Before that, I did my PhD and postdoc at King's College London, where I was developing new AI techniques for analysis of cardiac imaging – both ultrasound as well as MRI data.

How did you get into this field?

This was during my undergraduate in France, where I was learning about deep learning and AI – well, at that time, it was machine learning. Then I did an internship at Philips in France, and we started with some machine learning. I really liked it, it was in the area of cardiac, and then during my PhD, that was the main focus, developing new techniques for combining MRI and ultrasound for quantification of cardiac motion.

Did you want to be a doctor when you were a kid?

A little bit, yeah! *[she laughs]* I liked it, but then I also knew it was quite challenging and demanding, and I also liked the engineering side, so I think biomedical engineering is a good mix.

It's a good mix, and you're working to save people's lives.

[Esther smiles] Yeah, let's say it this way!

What attracts you to this field?

I was very lucky during my postdoc and PhD because I worked directly with doctors. We were trying to make something that would work in clinic, that people could use, and that may in the future help patients. That's what attracts me, especially thinking maybe one day I can make a contribution to clinics.

You're **just starting out, but do you think that during your career, you'll be able to help people in clinical settings or create technology that will help people feel better?**

I guess it's difficult to say, but I hope that one day I will develop some technology that could help – maybe not directly to save someone, but help the doctors make better predictions or better decisions that will benefit patients in the future.

Let's dream: What would 100% success be for you in your career?

[Esther thinks]

Nobel Prize in Medicine?

[she laughs] Exactly, that would be a really good dream! I think I'm a little bit less ambitious. I would be very happy to see a tool used in hospitals in different countries – not only in the UK where I'm based but also internationally.

You're not really English, are you?

No.

Where are you from?

Spain.

Where in Spain?

Barcelona.

Barcelona? You're Catalan?

Yes. *[she smiles]*

Okay, we've featured a few Catalans. We had Laura Leal-Taixé.

I've heard of her.

Why are you in the UK?

I did my first degree in Spain and then my second degree in France. When I was at Philips, they were collaborating with King's College as a PhD and told me, if you want, you can go there. That's how I ended up in the UK.

How nice, and you like it?

Yes, a lot.

Tell us one special thing about Spain, one about France, and one about the UK.

Well, I'm from Barcelona, so the beach, definitely! *[she laughs]* From France, I like the culture and the ambiance. I think it's really good. London, I don't know, I fell in love with London when I moved, so everything! *[laughs]*

Have you had a winter in London?

Oh, yeah, I have.

And you still like it?

[Esther laughs] Yes!

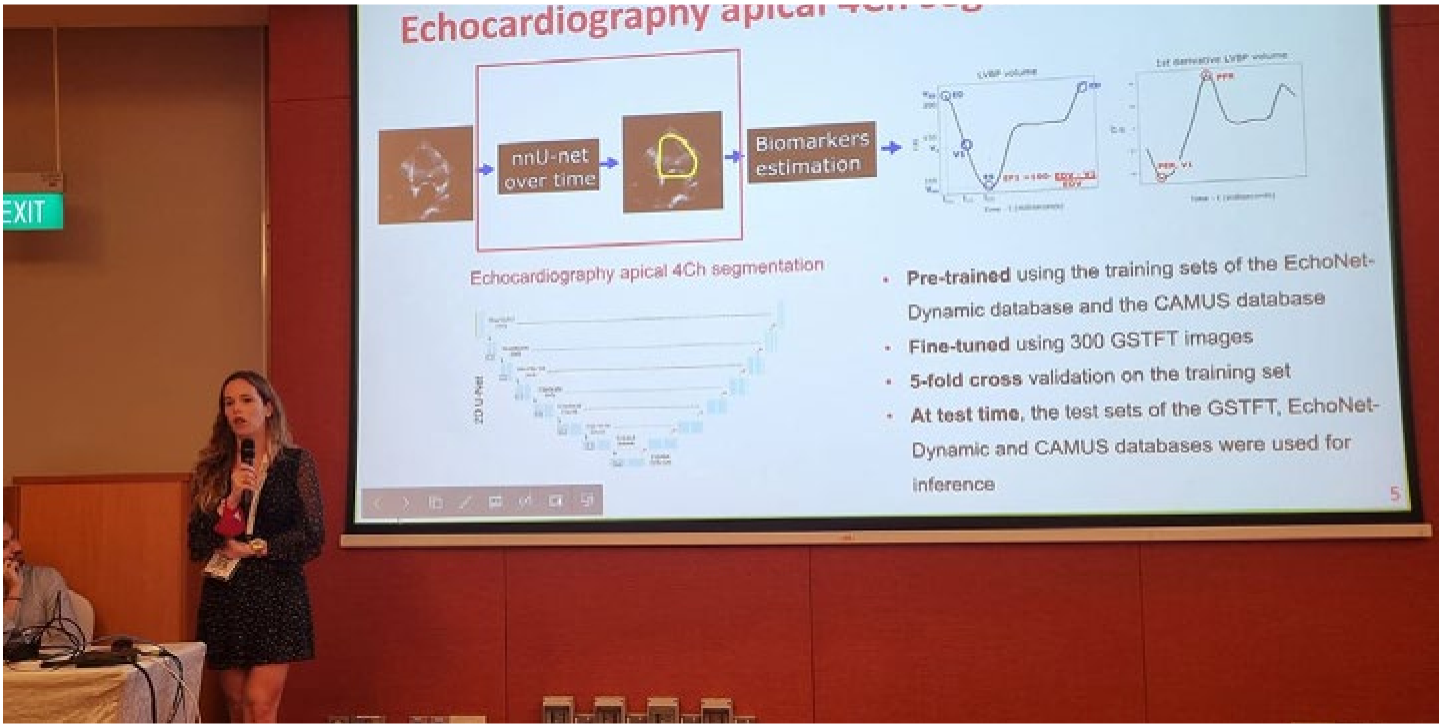
That's very nice. You are a special Catalan!

Yes, I know!

How is the university?

It's really good. I like everything about it. It's a really good community, and we work very closely together. Also, I like the way the department is structured because we're based in a hospital, so we have access to clinical data.

That reminds me of something Nassir Navab told me a few years ago. He wants his team to be in regular contact with doctors so that the solutions they develop



are very well anchored to the real world.

Exactly. That's exactly what I would like too, this very close relationship between doctors and engineers. You do something that they want, and they do something for you.

What is the thing that surprised you the most from your contact with real doctors?

Sometimes it's hard communicating. It's not the language itself, it's that we see things from different angles.

You're very certain of things, they're very certain, and you need to convey or make a language that both of you understand and learn from them the same way they need to learn from you.

Do you ever have the temptation to explain things to doctors?

Yeah, I try many times, and sometimes I succeed, and sometimes I fail! [*she laughs*]

Can they convince you?

In some areas, yes, because they're the experts. When they talk about pathophysiology, I need to trust them because it's not my area of expertise. The same way that when I talk about AI, they are more like, okay, if you say it.

If you had been a doctor, what area would you have chosen?

Cardiology, I guess.

What's in the heart that is special for you?

I like it because it's one of the organs that keeps moving throughout your life. It beats and contracts and expands. It's interesting how it's kind of the motor of all your body.

That's a nice definition. I'm sure that most cardiologists would agree with that. It's also very easy to find emojis of hearts.

The heart looks beautiful, right, in all the emojis? Also, it's about love.

What do you miss the most about Spain?

I think it's family and close friends. It's different being away. I was only with my partner at the beginning, so it was more

challenging than back home having a good background.

Are you Catalan or Spanish?

[Esther laughs] Both, let's say!

Okay, I will not ask anymore! [we both laugh] So, you miss family and friends – and food, certainly?

Well, we still cook a lot of Spanish food, but I miss fresh fruit and vegetables that taste nicer there.

Yours are the best tortillas in London?

Exactly, yes. I try!

What will be your next steps?

I just started in HeartFlow, it's only been

three months, so it will be nice to stay with them for a while. It's a new challenge, especially moving from academia to the industry.

What is one thing you need to learn from HeartFlow that you don't know yet?

This is a medical company and I've been mainly working in academia. It's very different. I would like to understand what the steps are to get from an idea to a product. How do you then sell it? Maybe one day I'll want to do it for myself, create something like a company, and I think it's important to understand every single step. How do you go from A to B, and what are the challenges and opportunities?

Do you think one day you'll be brave enough **to launch an idea of your own?**

That would be a good day. A good dream, yeah.

Are you not afraid?

Yeah, but you need to be brave in life, right? *[she laughs]*

Of course, tough ladies need to be brave.

Exactly.

Can you also envisage that something could go wrong, and you'd have to start over again?

Yes, exactly. But I guess if you have a strong background and experience, you can always go back to either industry or academia. You've got to try it at least.

It sounds like you feel lucky that you had what you had.

Yes, very lucky. From being given the opportunity to go to France to do the second degree and then, at that time, them contributing towards King's College for a PhD. I work very hard, but I also feel like I've been very lucky in the sense of the opportunities I've been given and the people who've supported me. I have good background support from my supervisors and everyone at King's that allows me to grow and become a better researcher.

Who are your supervisors?

Andrew King is the technical one, and Reza Razavi is the clinical one.

Of all the incredible things you've done to get to where you are, what was the biggest difficulty you had to overcome?

I think it was the step of trying to make what I was working on work in a clinical setting and all the challenges, from getting the data from the back systems to applying it to make something the clinicians wanted to use. That was the most challenging.

What forces did you find in yourself to overcome this challenge?

I guess it was the motivation. I really believed in the project. I thought it would be interesting. As I mentioned, I had a lot of collaborators who were very keen on helping and pushing – also, working very hard!

What's your secret?

I really believe in the projects I'm working on. I really think it will be something that can make a contribution, so I guess I have a lot of internal motivation. Also, I really like the imaging community and coming here for the conference and meeting with people.

If you had to select one achievement from everything you've done up to now, what would it be?

That's a very good question. Finishing my PhD was a big achievement, definitely. You work very hard for three or four years and at the end you have something which you can share with people and your family.

What did you do to celebrate?

I did two things. I decided that climbing Kilimanjaro in Tanzania was a good way to overcome the challenge! *[she laughs]*

You did?

Yes!

Wow!

Then, with friends and family, we had a small party all together.

What was most difficult, Kilimanjaro or making a PhD?

Both! *[we both laugh]*

Physically, Kilimanjaro, I guess intellectually, well… they are different!

What is your message to the community?

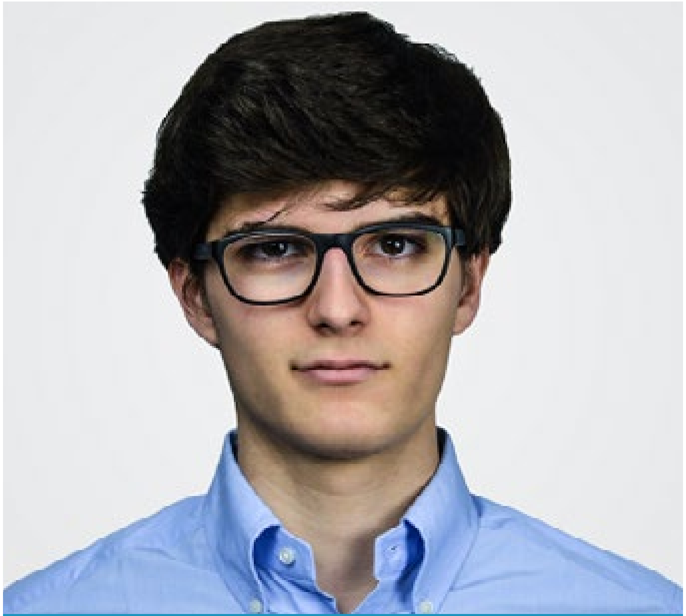
It's been a hard few years with Covid and everything, and it's really great to all be here again. I'm looking forward to chatting with everyone, getting to know each other again, and continuing to collaborate. I guess that's a positive message!

More than 100 inspiring interviews with successful Women in Science!


HAPPINESS IS…
HOLDING TICKETS
THAT SAY
VANCOUVER!
CVPR and MICCAI 2023 in Vancouver.
Come over. In-person. It will be fun!

D'ARTAGNAN: COUNTERFACTUAL VIDEO GENERATION

BEST OF MICCAI



Hadrien Reynaud

Thanos Vlontzos

Athanasios "Thanos" Vlontzos has just completed his PhD at Imperial College London and recently joined Spotify as a Research Scientist.

Hadrien Reynaud is currently studying for his PhD at Imperial.

Their paper explores a novel causal generative model applied to echocardiograms, and they spoke to us about it ahead of their poster session.

The ejection fraction is probably the most important metric in the clinical evaluation of cardiac function because it tells the clinician how much blood the heart pumps into the body for every beat. This paper proposes a novel neural network architecture that generates two videos for a given cardiac ultrasound scan, including one modified to change the ejection fraction.

"We want to let the clinician compare the true echocardiogram from the patient and an echocardiogram of the same patient with a different ejection fraction," Hadrien tells us. *"Same heart, same person, but different ejection fraction, which doesn't exist. It's purely hypothetical, but it lets the clinician see if the patient's heart is in bad condition, good condition, or too good condition, which can also be bad, by comparing the two results visually."*

Thanos adds:

"In real life, a cardiologist can't say, what if the patient had a 45% instead of a 40% ejection fraction? With our model, they can."

This work aims not to treat but to diagnose the patient. It allows them to see what problems they might have. It is abstract to talk about ejection fraction, but a visual

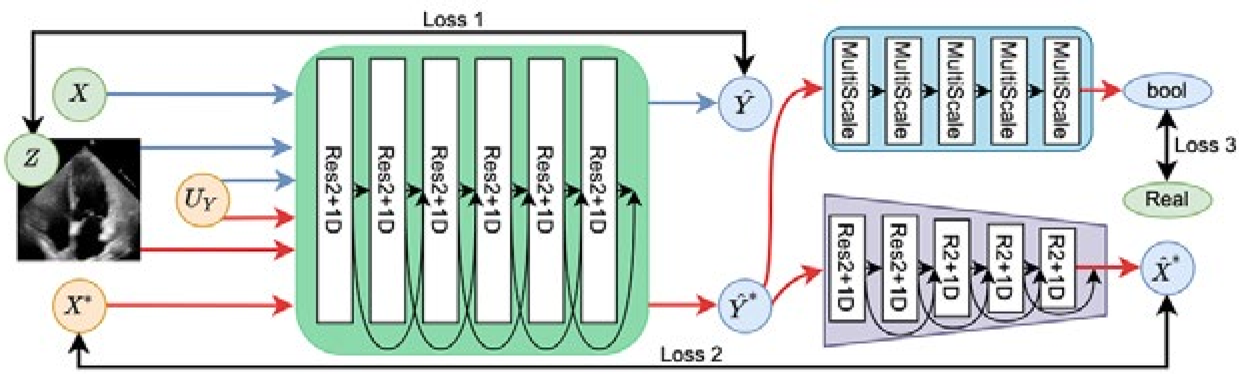


Fig. 1: The D'ARTAGNAN counterfactual framework. The green model is the generator, the blue model is the discriminator and the purple model is the expert LVEF regressor. The factual image flow is represented by the blue arrows and the counterfactual path by the red arrows.

This is a representation of the neural network we designed. It shows the factual and counterfactual flow of information, as well as the "trick" we used to train the counterfactual branch, ie: the combination of the expert model for the LVEF regression and GAN discriminator for visual quality.

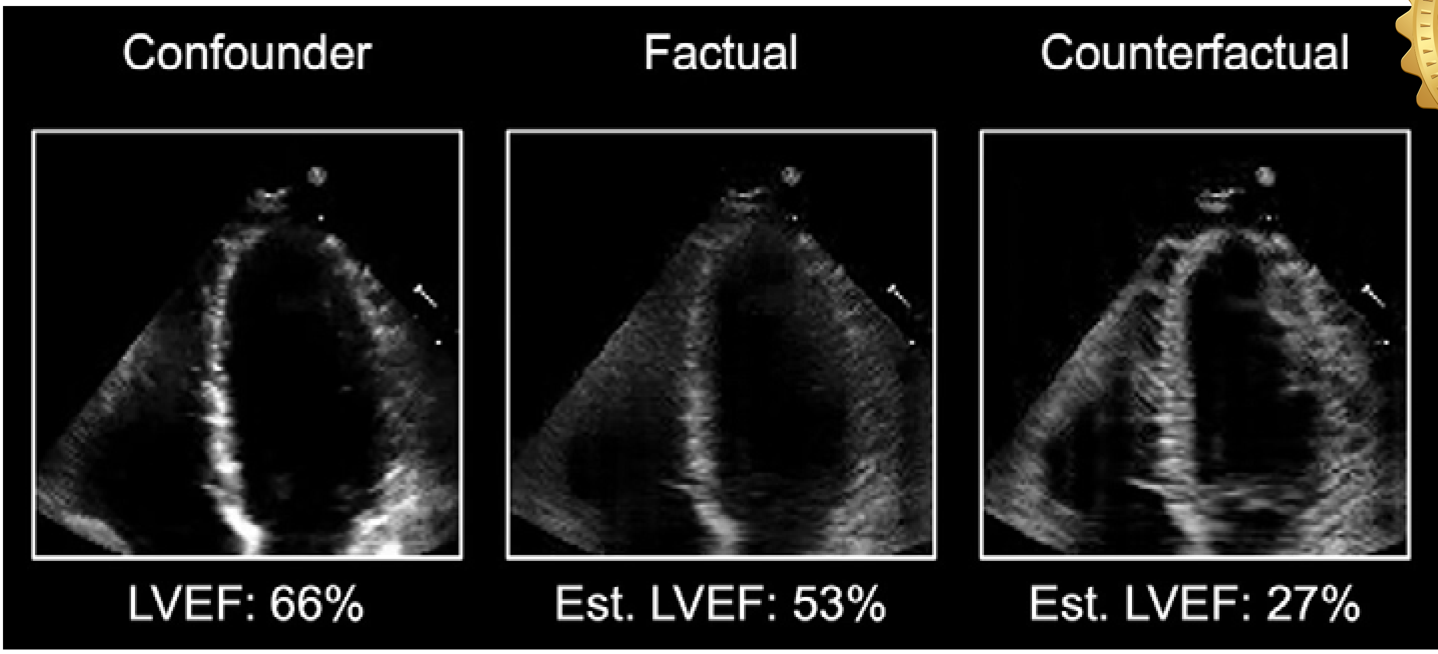
representation of what someone's heart is doing now and what it could be doing if it was healthier offers something tangible.

This work falls under the umbrella of counterfactual analysis. In the field of causality in machine learning, counterfactual questions explore alternative scenarios that might have happened had our actions been different. These scenarios can help inform clinicians by offering possibilities that may be hard for them to visualize.

"We've shown it in echocardiograms with the ejection fraction, but our method isn't only constrained there," Thanos points out. *"It could be a question like, for example, had I given the patient a different drug, would they have survived? We chose an echocardiogram with the ejection fraction because we felt it was clear-cut and visually easy for people to understand the power of these kinds of counterfactual questions, which otherwise they wouldn't be able to answer."*

Hadrien tells us that the biggest challenge on the technical side came from the fact that **the neural network is causal**. The network has two branches, the factual and the counterfactual, and to train that network, they needed ground truth to compute an error on the output of both branches. For the factual branch, that was not difficult because they had a video input and wanted the network to reconstruct that input. However, for the counterfactual branch, they had no way of generating what did not exist.

"That was the hardest part – we had to find a trick," he says. *"The solution we found was not to use videos as our ground truth, but instead use two neural networks to compute the loss on metrics that we wanted the network to learn. We wanted the counterfactual branch to generate a video that looked real and a video that had a different ejection fraction. We could train a network to ensure that the video produced looked like a real echocardiogram,*



This shows 3 triplets of echocardiograms: the input (confounder) and both outputs (factual and counterfactual). It demonstrates that the anatomy is preserved while the ejection faction changes (the amplitude of the movement is different).

and we knew how to make a network that would evaluate the ejection fraction in the counterfactual video, so we could backpropagate the loss through this expert network to enforce the ejection fraction."

On the theoretical side, the problem was that, by definition, the counterfactual was an alternative world they did not have access to but wanted to contemplate. Even figuring out a video would look a certain way, they had no way to tell if it was true. They had to apply some stringent conditions to allow them to make such claims.

*"Those conditions are essentially called **identifiability conditions,"*** Thanos explains. *"The question here is, can we identify the correct counterfactual from the data we observe? There are an enormous number of potential counterfactuals. In our case, because the echocardiograms and the videos were difficult to identify, we used a combination of theoretical constraints, plus we went to cardiologists and said, 'Does this look reasonable to you?' to inform our decision process."*

This process is not unique to machine learning but used in many fields that feature causality and counterfactuals, such as econometrics and epidemiology. It is always important to be careful and strict about any assumptions, how they are made, and how they are implemented. Results could be wrong, but theoretical constraints and expert knowledge can help mitigate this.

Did anything go wrong? Or did they see any surprising results?

"We were curious to see how the model would perform under extreme conditions, so we did produce some videos where, in real life, the patient would be dead!" Thanos laughs nervously. *"The video itself was able in terms of how the heart would react if the ejection fraction were, for example, 10%. However, a human being cannot handle a 10% ejection fraction, but that's a completely different story."*

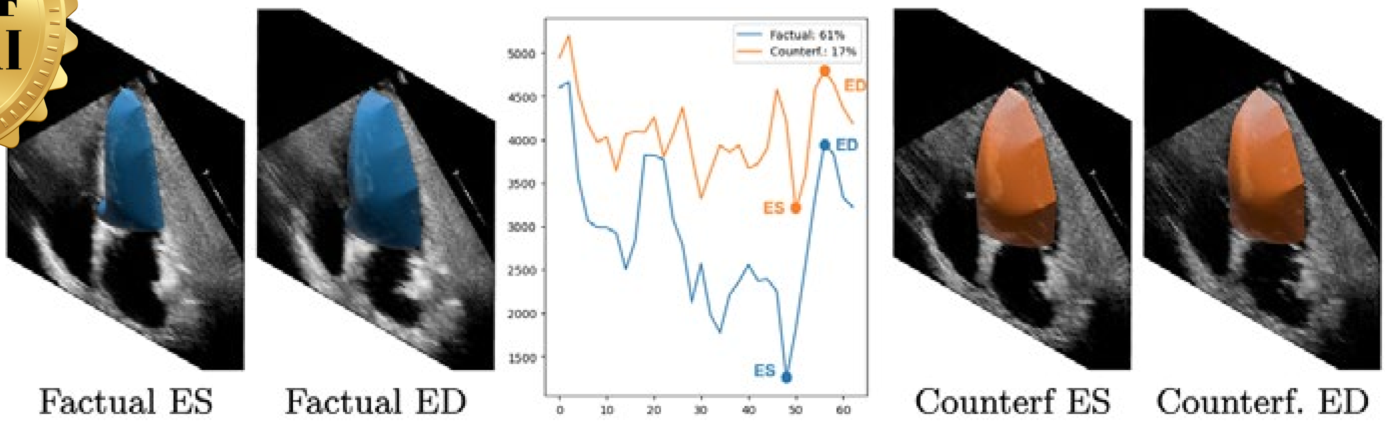


Fig. 3: Qualitative results for D'ARTAGNAN over the same confounder and noise. Left: factual ES and ED frames. Middle: left ventricle area over time, obtained with a segmentation network. Dots represent where the corresponding frames were sampled. Right: counterfactual ES and ED frames. Anatomy is preserved across videos, while the LVEF fraction is different.

This is a static image representing 2 key moments of the cardiac cycle, called the end-systole (ES) and end-diastole (ED) for the factual (=real world) echocardiogram on the left and the counterfactual (=fake, created by the model) echocardiogram on the right. The image shows that the model changed the real echocardiogram to lower the ejection fraction while retaining the anatomy of the patient in the original echocardiogram.

Co-authors of this paper include experts in causality and cardiology who have guided Thanos and Hadrien on this topic.

"When we were brainstorming how to approach the subject, the cardiologist told us that what we were proposing was interesting and novel," Hadrien reveals. *"They were keen to see what could be done and where it could be used. It was very encouraging."*

Bernhard Kainz, Reader in the Department of Computing at **Imperial** and one of the paper's co-authors, gives us his perspective:

"I'm heavily biased, of course, but this work is novel ***because it's the first MICCAI paper that shows how causality research and machine learning can be connected to provide new tools for diagnostic support.*** *Doctors ask, 'How would a patient's scan look if clinical parameter X was different?' We show several theoretical ways to learn and generate such counterfactuals and a practical application for diagnosing cardiac function with ultrasound."*

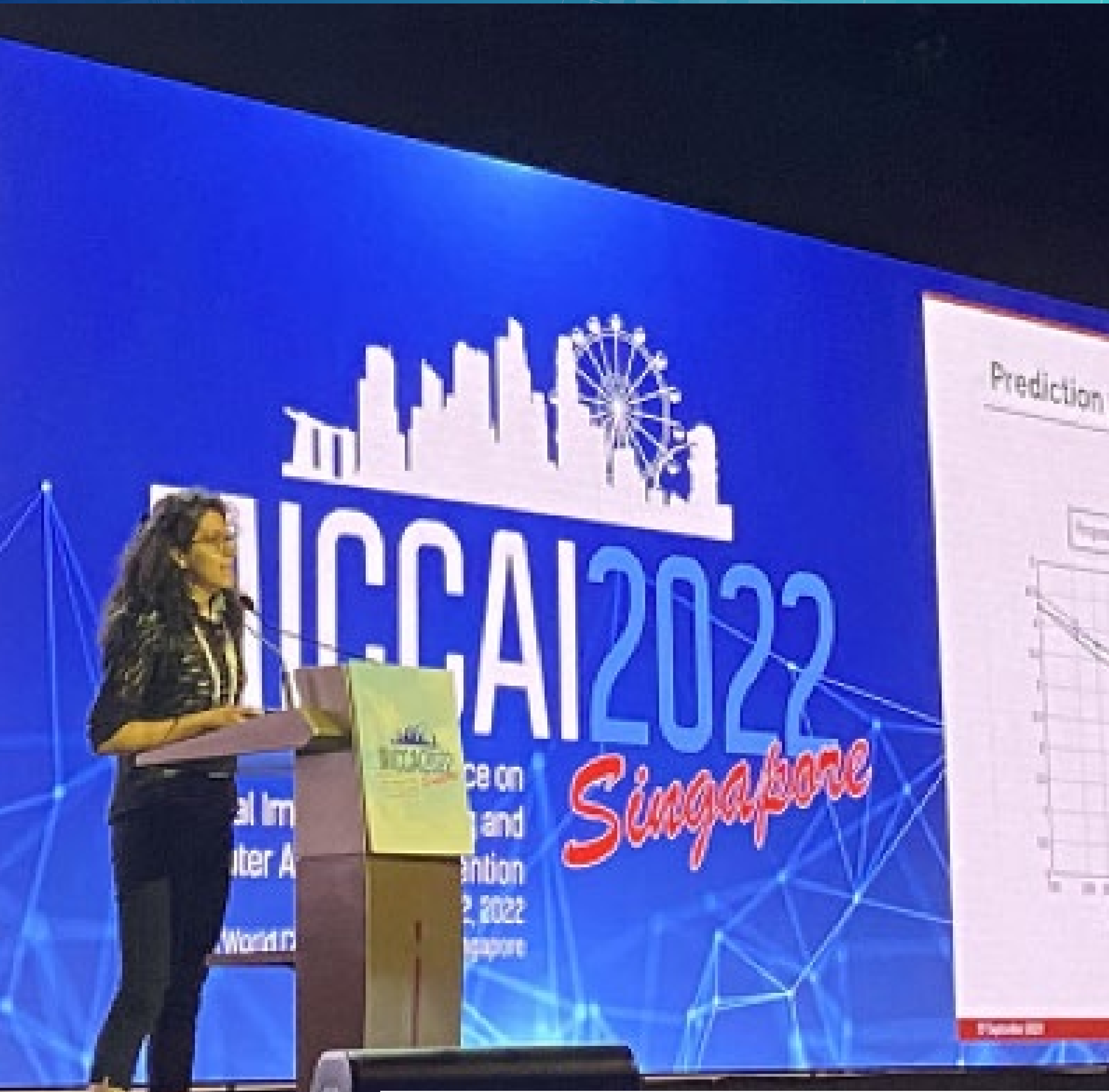
Thanos and Hadrien are already considering the next steps and are keen to iron out a small flaw they have identified in their approach.

"The causal framework is supposed to have variables which are completely independent of each other," Hadrien points out. *"In our case, the factual and something we call the confounder share some information. We're thinking of ways to change this. It's challenging and implies changing many things in the model, but that's one of the flaws. It's a work in progress for us."*

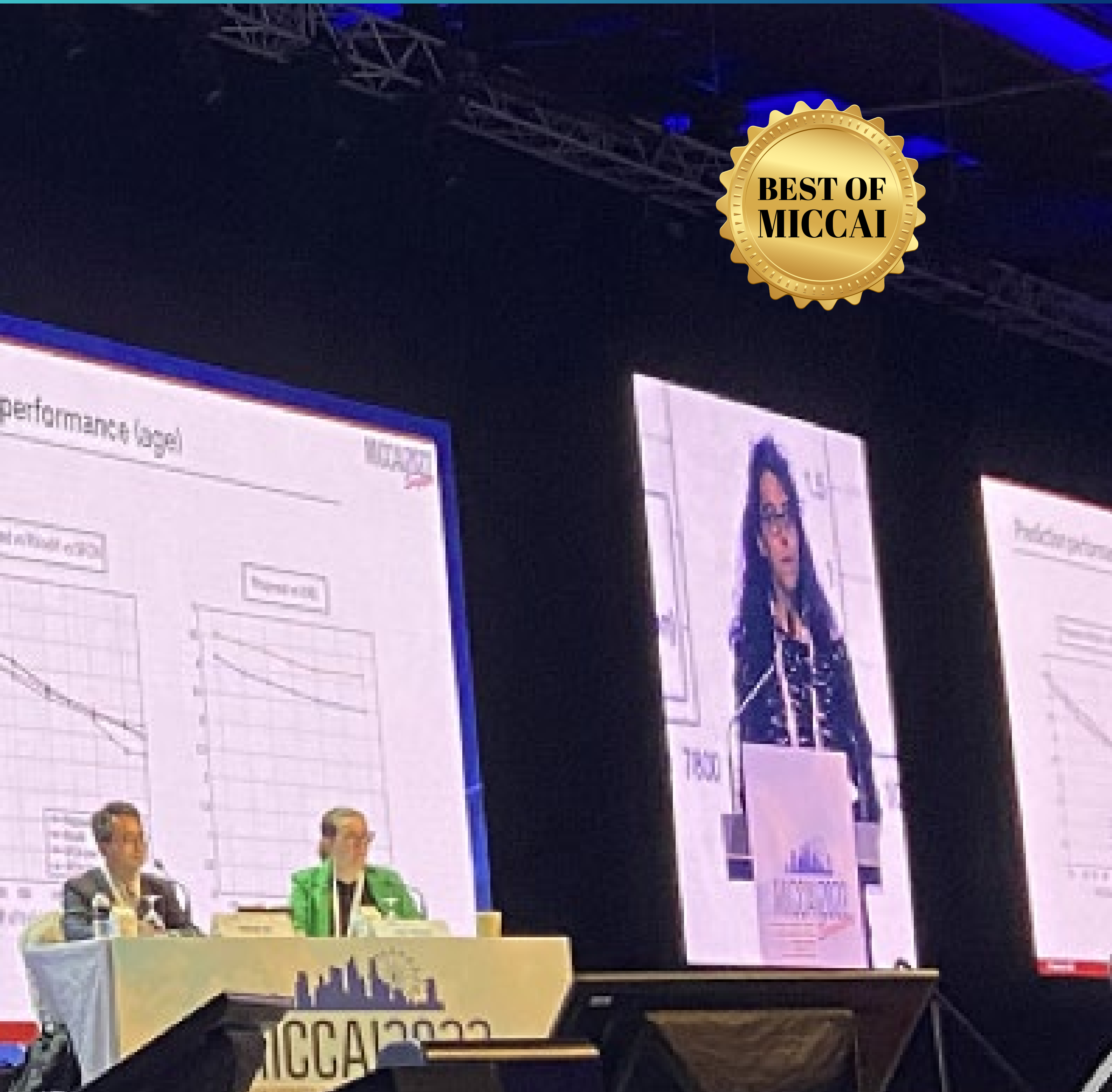
Could we see the result at MICCAI in Vancouver next year?

"We hope so!"

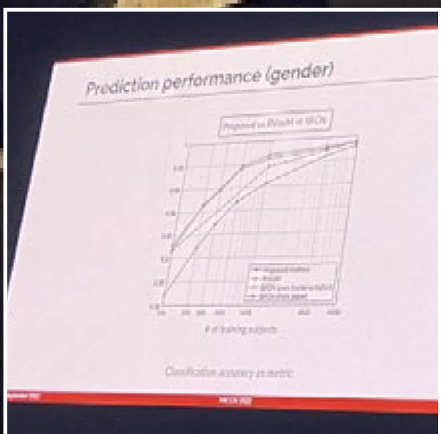
Oral Presentation



Chiara Mauri is a soon-to-be PhD graduate at **Technical Univ**
Denmark under the supervision of **Koen Van Leemput**. Ch
an oral presentation about her paper "**Accurate and Exp**
Image-based Prediction Using a Lightweight Generative M
novel way to obtain interpretable predictions based on bra
without sacrificing prediction accuracy.



versity of
iara gave
plainable
Model", a
ain scans,



Developing for the Medical AI Project Lifecycle

Michael Zephyr is a Developer Evangelist for Healthcare at NVIDIA and one of the Group Leads for the MONAI Adoption and Outreach Working Group.

Stephen Aylward is the Senior Director of Strategic Initiatives at Kitware. He is also a MICCAI Fellow and an adjunct professor of Computer Science at the University of North Carolina.

An exciting journey started three years ago when **NVIDIA** and **King's College London** came together during **MICCAI 2019** and formed **Project MONAI** as an initiative to **develop a standardized, user-friendly, and open-source platform for Deep Learning in Medical Imaging**. Soon after that, they established the MONAI Advisory Board and Working Groups with representatives from Stanford University, National Cancer Institute, DKFZ, TUM, Chinese Academy of Sciences, University of Warwick, Northwestern University, Kitware, and Mayo Clinic.

Throughout this journey, MONAI has deepened its offering in radiology, expanded to pathology, and most recently, included support for streaming modalities starting with endoscopy. Now three years later, MONAI has over 600,000 downloads. It is used in over 450 GitHub projects, has been cited in over 150 published papers, and academic and industry leaders are using MONAI in their research and clinical workflows.

We're excited to announce that MONAI is **continuing to expand open-source healthcare AI innovation** with v1.0. With a focus on providing a robust API that is designed for backward compatibility, this release ensures that you can integrate MONAI into your projects today and benefit from the stability of an industry-leading framework into the future.

Let's look at the features included in the MONAI Core v1.0, MONAI Label v0.5 releases, and a new initiative called the MONAI Model Zoo.

MONAI Core v1.0

With the release of v1.0, MONAI Core focuses heavily on a robust and backward-compatible API design and also includes additional features like MetaTensors, a Federated Learning API, the MONAI Bundle Specification, and an Auto3D Segmetentation framework.

MetaTensor

The MetaTensor enhances the metadata-aware imaging processing pipeline by integrating both torch tensors and imaging meta-information. This combined information is essential for delivering clinically useful models, supporting image registration, and joining multiple models into a cohesive workflow.

MONAI MetaTensor Docs

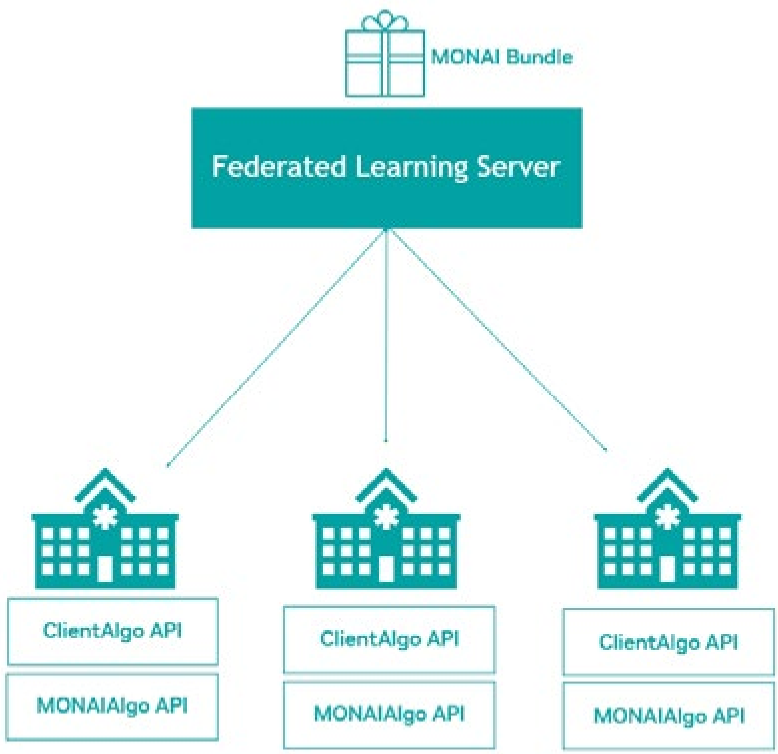
Federated Learning

The MONAI Federated Learning module provides a base API that defines a MONAI Client App that can run on any federated learning platform. With the new federated learning APIs, you can utilize MONAI bundles and seamlessly extend them to the federated learning paradigm.

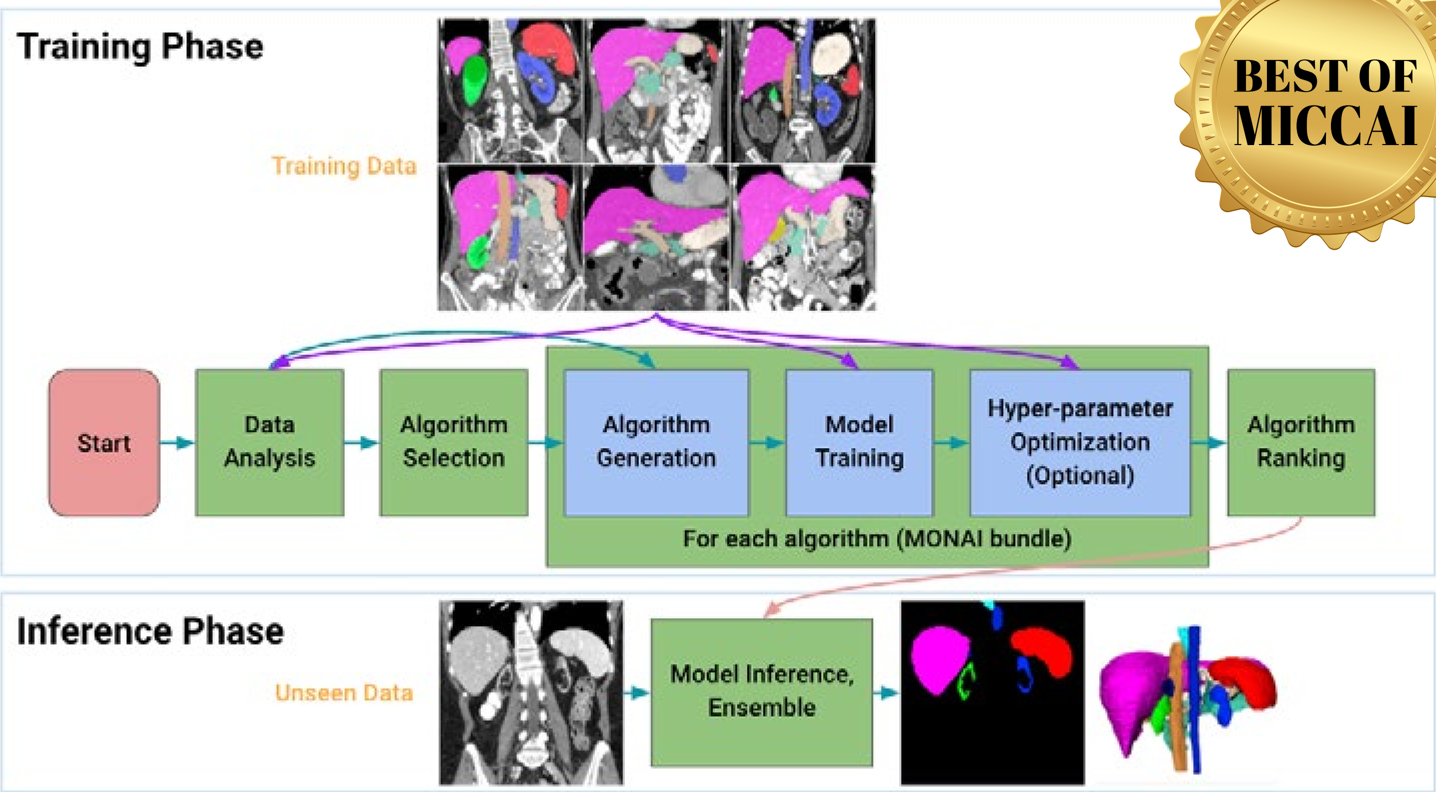
The first platform to support these new Federated Learning APIs is NVIDIA FLARE, the federated learning platform developed by NVIDIA. We welcome the integration of other federated learning toolkits to the MONAI Federated Learning APIs to help build a common foundation for collaborative learning in medical imaging.

MONAI Federated Learning Docs

NVIDIA FLARE + MONAI Example



MONAI and Federated Learning high-level workflow using the new MONAIAlgo FL APIs



Auto3D Segmentation Training and Inference workflow

BEST OF MICCAI

Auto3D Segmentation

Auto3D is a low-code framework that allows data scientists and researchers of any skill level to train models that can quickly segment regions of interest in data from 3D imaging modalities like CT and MRI.

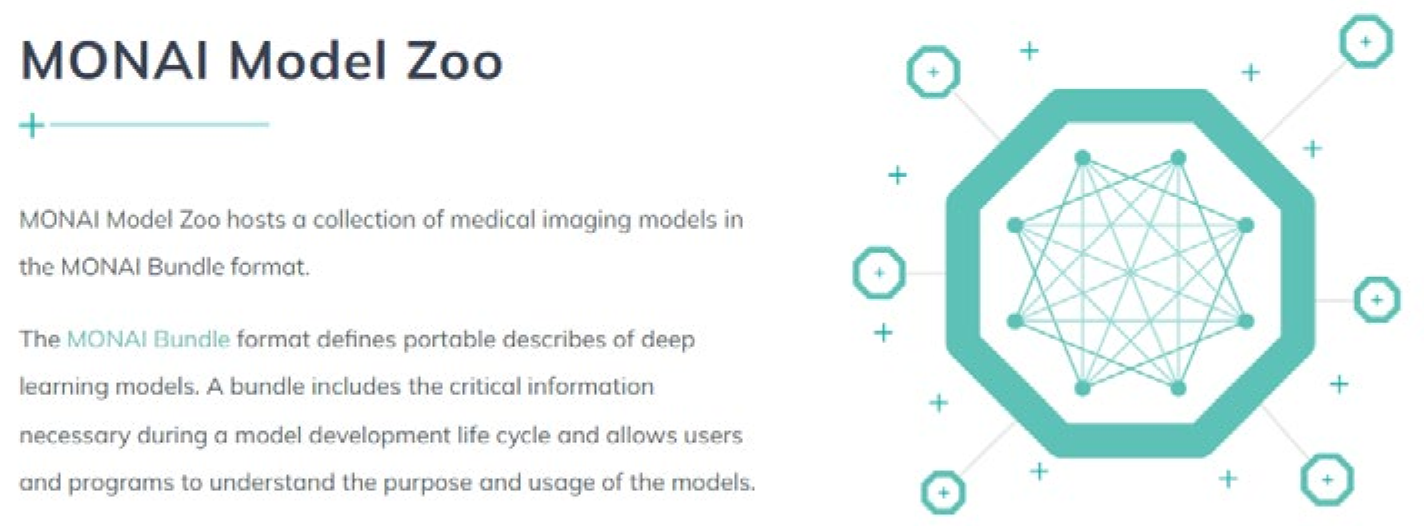
Developers can start with as little as 1-5 lines of code, resulting in a highly accurate segmentation model. By focusing on accuracy and including state-of-the-art models like Swin UNETR, DiNTS, and SegResNet, data scientists and researchers can utilize the latest and greatest algorithms to help maximize their productivity.

Auto3D Tutorial

MONAI Model Zoo

We're excited to announce the **MONAI Model Zoo**, a hub for sharing pre-trained models that allow data scientists and clinical researchers to jump-start their AI development.

In the first release, there are 15 pre-trained models from MONAI partners, including King's College London, Charité University, University of Warwick, Vanderbilt University, and Mayo Clinic.



MONAI Model Zoo Landing Page

These models utilize the MONAI Bundle specification, making it easy to get started in just a few commands. With the MONAI Bundle and Model Zoo, we hope to establish a common standard for reproducible research and collaboration, and we welcome everyone to contribute to this effort by **submitting** their pre-trained models for downstream tasks.

MONAI Label v0.5

MONAI Label now supports MONAI Core v1.0 and continues to evolve by improving overall performance, offering new models for radiology, and expanding into endoscopy with integration into the CVAT viewer for annotation and releasing new endoscopy models.

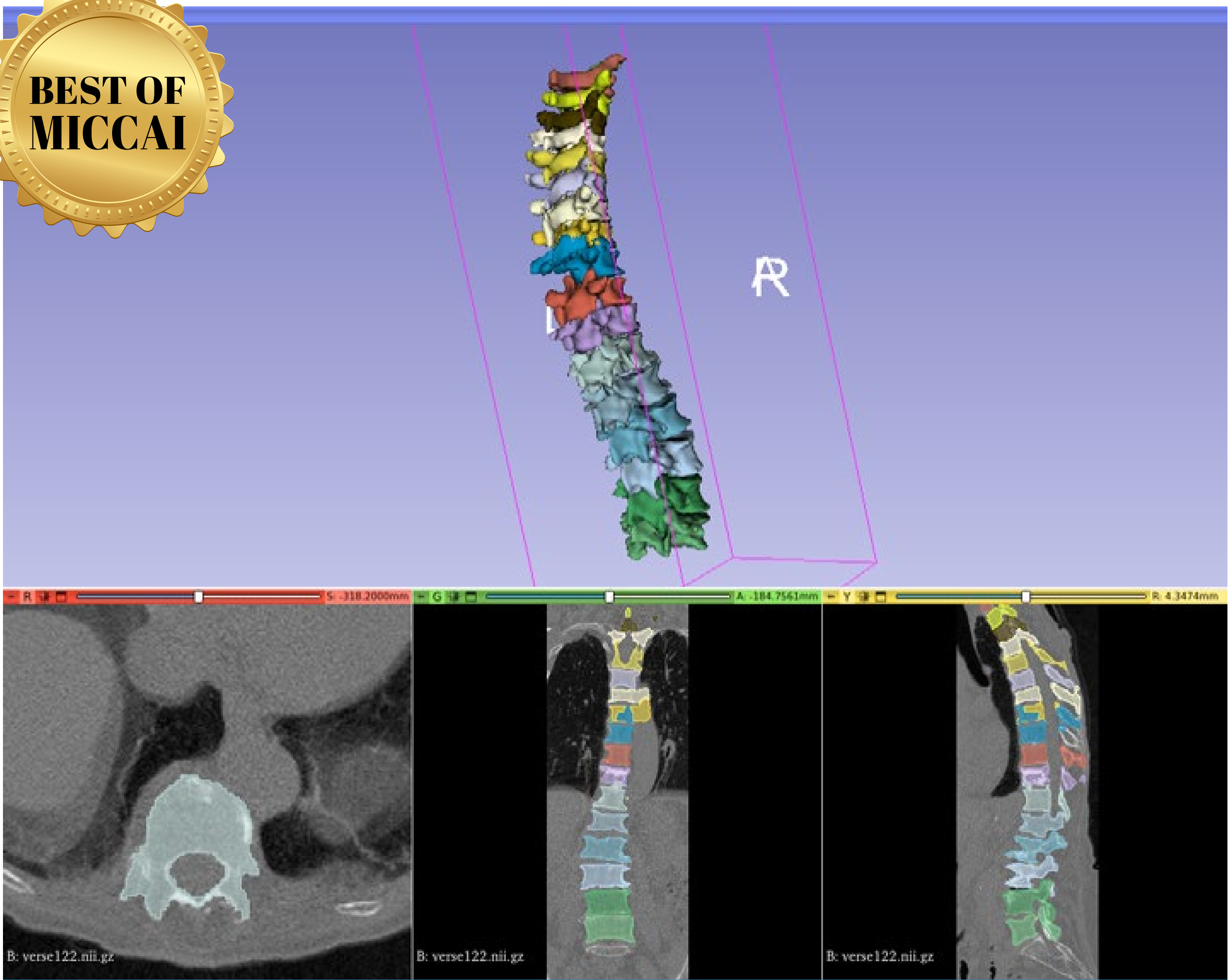
MONAI Label has been updated to support MONAI Core v1.0. For radiology, we've focused on improving overall performance and released a new vertebra model. For endoscopy, we've continued to improve CVAT integration and released three new models.

Endoscopy

MONAI Label now supports 2D segmentation for endoscopy. By continuing to expand on the previous CVAT integration, MONAI Label has integrated active learning into CVATs automated workflow.

Three new models are being released including Tool Tracking Segmentation, InBody vs. OutBody De-identification classification model, and DeepEdit for interactive tool annotation.

MONAI Label Endoscopy Sample Applications



MONAI Label's new Vertebra model segments several structures in CT images, shown running in 3D Slicer.

Conclusion

This is a momentous milestone for Project MONAI and we are looking forward to further serving the medical imaging community. We want to hear your feedback! Connect with us on **Slack** and **GitHub**. Please share your successes and report any issues you might have with MONAI.

Interested in joining the MONAI Community? Get started on our **MONAI YouTube Channel**, where we have tutorials, archived bootcamps, and walkthrough guides.

Stay tuned for the latest news on our hosted events! Whether you're new to MONAI or already integrating MONAI into your workflow, the **MONAI Website** and **Twitter** account are the best places to stay up to date!

BACK TO LIVE WORKSHOPS!

MAPPING IN CYCLES: DUAL-DOMAIN PET-CT SYNTHESIS FRAMEWORK WITH CYCLE-CONSISTENT CONSTRAINTS

BEST OF MICCAI

Jiadong Zhang

Jiadong Zhang is in the second year of his master study in the IDEA Lab at the School of Biomedical Engineering, ShanghaiTech University, supervised by Dinggang Shen.

His paper explores a novel PET-CT synthesis framework. He spoke to us ahead of his poster presentation.

PET is a vital medical imaging technique for people suffering from cancer and brain disorders. CT is needed for anatomical localization and also PET attenuation correction during PET scanning. But acquisition of CT increases risk of radiation exposure.

This work proposes using PET image to synthesize CT image for non-diagnosis purpose. Most existing works use learning-based image synthesis to build cross-modality mapping in the image domain, without considering the projection domain, leading to inconsistency between information in the two domains.

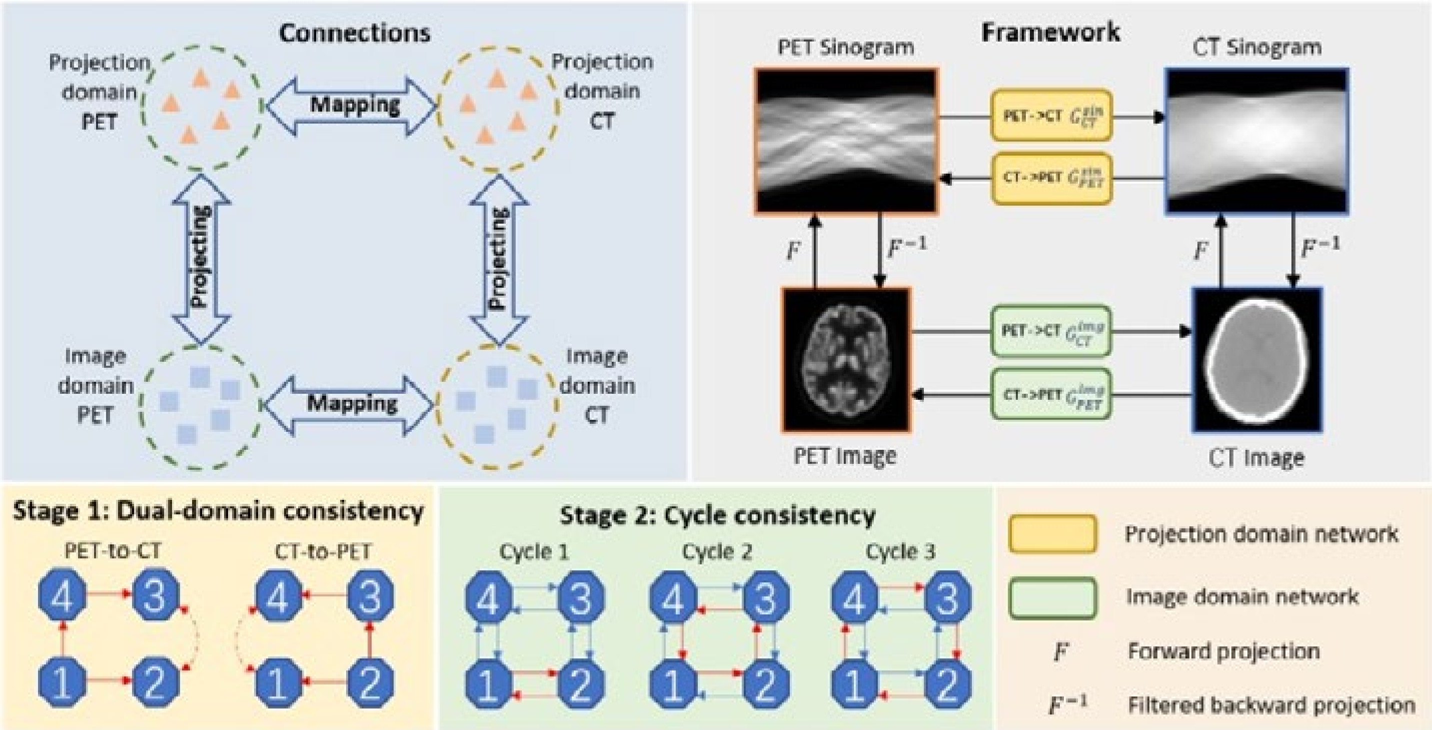


Fig. 1. The overview of our proposed dual-domain PET-CT synthesis framework. The 1st figure illustrates the connection between dual domains and two modalities. The 2nd figure illustrates the overall framework architecture. The 3rd figure illustrates dual-domain consistency in Stage 1. The 4th figure illustrates cycle consistency in framework in Stage 2.

"Our framework uses dual-domain knowledge in two different modalities, PET and CT, and keeps dual-domain consistency in the synthesis process," Jiadong explains.

"To the best of our knowledge, ours is the first paper to exploit dual-domain information for two different modalities. Other works just use one modality, such as low-dose and standard-dose CT. We hope this work can make a significant contribution to cross-modality synthesis tasks."

The framework combines four neural networks with a forward projection and a filtered back projection, using bidirectional mapping with multiple closed cycles. These cycles can further serve as cycle-consistent constraints to keep the anatomical structures consistent in the synthesis process for better performance.

"Synthesising CT image from PET image is challenging because some anatomical structures are not visible in PET image," Jiadong tells us.

"When using other medical image synthesis methods, such as CycleGAN, the anatomical structure in dual-domain is inconsistent, negatively impacting the results. Our work addresses this issue."

Jiadong points out that they have designed a general framework that could be used for many other applications.

"So far, we have only explored our framework on PET-CT synthesis," he says.

*"Next, we want to explore the performance

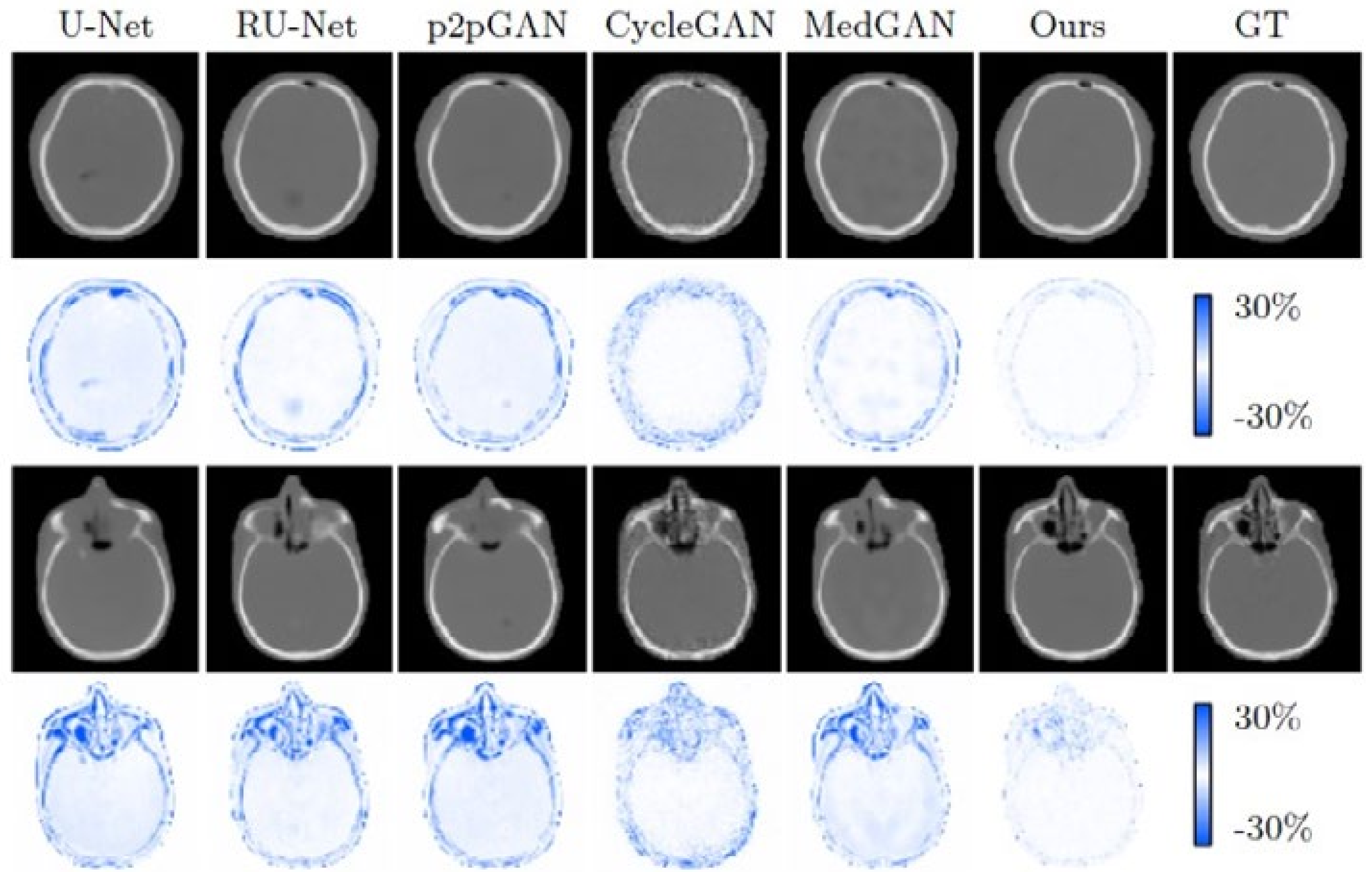


Fig. 2. Visual comparison of synthesized CT images by different methods on two typical samples. For each sample, the first row shows CT images, and the second row shows the corresponding error maps.

of this method on other tasks, such as low-dose PET reconstruction, low-dose CT reconstruction, and multimodal MR synthesis. We will report related results in our future journal paper."

Or a paper at MICCAI next year?

"Yes, maybe!"

Before we let him go, Jiadong is keen to speak about his supervisor, **Dinggang Shen**.

"Dinggang Shen is a great professor who has taught me many research and life sessions," he smiles.

"In research, I have learned a lot from Prof. Shen in each course, seminar, and group meeting, including how to be critical thinking, write a good paper, and make an impressive presentation. More importantly, he always emphasizes that we should work closely with physicians for doing clinically significant works. He says all good researches always come from real clinical problems and then apply back to solve clinical problems. Also, he provides me with multiple excellent chances to work or communicate with many talented people, which makes me enjoy the journey of my master study. Generally, I am growing very fast in the ***IDEA*** *Lab, and I am glad that I am doing such meaningful health and medicine-related work here."*

BACK TO POSTERS AGAIN

STATISTICAL ATLASES AND COMPUTATIONAL MODELING OF THE HEART

BEST OF MICCAI

By Marica Muffoletto

Marica is currently working at her CDT in Smart Medical Imaging at King's College London.

She is also Engineering Editor at Computer Vision News since 2020.

Find her on Twitter @maricas8.

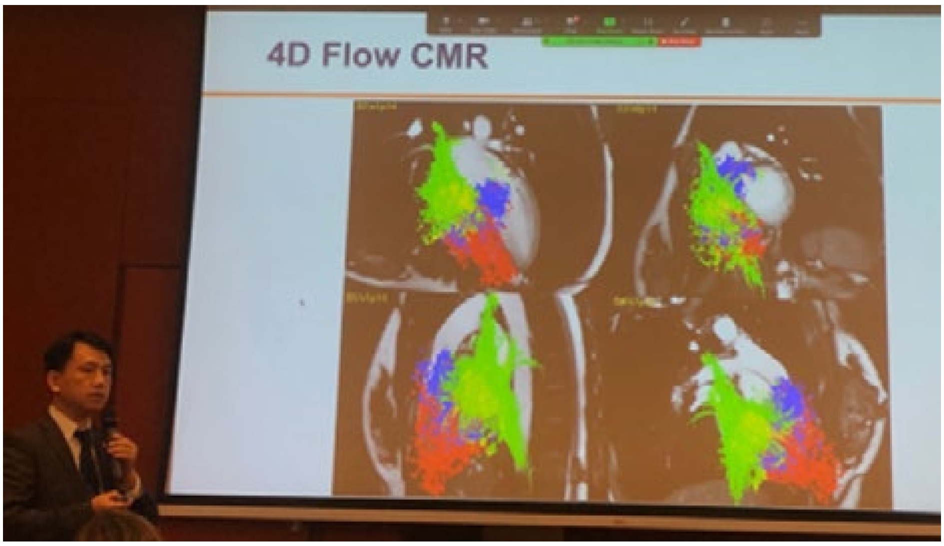
MICCAI has finally started here in Singapore. The first day hosts a bunch of workshops, and I was visiting STACOM - **Statistical Atlases and Computational Modeling of the Heart**, which has been around for 13 years now, and is organized by a great team of researchers- Alistair Young (King's College London), Maxime Sermesant (Inria, France), Oscar Camara (Universitat Pompeu Fabra, Spain), Esther Puyol Antón (KCL), and Avan Suinesiaputra (KCL).

We started with a talk from **Victoriya Kashtanova** on cardiac electrophysiological modelling who combined single EP Mitchell-Schaeffer models with a DL-based approach (ResNet) to achieve personalization of the action potential. **Carlos Albors Lucas** followed with his work on computational models to analyze haemodynamics from dynamic CTs. He presented his method DM-DCT, highlighting the importance of LA wall motion for thrombus estimation. **Lei Li** showed how to combine CMR, ECG and patient metadata (age, sex, BMI) using a VAE and a point cloud input to predict conduction velocity and root nodes, to identify myocardial infarction. After **Xiaoran Zhang** who talked about his project on Biomechanics informed Modeling for cardiac image registration,

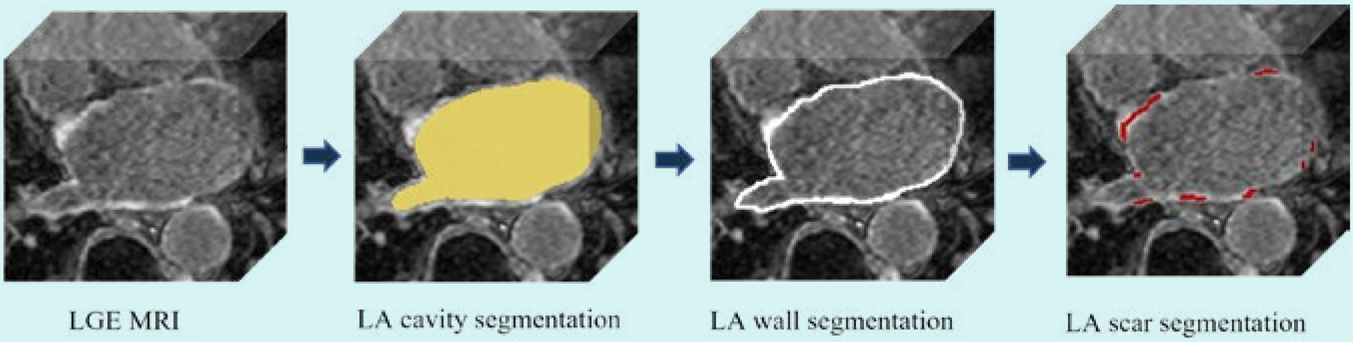
we ended the first session with ("finally!" cit. from organizer) the first talk of the day on statistical shape modelling led by **Jadie Adams**.

Jadie's work, called **Spatio-temporal Cardiac Statistical Shape Modeling: A data driven approach** is the perfect ending to a very interesting section and she indeed won a well-deserved prize for it! In her research, she focuses on spatio-temporal data and how to generate point distribution models (PDM) by an optimization scheme across shape & configuration spaces using a software called **ShapeWorks**. Simultaneously optimizing across subjects and time points, she achieves astounding results. After she showed us that generative modelling can generalize her proposed approach by capturing the temporal projection of space, we really hope to see follow up work on incorporating latent diffusion modelling (LDM) to her pipeline.

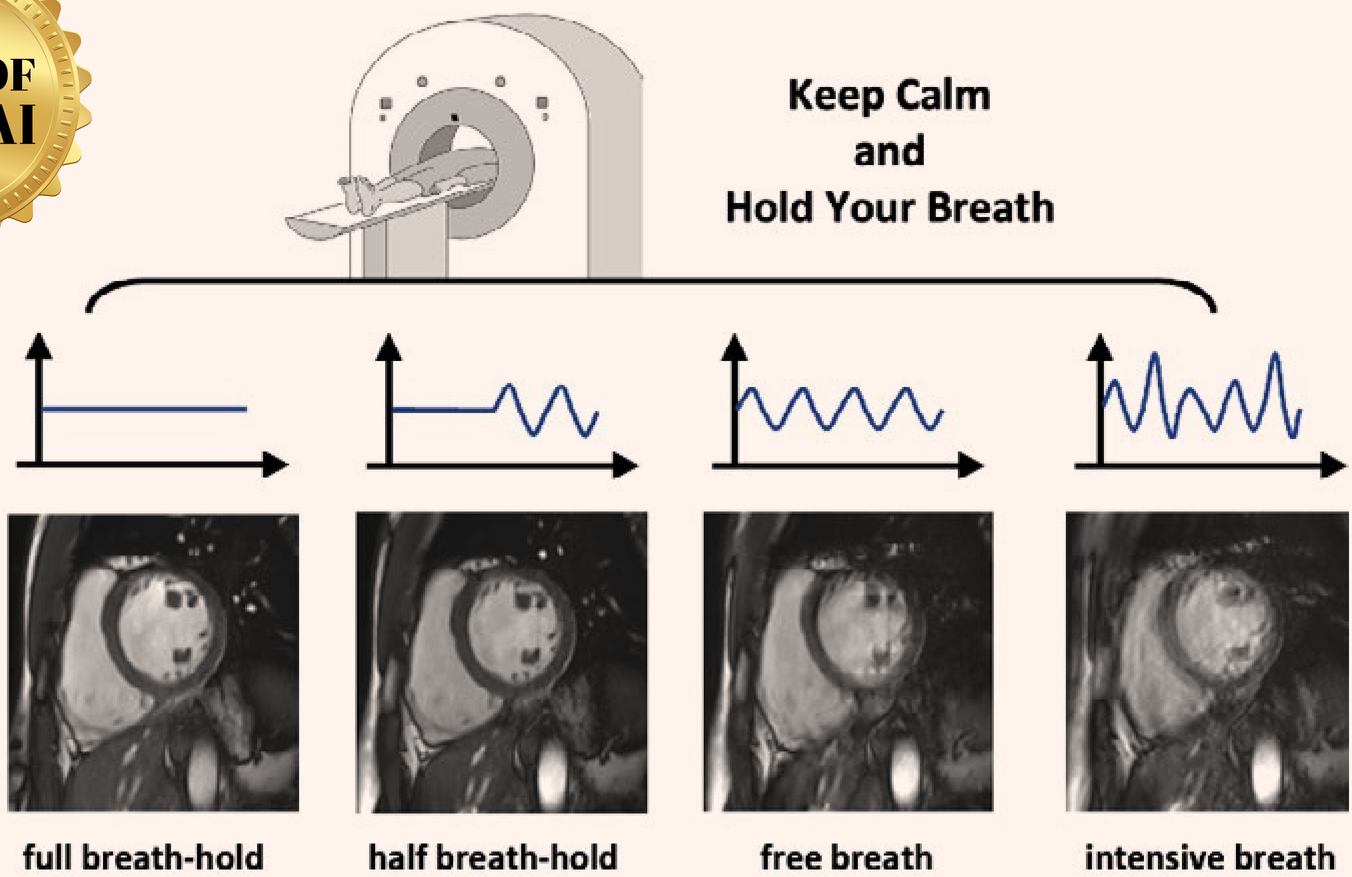
Next, the local researcher and Keynote speaker **Liang Zhong** (below) introduced his project **INITIATE**. His talk focused on the importance of diagnosing and treating **Congenital Heart Diseases (CHD), a lifelong chronic condition, and on the significance and challenges of a** correct analysis of the Right Ventricular function. INITIATE is born with the intent of filling the lack of an Asian centric database for CHD and aims to build 4 main toolkits – 1) segmentation and modelling, 2) Feature Tracking, 3) 4D Flow and 4) Quantitative markers.



The LAScarQS 2022 Challenge organised by **Xiahai Zhuang** from the Fudan University (China) focuses on the LA wall segmentation and the scar quantification.



Yuchen Zhang, winner of the first position, showed his boundary-focused nnUNet, where he combined 2 nnUNets, a TopK + Dice Loss and a signed distance map of the boundaries to explicitly model spatial relationship. He was followed by the author of **UGformer**, a 2-stage segmentation model made of an enhanced transformer block and a GCN-based structure to optimize the global space of intermediate feature layers. **Sihain Wang** showed his multi-depth segmentation network to address variation in scar size and shapes with an additional plug-and-play Sobel fusion model, which extracts LA boundary information to improve scar segmentation. The other discussed techniques included a Multi-scale Weight Share Network **(MSWS-Net)**, which employs a UNET modified to extract features at multiple scales and **LASSNet**, a 4 steps DNN model.



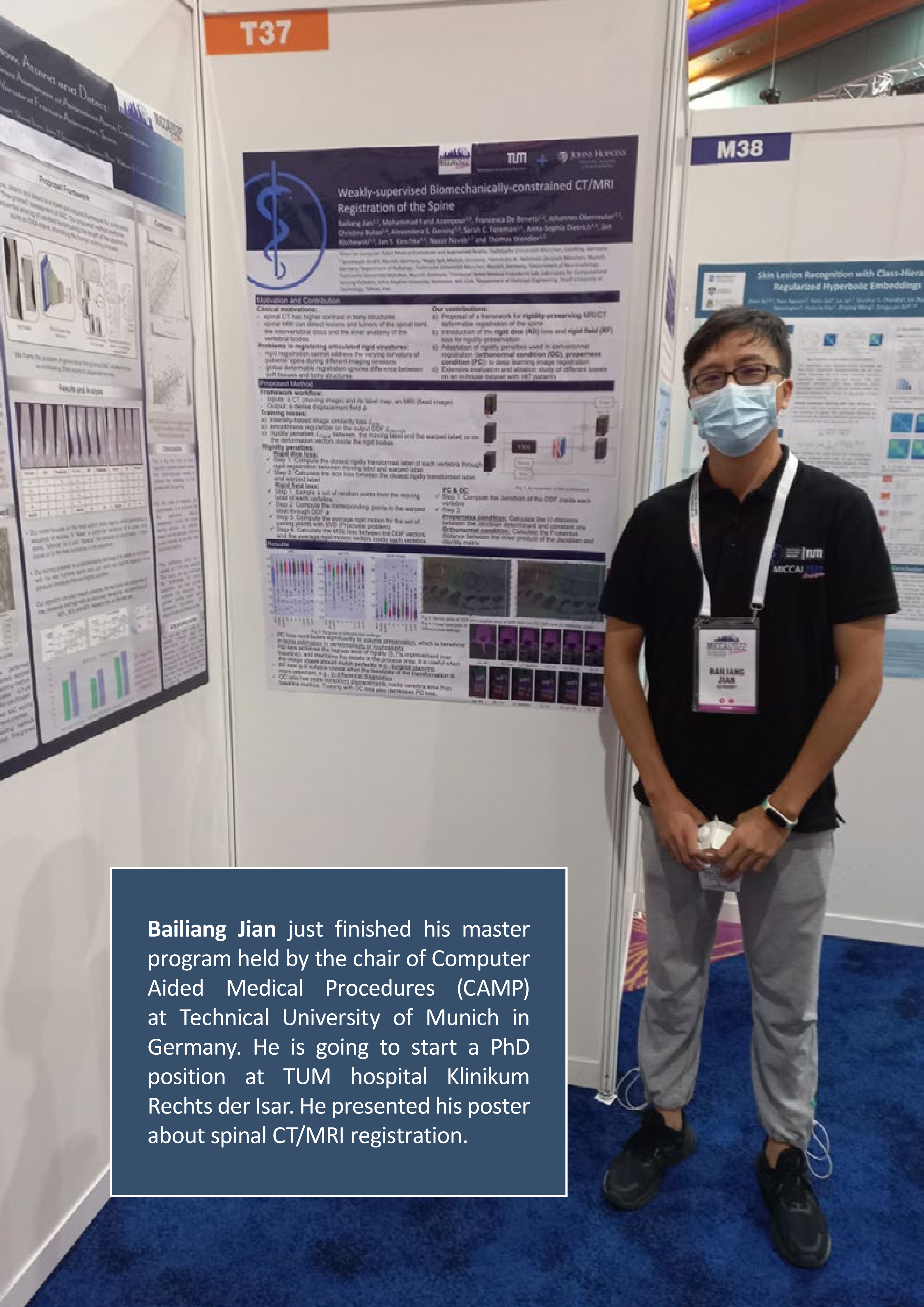
The second challenge (**CMRxMotion**) was organized by yet another participant from Fudan University, Shuo Wang and hosted by Chen Chen and Ouyang Cheng from Imperial College. It addressed respiratory motion artefacts with two main tasks: 1) CMR image quality assessment and 2) robust CMR segmentation.

This session included talks on a wide range of techniques- from recurrent neural networks and “insane” data augmentation to deformable convolutions, multi-task learning and ensemble classification frameworks.

Yasmina Al Khalil started introducing the method OPENGTN, which won 3rd Place on both tasks, and included two sections, an auto-encoder trained to reconstruct images with noise for prediction of quality control, and then an ensemble of models to improve robustness through data augmentation helped by region-based training which segmented apical, middle, and basal slices separately. The 2nd place on Task 1 was won by the Philips CTS method from **Xiuzheng Yue**, combining deep learning for global view and machine learning for LV radiomics feature extraction through voting, while the 1st place was achieved by UON_IMA, from **Ruizhe LI**, where the author used a biased voting strategy to aggregate the decisions from different patch-based models.

The last session covered some echo-based projects such as **Unsupervised Echocardiography Registration through Patch-based MLPs and Transformers**, a patch-based MLP/Transformer method to extract features from echo data, and another one by **Matthias Ivantsits** (image next page) who built an end-to-end oriented mitral valve DL surface reconstruction, applied to 3D TEE data and employing a Voxel-Encoder, Voxel-Decoder and a Mesh-Decoder.

The paper **Efficient MRI Reconstruction with Reinforcement Learning for Automatic Acquisition Stopping** showed a policy network with a state signal value and a penalty term to estimate the reconstruction, and we finished with **Marcel-Beeze**, presenting a project on reconstructed 3D cardiac anatomy meshes for 3D shape-based and contrast-based major adverse cardiac event (MACE) risk prediction; and **Buntheng LY**, who applied LIME and Integrated gradients methods to meshes obtained by Graph Convolutional Network (GCNs).



Bailiang Jian just finished his master program held by the chair of Computer Aided Medical Procedures (CAMP) at Technical University of Munich in Germany. He is going to start a PhD position at TUM hospital Klinikum Rechts der Isar. He presented his poster about spinal CT/MRI registration.

FEDHARMONY: UNLEARNING SCANNER BIAS WITH DISTRIBUTED DATA

BEST OF MICCAI

Nicola Dinsdale

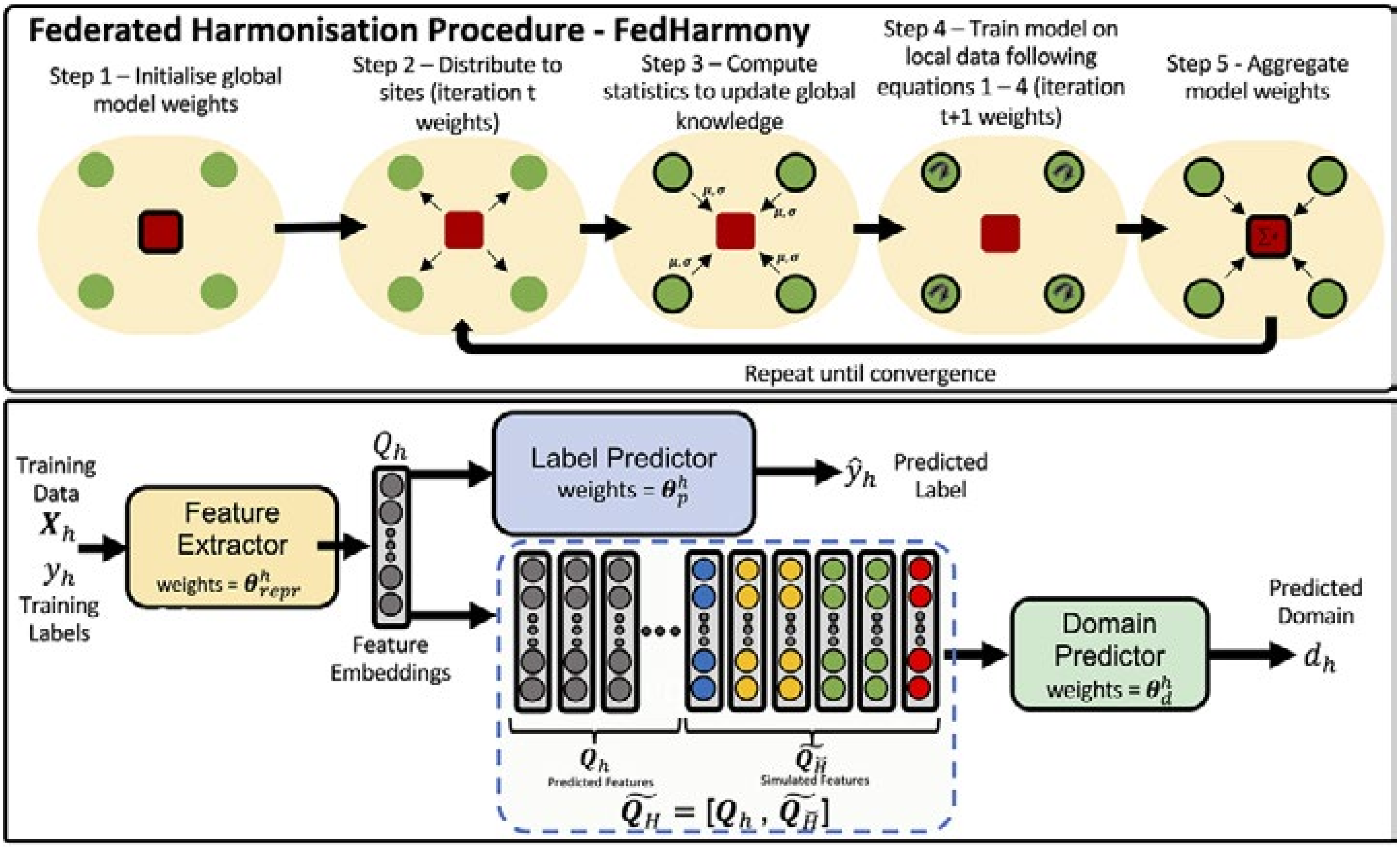
Nicola Dinsdale is a postdoctoral research associate at the University of Oxford under the supervision of Ana Namburete. Her paper proposes a novel solution to the MRI data harmonization problem in a distributed setting. She speaks to us ahead of her poster this afternoon.

When investigating biological problems of interest, combining data across multiple sites and scanners is necessary to **increase statistical power and the breadth of biological variability**. However, doing this presents two problems: the harmonization problem and data privacy concerns.

The harmonization problem occurs because different scanners give different signals. The same subject acquired on different scanners will look subtly different due to the MRI scanner itself, rather than anything interesting in the person's biology. **When data is combined across scanners, this will increase noise**.

Data privacy is an issue because medical imaging data is inherently personal information, so sharing this across sites could be a contravention of privacy legislation.

"Our approach is trying to overcome these



two barriers to enable us to answer the question of interest while removing scanner effects and protecting individual privacy," Nicola tells us.

"The idea is that you should be able to take data from different hospital sites and learn from it without ever having to move it. This method should allow us to combine data across different institutions."

The paper proposes a federated framework in which the data doesn't move. However, for harmonization, data must be compared between sites to be able to remove the scanner information. To solve this, it proposes **domain adaptation in a federated setting while reducing the amount of shared information**.

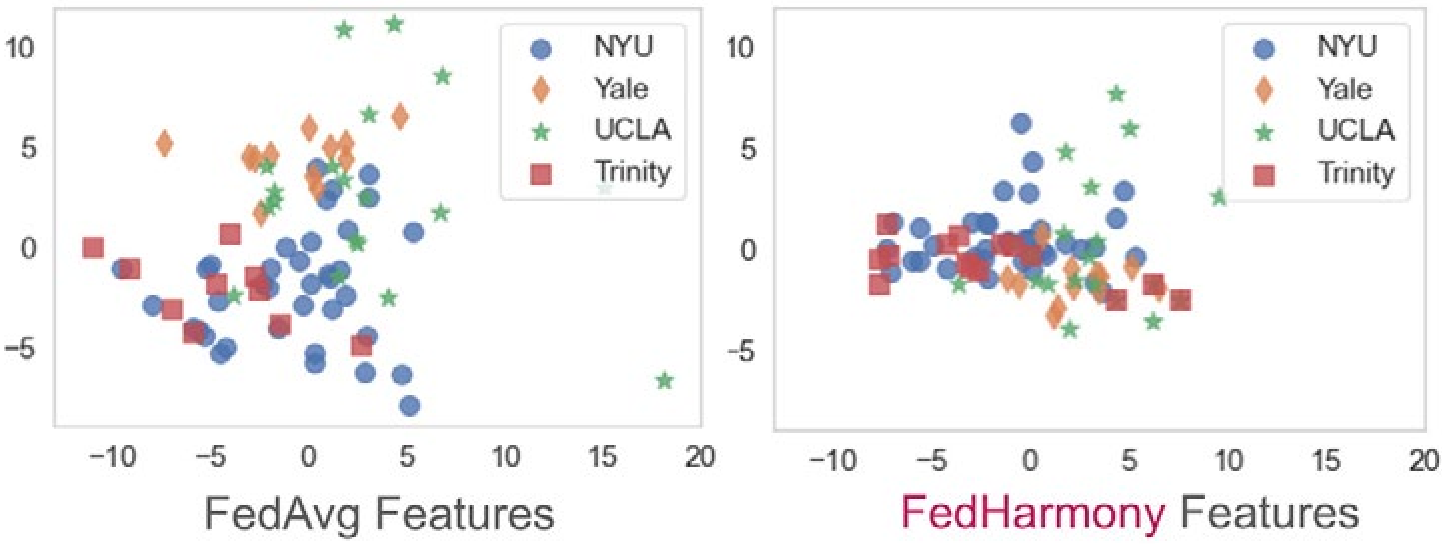
"We reduce the amount of shared information to just a mean and standard deviation per feature per site," she reveals.

"What this means is that rather than sharing the whole amount of data, or an entire feature embedding, you're sharing about 96 pieces of information."

Encoding the information while ensuring privacy was challenging because standard approaches do not protect privacy. However, by encoding it as the mean and standard deviation, this method could share just that information and then create example features by pulling them from a Gaussian distribution.

"We're modeling features as a Gaussian distribution," Nicola explains.

"It's a deep learning-based approach. We use an iterative framework that allows us to remove the information adversarially. We do the task we're interested in while removing the scanner information, but we use the Gaussian distribution to generate



features *for the sites that we're not currently training at to protect the privacy of the individuals."*

Could this novel element be what led to the paper being accepted at MICCAI this year?

"Yeah, I think it goes substantially beyond the existing approaches because they share the whole feature embedding," she responds.

"From that feature embedding, you would be able to reconstruct the image, especially if it was a segmentation task, so it doesn't protect privacy."

The paper demonstrates this new framework for the task of age prediction. Going forward, Nicola hopes to generalize it to different architectures and tasks, such as segmentation.

Perhaps that will earn her an oral presentation in Vancouver next year?

"Hopefully!" she laughs.

Nicola is currently based in a new group at Oxford called the **Oxford Machine Learning in NeuroImaging Lab (OMNI)**, working under **Ana Namburete**. Its work focuses on approaching real-world problems to make deep learning methods help clinically.

"Professor Namburete is a great supervisor because she helps us to investigate the problems that we're interested in but allows us to have our own research directions and creativity," she tells us.

Nicola is not the only OMNI member showcasing their work at MICCAI this year. **Linde Hesse's** poster on Monday, **INSightR-Net: Interpretable Neural Network for Regression using Similarity-based Comparisons to Prototypical Examples**, explored how to make networks interpretable for regression tasks. Also, **Pak Hei Yeung**'s poster on Monday, **Adaptive 3D Localization of 2D Freehand Ultrasound Brain Images**, explored domain adaptation using cycle consistency for ultrasound.

"I really like programming and computer vision, but I'm motivated by exploring real-world problems," Nicola adds.

"I like being able to identify barriers to current methods being applied clinically, and this was the first obvious barrier to approach."

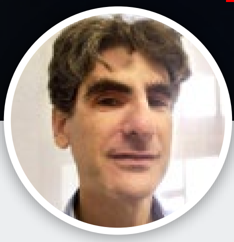
FIRESIDE CHAT

THE CHALLENGES AND FUTURE OF SURGICAL ROBOTICS

Missed the Meetup?

Don’t miss the video!

Be sure not to miss next time :-)



MODERATOR

Moshe Safran

CEO

RSIP Vision USA

GUEST

Eric Taylor

Senior Director, R&D

Surgical Robotics

Medtronic

RSIP VISION Visual Intelligence for MedTech

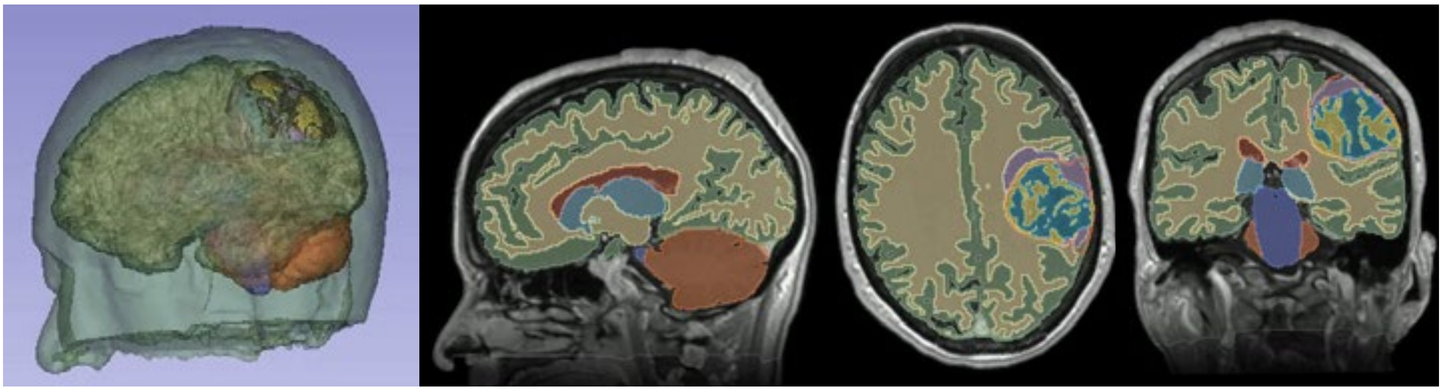
Medtronic

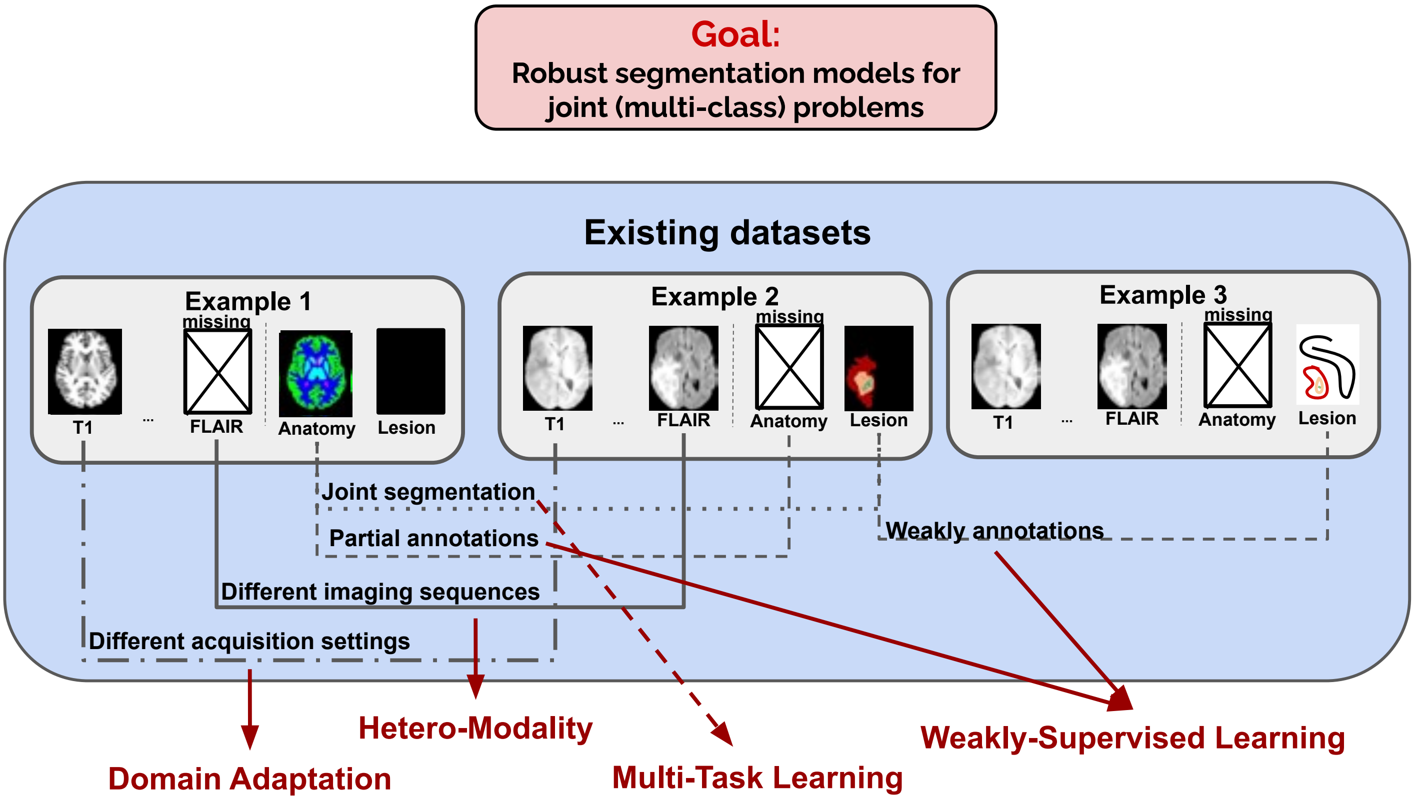
Congrats, Doctor!

Reuben Dorent has recently completed his PhD at King's College London. His research aims to improve brain tumour surgery using medical image analysis. In particular, Reuben's research tackles a fundamental challenge for translating algorithms to clinical practice: the lack of large annotated medical datasets. His research lies at the intersection of weakly-supervised learning, domain adaptation and Bayesian modelling. Reuben is now a postdoctoral fellow at Harvard Medical School in image registration for computer-assisted surgery. Congrats, Doctor Reuben!

Brain tumours strike people in the prime of life. Surgical resection – the primary treatment of most brain tumours – aims at maximising the extent of tumour resection while preserving the patient's cognitive function. To optimise this tradeoff, Reuben's PhD research aimed at developing **automated brain structures and pathology segmentation** for improved surgical guidance.

While Neural Networks have become the state-of-the-art for most image segmentation tasks, annotated databases required to train them are usually dedicated to a single task, leading to **partial annotations** (e.g. brain structure or pathology delineation but not both). Moreover, the information required for these tasks may come from distinct magnetic resonance (MR) sequences, leading to datasets with heterogenous sets of image modalities (**hetero-modality**). Similarly, the scans may have been acquired at different centres, with different MR parameters, leading to differences in resolution and visual appearance among databases (**domain shift**). Given the large amount of resources, time and expertise required to carefully annotate medical images, **it is unlikely that large and fully-annotated databases will become readily available for every joint problem**. For this reason, there is a need to develop **collaborative** approaches that exploit existing **heterogeneous** and **task-specific** datasets and **weak annotations** instead of time-consuming pixel-wise annotations.





Learning from partially annotated datasets.

Since there is no large annotated dataset for joint brain structure and tumour segmentation, we proposed to exploit annotated databases that are partially annotated and hetero-modal. Starting from a variational formulation of the joint problem, we leveraged the disjoint nature of the label sets to propose a practical decomposition of the joint loss. We then minimised the expected risk under the constraint of missing modalities via a tractable upper bound. The proposed approach achieved higher accuracy than well-established atlas-based approaches while not requiring manual tumour delineation.

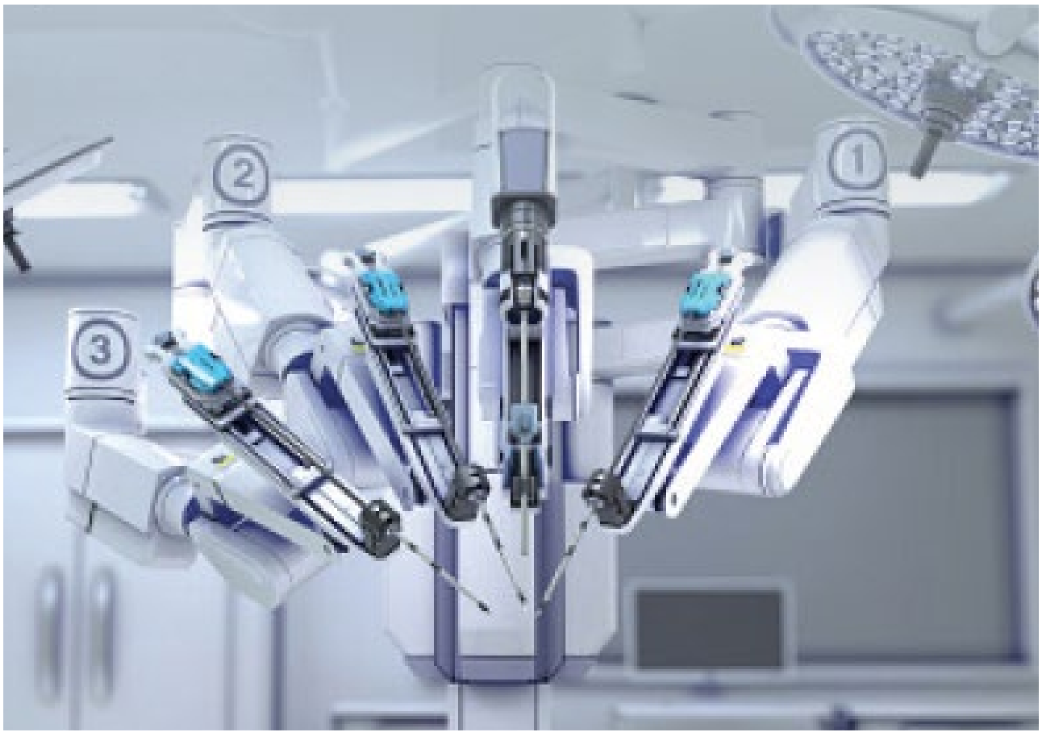
Handling missing imaging modalities.

We proposed a principled formulation using probabilistic graphical modelling to handle missing imaging modalities at inference time. Specifically, all imaging modalities (including segmentation) are assumed to be conditionally independent via a multi-scale latent representation. As a result, the proposed framework successfully performs image segmentation and image reconstruction with incomplete sets of input images.

Improving robustness using weak or missing annotations.

We explored weakly-supervised and unsupervised approaches to ensure that a network trained on a data distribution can successfully generalise on another one. This led to the creation of the first medical benchmark for cross-modality domain adaptation (crossMoDA). The level of performance reached by the top-performing teams from all over the world is strikingly high and close to full supervision.

Next steps: Exploiting the developed algorithms for pre- to intra-operative image registration!



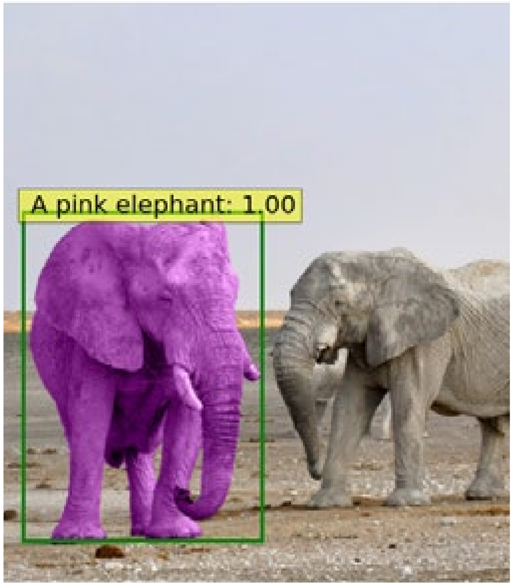
IMPROVE YOUR VISION WITH

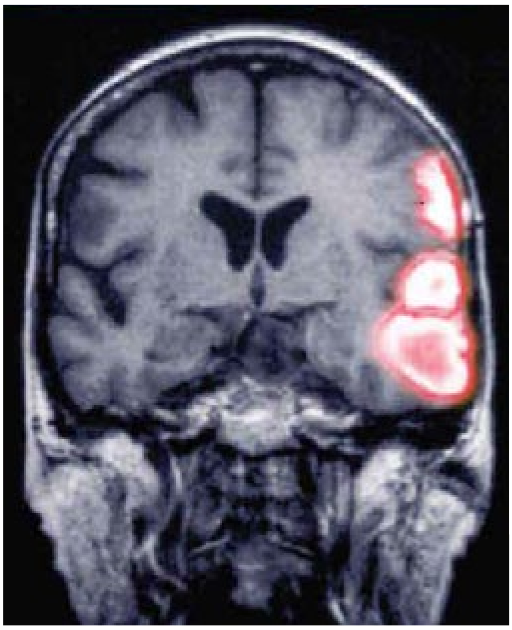
Computer Vision News

SUBSCRIBE

to the magazine of the algorithm community

and get also the new supplement Medical Imaging News!





RSIP VISION