December 2020

Computer Vision News

The Magazine of the Algorithm Community

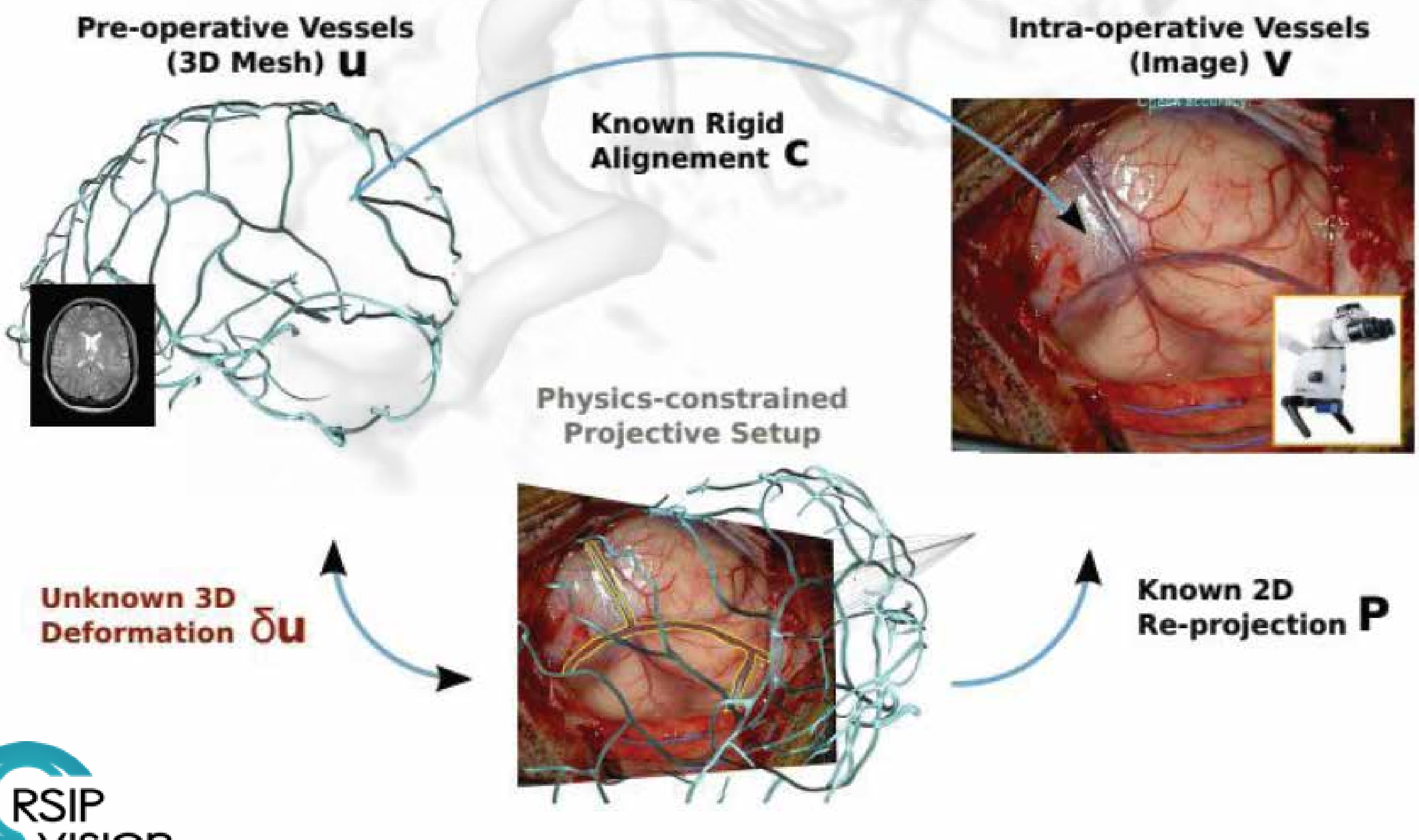
Research
RL-CycleGAN

Women in Computer Vision

Computer Vision Tool
Self-Organizing Maps (SOMs)

RSIP Vision Events

Upcoming Events



Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editors:
Marica Muffoletto
Ioannis Valasakis

Head Designer:
Rotem Sahar

Publisher:
RSIP Vision
[Contact us](#)
[Free subscription](#)
[Read previous magazines](#)

Dear reader,

You will read on the pages that follow our review of a great research paper by **Google scientists: RL-CycleGAN**. Further on, this December issue of **Computer Vision News** includes the 2nd part of our **BEST OF MICCAI 2020** selection, reviewing top papers from the conference. **[If you missed part 1, it is here!](#)** We conclude with top articles about **AI in Ultrasound**: major researchers from academy and industry tell us about current state-of-the-art. Finally, don't miss on page 51 the invitations to **RSIP Vision**'s upcoming Meetup and Webinar. **Register for free now!**

Call RSIP Vision for your next Deep Learning project. LinkBio just did - read what they think about it, below :)

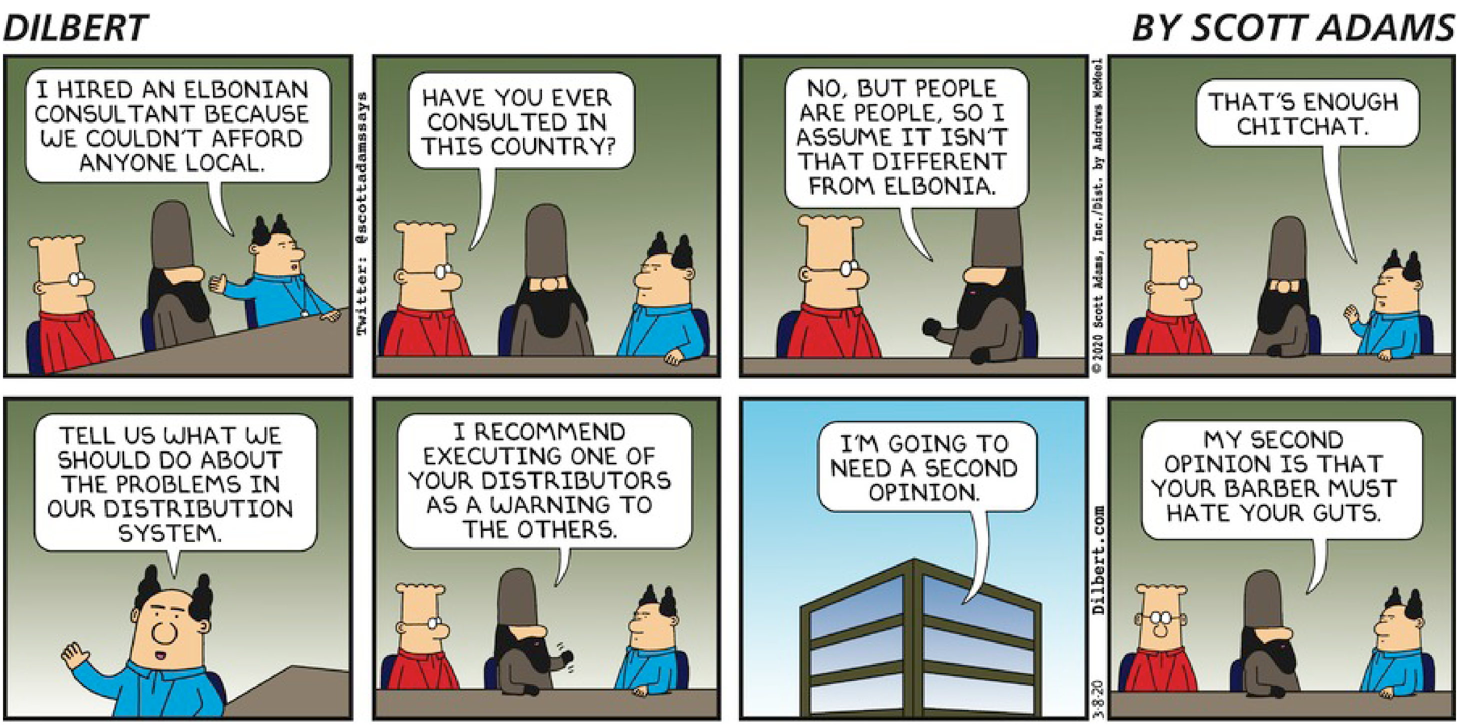
Enjoy the reading and **[subscribe for free](#)**!

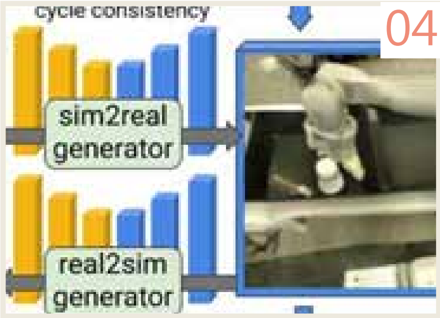
Ralph Anzarouth
Editor, **Computer Vision News**
Marketing Manager, **RSIP Vision**

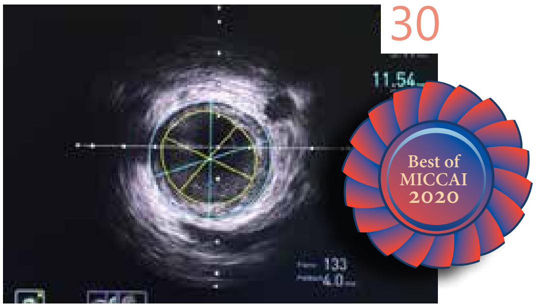
Feedback of the Month

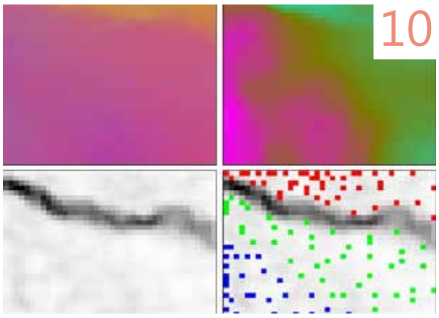
Looking for innovative technologies to enhance our Digital Surgery Platform, we found RSIP Vision's algorithmic solutions to be robust, clinically accurate, and quick. Together with an expert and responsive team, they offer a unique solution!

Riccardo Signoretti,
Vice President, Digital Surgery Platform
LinkBio Inc - Part of Waldemar Link GmbH & Co. KG

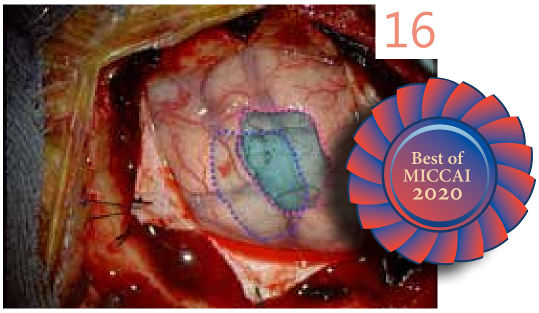



04
cycle consistency
sim2real generator
real2sim generator

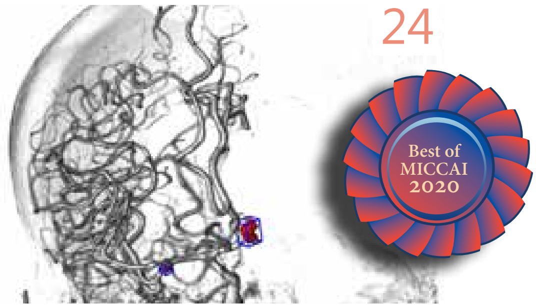

30
Best of MICCAI 2020


10


44


16
Best of MICCAI 2020


50
AI-Assisted Surgery
Perspectives and Challenges
Tuesday Dec 15th
10 am PT
1 pm ET


24
Best of MICCAI 2020


51
Ai4 2021
CYBERSECURITY SUMMIT
EXPLORING ARTIFICIAL INTELLI
CYBERSECURITY
FEB 3-4, 2021
TAKING PLACE DIGITALLY

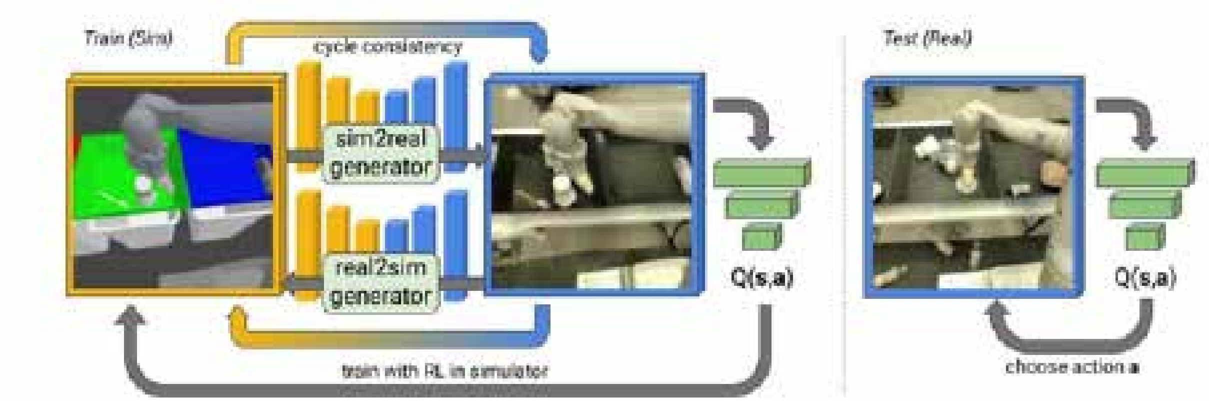
RL - CycleGAN

by Marica Muffoletto

Every month, Computer Vision News selects a research paper to review. For the end of the year, we have had plenty of choice given the recent works presented at MICCAI and CVPR, and we decided to get inspired from the latter and review the paper RL-CycleGAN: Reinforcement Learning Aware Simulation-To-Real. We are indebted to the authors (Kanishka Rao, Chris Harris, Alex Irpan, Sergey Levine, Julian Ibarz, Mohi Khansari), for allowing us to use their images to illustrate this review. You can find their paper at this link.

Introduction:

This paper focuses on the problem of training deep neural networks to perform certain complex tasks, such as grasping objects. To this end, **Reinforcement Learning** (RL) is often used to learn visual representations, but this can require lots of task specific data, hard to retrieve. A possible solution is to train such systems using simulations and then transfer the learned representations on real problems. Of course, there is an issue linked to the **simulation-to-reality gap** which needs to be taken in consideration, and it's usually accounted by manual task-specific engineering. To make this task automatic, the authors propose a method that employs generative models to translate simulated images into realistic ones combined with a RL-scene consistency loss for image translation, to enforce that the Q-values predicted by an RL-trained Q-function are invariant under the domain adaptation (DA) transformation. This is accurately named **RL-CycleGAN: Reinforcement Learning Aware Simulation-To-Real**, an extension of the CycleGAN for unpaired image-to-image translation introduced here.



The authors make use of two important concepts from the **Deep Learning** world which we are probably already familiar with: 1) the Domain Adaptation technique, and 2) Reinforcement Learning. Domain adaptation methods aim at training models using many examples from a source domain (here simulation) and few examples from a target domain (here reality). They can be based on a pixel-level adaptation, as does the **CycleGAN**, or on a feature-level adaptation. Let's look at both of these techniques more in depth!

More on CycleGAN

As per the original paper, CycleGAN includes 2 GANs, here called **Sim2Real** and **Real2Sim**. These learn the mappings from the simulation to real domains and vice versa. The architecture also includes 2 adversarial discriminators. Below we report the mathematical formulations, assuming that X and Y are the two image domains (simulation and real).

Generator Sim2Real	G: X -> Y
Generation Real2Sim	F: Y -> X
Adversarial discriminator 1	Dx: {x} from {F(y)}
Adversarial discriminator 2	Dy: {y} from {G(x)}
Adversarial Loss for Sim2Real	$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y\sim Y}[\log D_Y(y)] + \mathbb{E}_{x\sim X}[\log\left(1 - D_Y\left(G(x)\right)\right)]$
Adversarial Loss for Real2Sim	$\mathcal{L}_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x\sim X}[\log D_X(x)] + \mathbb{E}_{y\sim Y}[\log\left(1 - D_x\left(F(y)\right)\right)]$
Cycle Consistency Loss	$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x\sim D_{sim}} d\left(F\left(G(x)\right), x\right) + \mathbb{E}_{y\sim D_{real}} d\left(G\left(F(y)\right), y\right)$

Sim2Real is trained by: $min_G \; max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y)$

Real2Sim is trained by: $min_F \; max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X)$

The last item in the table is the cycle consistency loss, used to ensure that, from the generated images, the original ones can always be recovered, i.e.: $x \to G(x) \to F\left(G(x)\right) \approx x\,, y \to F(y) \to G\left(F(y)\right) \approx y$

6 Research

Reinforcement learning techniques can be used for a variety of tasks. Training deep neural network models to grasp objects is among these, and it requires both the ability to learn visual representations and to change the environment accordingly.

More on Q-learning

This is a reinforcement learning technique which learns a Q-function, using the below equations, to define the best policy to apply in order to maximize the total expected future reward. The definitions are reported below.

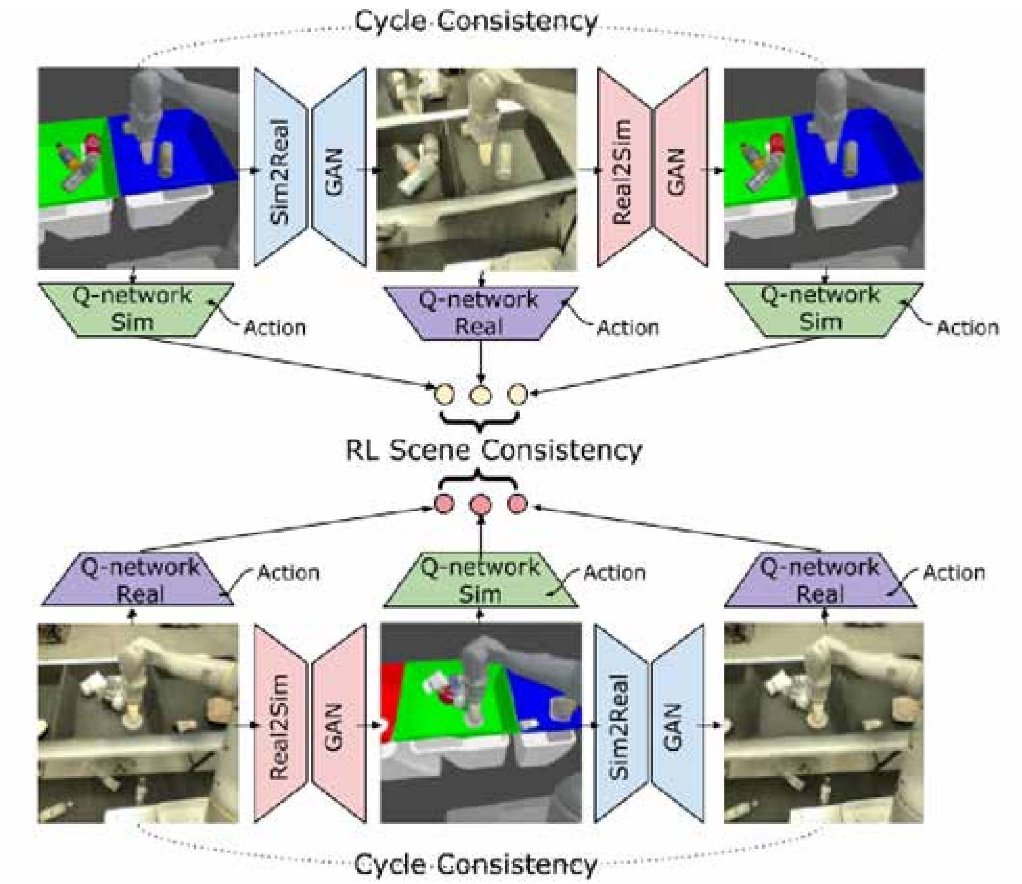
Environment of states	s = input image
Actions	a = candidate action
Rewards	r
Next states	s’
Estimate of next state’s value	V(s’)
Discount factor	γ
Distance metric	d
Q-function (updated to minimise TD loss)	$Q(s,a)$
Temporal difference (TD) loss	$d(Q(s,a), r+\gamma V(s'))$
Policy	$\pi(a\|s) = argmax_a Q(s,a)$

The combination of a CycleGAN with a reinforcement learning technique is employed in order to inform the GAN about which components of the image are relevant and avoid hiding information in the adapted image that might be fundamental. This is done by **enforcing RL consistency losses** on all the inputs and generated images.

RL-CycleGAN

The RL-CycleGAN jointly trains the RL model with the CycleGAN.
The Q-learning estimates Q-values from each image (x, G(x), F(G(x)), y, F(y), G(F(y))) in the CycleGAN model. These are indicated by q_x, q'_x, q''_x (triple for the 1st originally simulated scene), and q_y, q'_y, q''_y (triple for the 2nd originally real scene). Similar Q-values are encouraged within a triple by the RL-scene consistency loss:

$$\mathcal{L}_{RL-scene}(G,F) = d(q_x, q'_x) + d(q_x, q''_x) + d(q'_x, q''_x) + d\left(q_y, q'_y\right) + d\left(q_y, q''_y\right) + d\left(q'_y, q''_y\right)$$



Two different Q-networks (Q_{sim}, Q_{real}) are trained for simulation-like (x, F(G(x)), F(y)) and real-like images (G(x), y, G(F(y))), using the standard TD-loss. To compute the Q-function loss, the generator or pair of generators are applied to both current image x and next image x' first, and then combined with the TD-loss.

$$\mathcal{L}_{RL}(Q) = \mathbb{E}_{(x,a,r,x')}\, d(Q(x,a), r + \gamma V(x')$$

The total loss of the RL-CycleGAN is ultimately defined as:

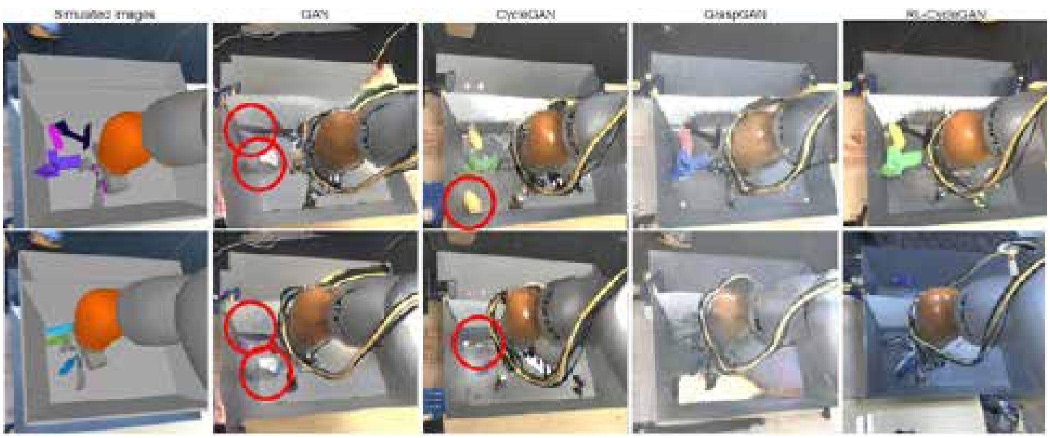
$$\begin{aligned}\mathcal{L}_{RL-CycleGAN}(G,F,D_X,D_Y,Q) \\ &= \lambda_{GAN}\mathcal{L}_{GAN}(G,D_Y) + \lambda_{GAN}\mathcal{L}_{GAN}(F,D_X) \\ &+ \lambda_{cycle}\mathcal{L}_{cyc}(G,F) + \lambda_{RL-scene}\mathcal{L}_{RL-scene}(G,F) \\ &+ \lambda_{RL}\mathcal{L}_{RL}(Q)\end{aligned}$$

The **RL-CycleGAN** is evaluated through three experiments on two different setups for robot grasping. To demonstrate the method invariability to robot and task, the two setups have different purposes. The former aims to generalize grasping of unseen objects, and the second increases the complexity of the environment, since grasping is performed from three bins with the robot placed at different locations relative to the bin. The performance is measured in terms of grasp success rate on the two robotic grasping systems.

Experiment 1:

This experiment is built to analyze if the main aim of bridging the gap when using simulated experience is fulfilled. It compares several methods built on GAN approaches and measures the success rate of each. The grasping model trained with the RL-CycleGAN performs the best (**70% success**). *"It preserves task-salient information and produces realistic images and does so with a general-purpose consistency loss that is based directly on the similarity of Q-values, without requiring manual identification of task-salient properties (e.g., object geometry)."*

The difference with a regular GAN (with a success rate of 29%) is huge, and this can be fully appreciated visually in the figure below.



Experiment 2:

This experiment investigates the effect of mixing real data and simulated data to train the grasping model, showing that the performance is hugely improved by including simulations using simulation-to-real methods. And specifically, for the first setup (grasping of unseen objects), even with a large available dataset of 580,000 real world trials, RL-CycleGAN success rate rises from 87% to 94%. Similar results are obtained on the second setup.

Experiment 3:

In the end, the authors experiment with fine-tuning the grasping models, using on-policy real data. The amount of off-policy real data to train RL-CycleGAN is then reduced to 5,000 for this, and it is found that, in comparison with state-of-the-art methods based only on on-policy data, it reaches the same performance using less episodes and no domain randomization.

Conclusion:

Tested on two different robot grasping setups, RL-CycleGAN achieves incredible performance. This work shows significant improvement on real world vision-based robotics. You can just have a look at the examples of robot grasping performed by RL-CycleGAN and make up your own mind on this interesting work.

We are closing this year's reviews on research with this exciting insight on robotics and deep learning. We look forward to yet many more stimulating papers in 2021!

Self-Organizing Maps (SOMs) with PyTorch

by Ioannis Valasakis (@wizofe)

Welcome back! :)

A term which isn't usually heard in the field of Neural Networks is that of a self-organizing map (SOM). Although it's difficult to describe it briefly, a SOM is a type of Artificial Neural Network (ANN) which is used as way to reduce dimensionality. This is achieved by creating a discretized representation of the input which is called a map. The main difference to the often used ANNs is that instead of back-propagation they use graphs to indicate the properties of the input space.

To explain it, we'll take a simple dataset, the famous Iris from R.A. Fisher. You can easily find and download this dataset, but it is also available through the Scikit-Learn Python package. In that case it can be loaded as follows:

```python
from sklearn import datasets

# load R.A. Fisher's Iris dataset
iris = datasets.load_iris()
```

The Python program that follows creates a 40x40 SOM of the dataset as a data structure. In the Figure 1, a U-Matrix is created with the black cells showing the similarity of the items and the white cells showing the borders between the item clusters. It can be shown that 3 different clusters (classes) are defined.

The dataset has four dimensions and there are three different labels (numbered as 0, 1 and 2). In order to visualize the dimensions, the SOM is used to reduce the dimensions to two.

Creating SOMs using Python

The Iris dataset includes 150 items. Some representative lines are shown below:

```
5.1,3.5,1.4,0.2,0
4.9,3.0,1.4,0.2,0
7.0,3.2,4.7,1.4,1
6.3,3.3,6.0,2.5,2
```

Each line represents an iris flower. The flower names are 0: setosa, 1: versicolor, 2: virginica and are represented in the last item of each line. The first four values on each line are the flower's sepal length, sepal width, petal length and petal width. Therefore, the data has four dimensions. SOMs usually need normalized data although in this case the predictor values all have roughly the same magnitude, so you could omit that step.

How many dimensions a SOM must have is a very subjective matter and very often decided after experimentation. The main idea is that each node vector represents some of the data items and on the same time the map nodes are geometrically closer, therefore representing similar items.

The SOM will be constructed using the following pseudo-code:

```
Define a n x n map with random node vector values
loop while s < StepsMax times
  Calculate and define what a "close" node means, based on s
  compute a learn rate, based on s
  pick a random data item
  determine the map node closest to data item (BMU)
  for-each node close to the BMU
    adjust node vector values towards data item
end-loop
```

The main skeleton

Here is the main skeleton of the SOM creation:

```python
import numpy as np
import matplotlib.pyplot as plt

def closest_node(data, t, map, m_rows, m_cols): . .
def euc_dist(v1, v2): . .
def manhattan_dist(r1, c1, r2, c2): . .
def most_common(lst): . .

def main():
  # 0. Definitions
  np.random.seed(1)

  dims = 4
  rows = 40
  cols = 40

  range_max = rows + cols
  lr_max = 0.4
  Steps_max = 2500

  # 1. Data loading
  # 2. SOM construction
  # 3. U-Matrix display and construction
  # 4. Reduced data display and construction

if __name__=="__main__":
  main()
```

The demo program is `isom.py` and after importing the needed packages, it defines four functions which are explained in the next paragraphs.

The function `closest_node()` gives indices for the row and column for a SOM with size `m_rows x m_cols`. As explained earlier, those are the coordinates of the map cell, the vector of which is closest to the data item at `data[t]`. Best matching unit (BMU) is the nearest cell vector to a specified data item.

The function `euc_dist()` is the Euclidean distance between two vectors, which is defined as the square root of their location difference, or by the Cartesian coordinates of the points using the Pythagorean theorem.

The function `most_common()` provides the most common value, using as input a list of integer values. The variables in the section "0" hold the dimensionality of the dataset, and the number of rows and columns of the SOM. `Range_max` is the maximum Manhattan distance for any two cells in the SOM, `lr_max` is the initial learning rate used when constructing the SOM and `steps_max` specifies the number of training iterations to perform.

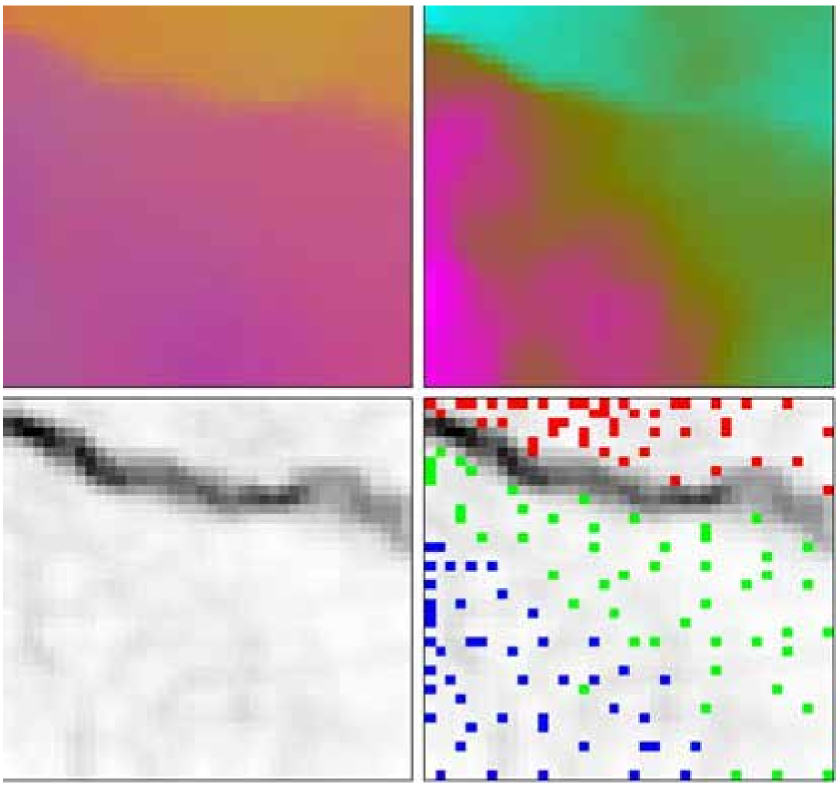


Image 1: A 40x40 grid of neutrons (SOM) using Fishers Iris flower data set (250 iterations).
A color image formed by first three dimensions of the four dimensional SOM weight vectors (top left), pseudo-color image of the magnitude of the SOM weight vectors (top right), U-Matrix (Euclidean distance between weight vectors of neighboring cells) of the SOM (bottom left) and overlay of data points (red: I. setosa, green: I. versicolor and blue: I. verginica) on the U-Matrix based on the minimum Euclidean distance between data vectors and SOM weight vectors (bottom right) *Source: Wikipedia*

Loading the Iris dataset

Using the scikit-learn Python package, the dataset can be loaded in the memory, as shown earlier.

Defining the SOM

The SOM is created by using the following statement:

```
print("Constructing a 40x40 SOM from the iris data")
map = np.random.random_sample(size=(rows, cols, dims))
```

The call to random_sample() generates a 40 x 40 matrix where each cell is a vector of size 4 with random values between 0.0 and 1.0. The creation of the SOM starts with statements:

```
for s in range(steps_max):
  if s % (steps_max/10) == 0: print("step = ", str(s))
  pct_left = 1.0 - ((s * 1.0) / steps_max)
  curr_range = (int)(pct_left * d)
  curr_rate = pct_left * learn_max
```

The `pct_left` calculates how many iteration steps are remaining (%).The `curr_range` is the maximum Manhattan distance at step s which defines how near is it.

Following, a random data item is selected and the best matching unit map node/cell is determined:

```
  t = np.random.randint(len(data_x))
  (bmu_row, bmu_col) = closest_node(data_x, t, map, rows, cols)
```

Each node/cell of the SOM is examined and the vector in the current node is updated if the current node is defined as "close" to the BMU.

```
for i in range(Rows):
   for j in range(Cols):
     if manhattan_dist(bmu_row, bmu_col, i, j) < curr_range:
       map[i][j] = map[i][j] + curr_rate *
      (data_x[t] - map[i][j])
```

Define the U-Matrix

The 40 x 40 vectors hold a value that corresponds to one or more data items which helps define the U-Matrix as following:

```
print("Define U-Matrix using the SOM")
  u_matrix = np.zeros(shape=(rows, cols), dtype=np.float64)
```

Each 40 x 40 cell of the U-Matrix holds an initial 0.0 value which is updated in every iteration:

```
for i in range(rows):
    for j in range(cols):
      v = map[i][j]
      sum_dists = 0.0
      ct = 0
```

The value in the SOM (v) corresponds to the current U-Matrix cell. Every adjacent cell in the SOM (above, below, left, right) is processed and the sum of the Euclidean distances is computed:

```
if i-1 >= 0:    # above
  sum_dists += euc_dist(v, map[i-1][j])
  ct += 1
if i+1 <= Rows-1:    # below
  sum_dists += euc_dist(v, map[i+1][j])
  ct += 1
if j-1 >= 0:    # left
  sum_dists += euc_dist(v, map[i][j-1])
  ct += 1
if j+1 <= Cols-1:    # right
  sum_dists += euc_dist(v, map[i][j+1])
  ct += 1
```

To visualize what is happening for a cell in the SOM with value (2.0, 1.0, 1.5, 0.7) and the Euclidean distances to the four neighbor cells are 7.0, 12.5, 11.5, 5.0, then the corresponding cell in the U-Matrix holds 36.0 before averaging and then 9.0 after averaging:

```
u_matrix[i][j] = sum_dists / ct
```

When the value in U-Matrix is small, it means that: the A small value in a U-Matrix cell indicates that the relative cell in the SOM is very near to its neighbors and therefore the neighboring cells are part of a similar group. The U-Matrix is displayed:

```
plt.imshow(u_matrix, cmap='gray')
plt.show()
```

Visualize the dimensionality reduction

When the data has class labels (as in the previous example), the SOM is able to reduce the dimensionality, so that data can be visualized and represented as a two-dimensional graph. The code to do this is shown below:

```
print("Assign data labels to map nodes")
mapping = np.empty(shape=(rows, cols), dtype=object)
for i in range(rows):
  for j in range(cols):
    mapping[i][j] = []
```

The class label is associated to each cell depending on the proximity of the corresponding cell in the SOM and then it's added to the cell list:

```
for t in range(len(data_x)):
  (m_row, m_col) = closest_node(data_x, t, map, rows, cols)
  mapping[m_row][m_col].append(data_y[t])
```

The most common class label is extracted from the list (in the current cell scope) and placed into the following matrix:

```
  label_map = np.zeros(shape=(rows, cols), dtype=np.int)
    for i in range(rows):
      for j in range(cols):
        label_map[i][j] = most_common(mapping[i][j], 3)
```

The 40 x 40 matrix `label_map` has the value of -1 if no data items are associated with the cell, or a value 0, 1 or 2 which indicates the most common class label associated with the cell. Now the reduced dimensionality matrix can be displayed:

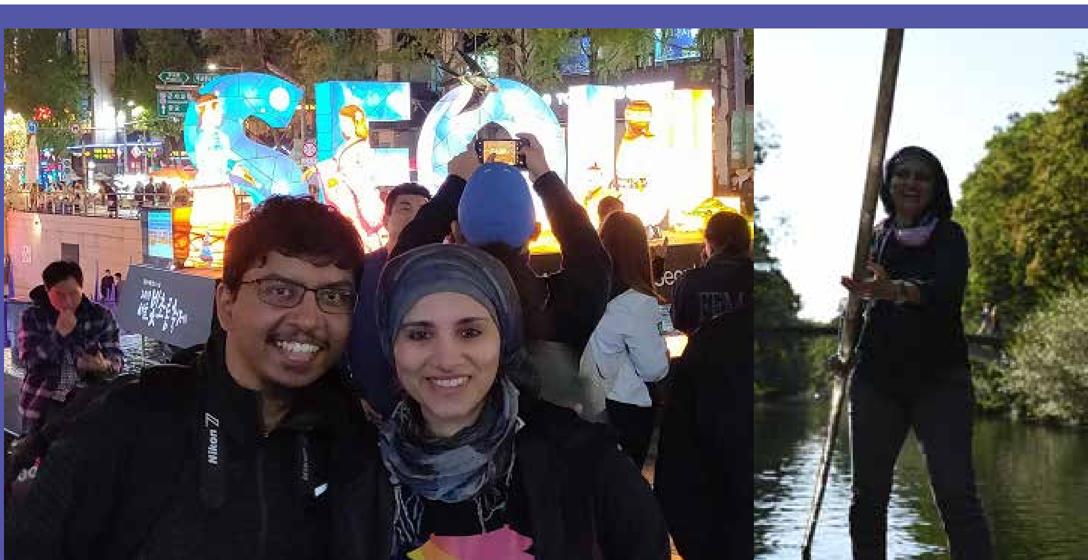
```
    plt.imshow(label_map, cmap=plt.cm.get_cmap('terrain_r', 4))
    plt.colorbar()
    plt.show()

  if __name__=="__main__":
    main()
```

The 4 arguments passed to the function `get_cmap()` take into account the four colors (as it's also displayed previously). Those are one for each class and the extra color to show that there's no association with any.

Finale!

There are many possibilities to explore using SOMs. Exploring hyper-parameters, using implementations with the fast.ai library (there is a nice article written by Ricardo Sayn on that) and more. Most importantly, even if they are not used for a real-life scenario, SOMs are a great way to define strategies, use neighborhood functions and explore dimensionality reduction! See you next month! :)



"Meet a Rockstar on page 44!!!"

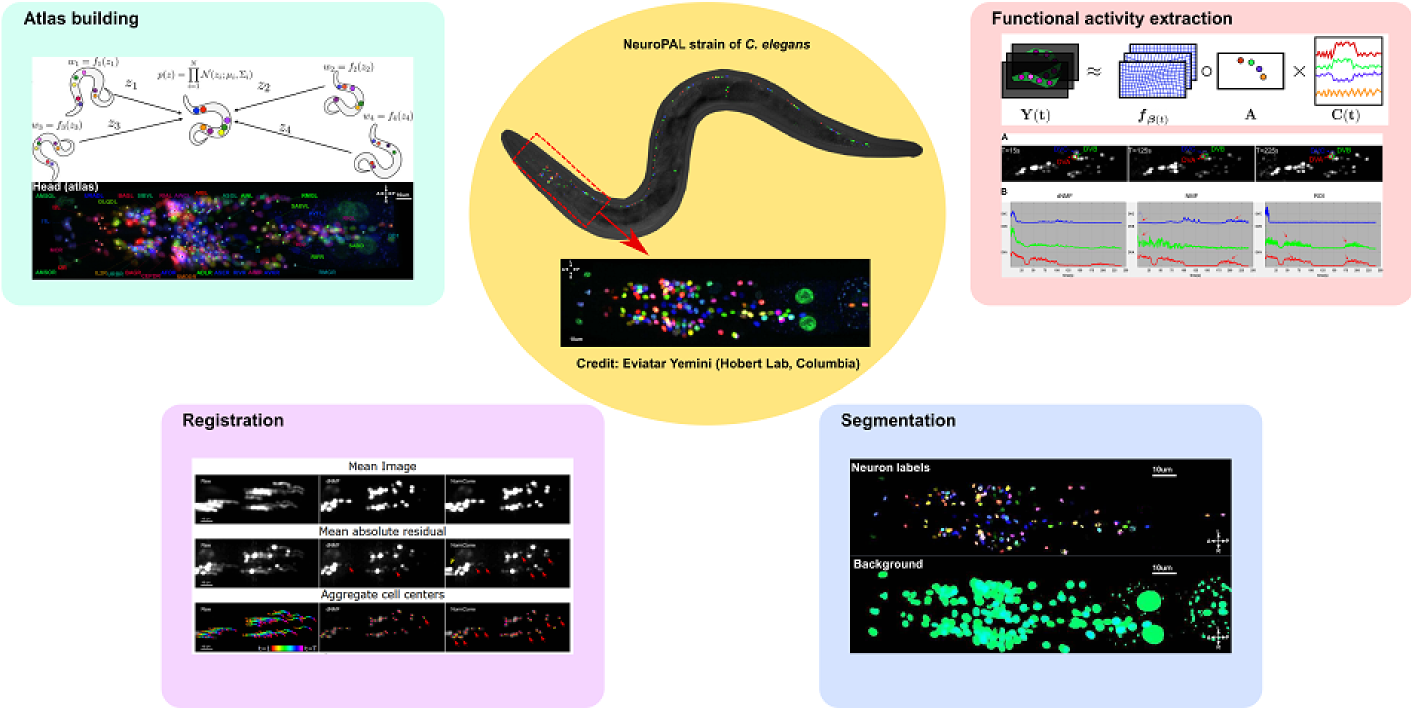
Statistical Atlas of C.elegans Neurons
Probabilistic Segmentation and Labeling of C. elegans Neurons
Demixing Calcium Imaging Data in C. elegans via Deformable Non-negative Matrix Factorization

Erdem Varol is a postdoc scholar and Amin Nejatbakhsh is a fourth-year graduate student at Columbia University in New York City. They are presenting a series of three papers at MICCAI this year exploring brain activity in the well-studied nematode C. elegans. They spoke to us ahead of their oral presentation at MICCAI.

Erdem Varol

Amin Nejatbakhsh

C. elegans is a worm that grows to about 1mm in length and lives in the soil in many parts of the world. It is a simple organism with a nervous system and a brain that can be perturbed to explore how it behaves under different conditions. It has a very low number of neurons and **parts of its brain are visible at a single neuron level**, which is not possible in humans. To study this organism's brain at such a fine resolution, you need many **computer vision tools**, which have been lacking until now. Erdem and Amin have developed a suite of tools to help, including atlas **building, registration, segmentation, and signal extraction**.



As each of these tasks are so involved, they have decided to dedicate **a paper to each of them!**

"Each of these papers are like an atomic part of analysis and data extraction from videos," Amin tells us. ***"A similar pipeline is developed for humans****. For example, fMRI data analysis, where you start with a bunch of images from the brain, build an atlas, and then go ahead and try to extract signals. We thought that some of the methods we were developing were applicable to other model organisms, so we wanted to keep it as generic as possible so that the community could at least use some of the building blocks of our method on other types of problems or animals."*

In the context of **neuroscience**, people usually start with simple questions and simple model organisms. In the atlas building problem, the team are studying the **viability of neuronal positions in the worm body** across different animals. For the signal extraction part, they are trying to understand how neural activity is giving rise to behavior. That is a fundamental problem in systems and **computational neuroscience**.

The team had a '**eureka**' moment when they used a very new area of computational research called **optimal transport**. They applied it to a problem that it had not been applied to before – segmenting neurons – and it yielded **excellent results**.

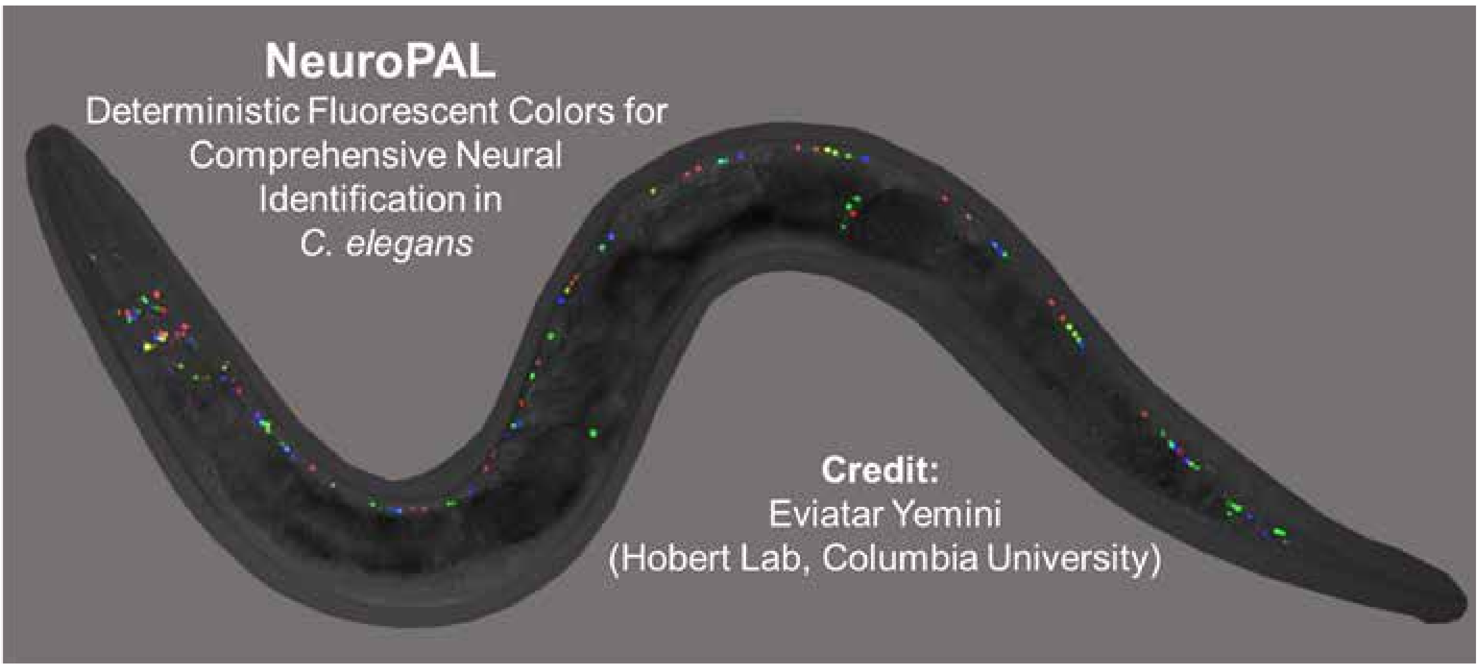
"This problem is more challenging than the usual problems people deal with and there are several reasons for this," Amin explains. *"The images you get do not have a perfect spatial resolution. As*

it is a worm, the videos we work with have non-linear motion, so we need to model that based on ***non-linear modelling tools****. That is one of the challenges we have been dealing with. Another challenge is a* ***lack of training data****. In many other cases, you would have substantial training data that you can use for building neural networks or machine learning tools, but in this case, it is a fairly recent technique. Figuring out* ***how to combine tools and techniques from different topics*** *was pretty challenging."*

In the segmentation paper, the team model the image of the C. elegans as a **Gaussian mixture model**, which has been used in this context for segmentation Ibefore, but part of the novelty of this work is that it combines the literature of Gaussian mixture models with optimal transport, which is a more recent topic possessing optimization techniques that people are using to build **efficient computational models**. It could be used as a general method for doing inference in the Gaussian mixture models. The team see this as their contribution to the field of statistics.

"Our imaging set-ups and computer vision algorithms are fairly new and that is why we had to build a lot of these tools from the ground up," Erdem tells us. *"C.elegans, especially the type of worms that we are using, have only been introduced in the last five years. There are five or so groups now working on similar problems, but the number isslowly but surely increasing. We hope to* ***be a pioneer in this area in terms of developing this first set of techniques*** *to motivate others to expand on them and improve on our results."*

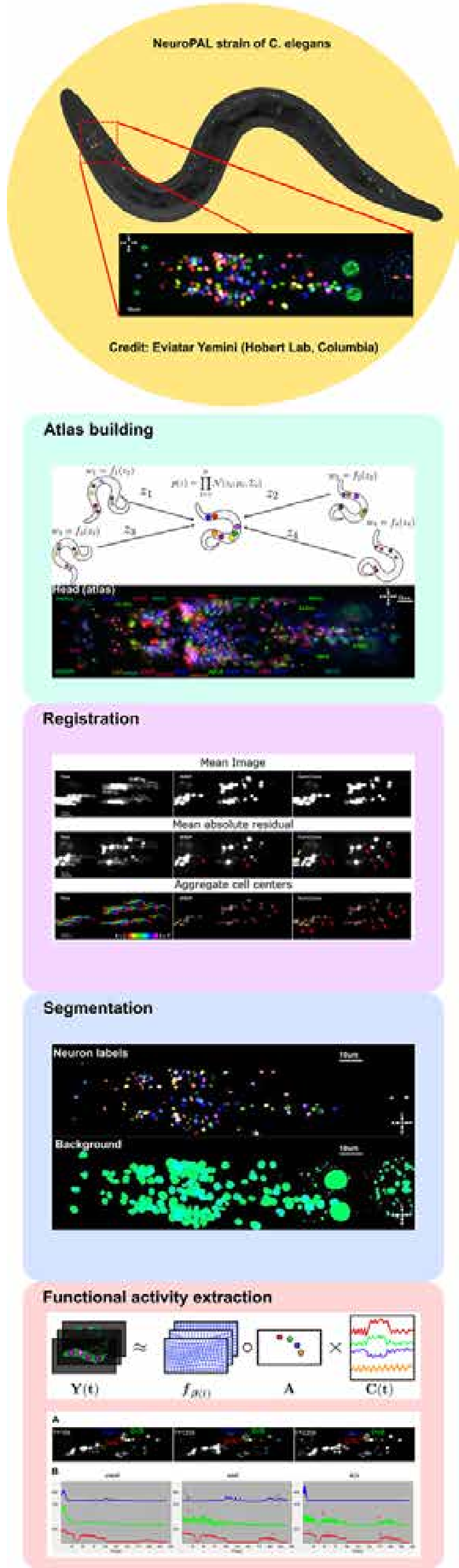
The pipeline the team are building allows scientists to **automate many things they have been doing manually**. Before, they



would need to go through each of the images or timeframes in a video, find the neurons and extract signals from them. Now, it is an **automated process**. They can just input the videos and grab the neural activities from them, which means experiments can be conducted in a high-throughput manner, leading to greater scientific insight.

"What I am most proud of about this work is not only does it have a great deal of potential in terms of accelerating the scientific hypothesis testing cycle, which is of great use to neuroscientists, but it can be of high interest to ***computer vision practitioners and method developers*** *as well,"* Erdem enthuses. *"The computational techniques we put forward could inspire other computational techniques for a completely different set of problems."*

Erdem and Amin have made all the **codes and software available online**. They would like to encourage people who work with C. elegans or other model organisms to explore their papers and see if they can use anything from the pipeline in their own problems.



Deformation Aware Augmented Reality for Craniotomy using 3D/2D Non-rigid Registration of Cortical Vessels

Nazim Haouchine is a post-doctoral fellow at the Brigham and Women's Hospital and the Harvard Medical School in Boston. He spoke to us ahead of his oral presentation to support neurosurgeons with a specific problem encountered when performing craniotomy.

To remove a tumor from the brain, a neurosurgeon needs to first open the skull. This operation is called a craniotomy. Before surgery, patients have an MRI scan, so the surgeon can understand where the tumor is and how to access it. However, in the operating room (OR), a phenomenon called brain shift can occur where the brain deforms, which means everything that was planned before the operation is no longer valid.

Marta Kersten-Oertel, a Canadian scientist, told Computer Vision News a couple of years ago that the brain has parts that you can work on and other parts that you absolutely should not touch.

"Exactly," Nazim agrees. *"I know the work by Marta Kersten-Oertel and ours is* ***closely related****. You must be precise. You have to avoid vessels and risky areas*

Nazim Haouchine

and ***access the tumor without damaging the brain****."*

This work explores how **pre-operative planning can be updated in real-time** to consider intraoperative brain shift, without the need for additional scanning and in the simplest way possible so that it can be applied generally around the world.

"Our core idea was to use ***cortical vessels****,"* Nazim explains. *"These are vessels that are visible* ***on the surface of the brain*** *and are strong features from a* ***computer vision*** *point of view. You also have them on the MRI, so if you extract them, you can work with them to perform a* ***graph registration***

*that will **estimate the brain shift** and that will **align and register the brain**."*

Another important aspect of this technique is that it only uses one image, which turns it into an ill-posed problem, but makes it appealing for future use in the operating room. Nazim solves the problem with a **graph-based registration** using physics-based modeling and a **graph in 2D extracted from cortical vessels**. Those cortical vessels are extracted in the OR using a **convolutional neural network. Deep learning** is very successful nowadays in generating robust segmentation. He built the first segmentation technique on cortical vessels so that it can be used and generalized later.

Nazim previously worked on a similar technique related to the liver at **Inria Strasbourg**, where he brought together computer vision, computer graphics, physics, and space simulation. *"When I came here to the **Brigham** they had this problem, and it is a well-known problem, but we have taken a fresh view on it and I wanted to bring to it some of what I had learned before,"* he tells us. *"We tried to do something different from others using simple methods and existing data without adding hardware. It makes a really nice pipeline. The **collaboration between neurosurgeons and scientists** makes me very proud actually. We are on the edge of trying this in the OR in the real world."*

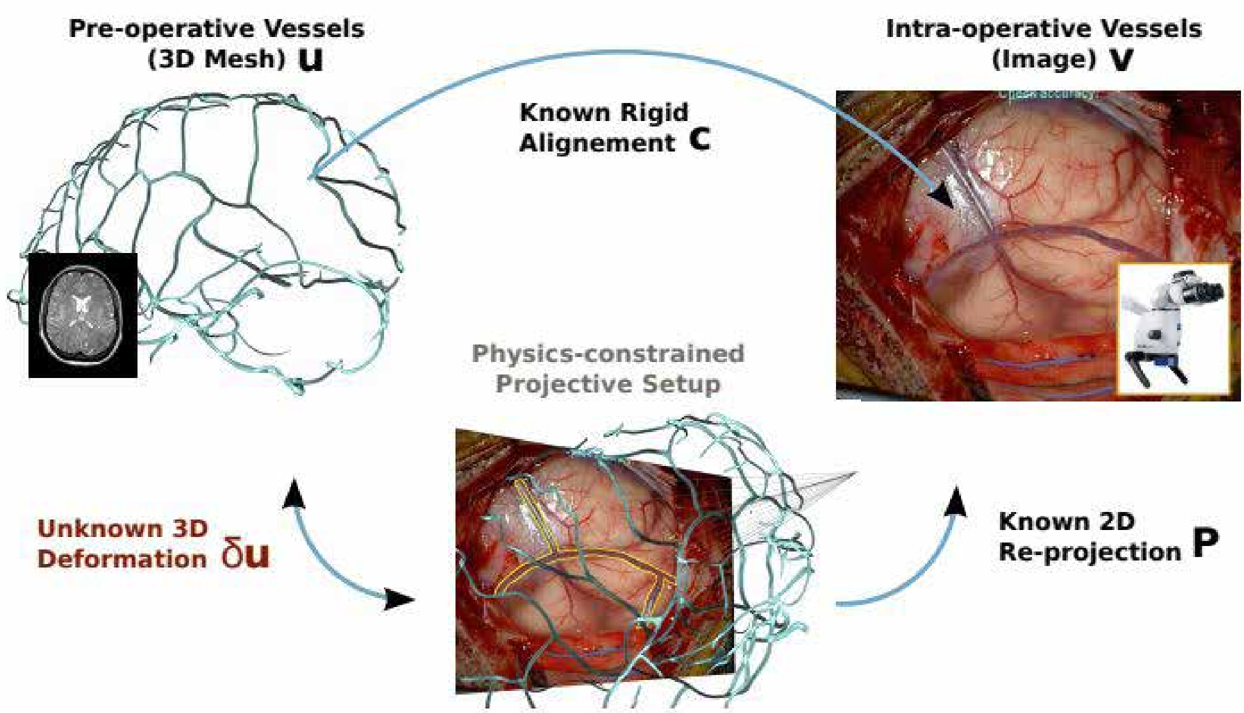
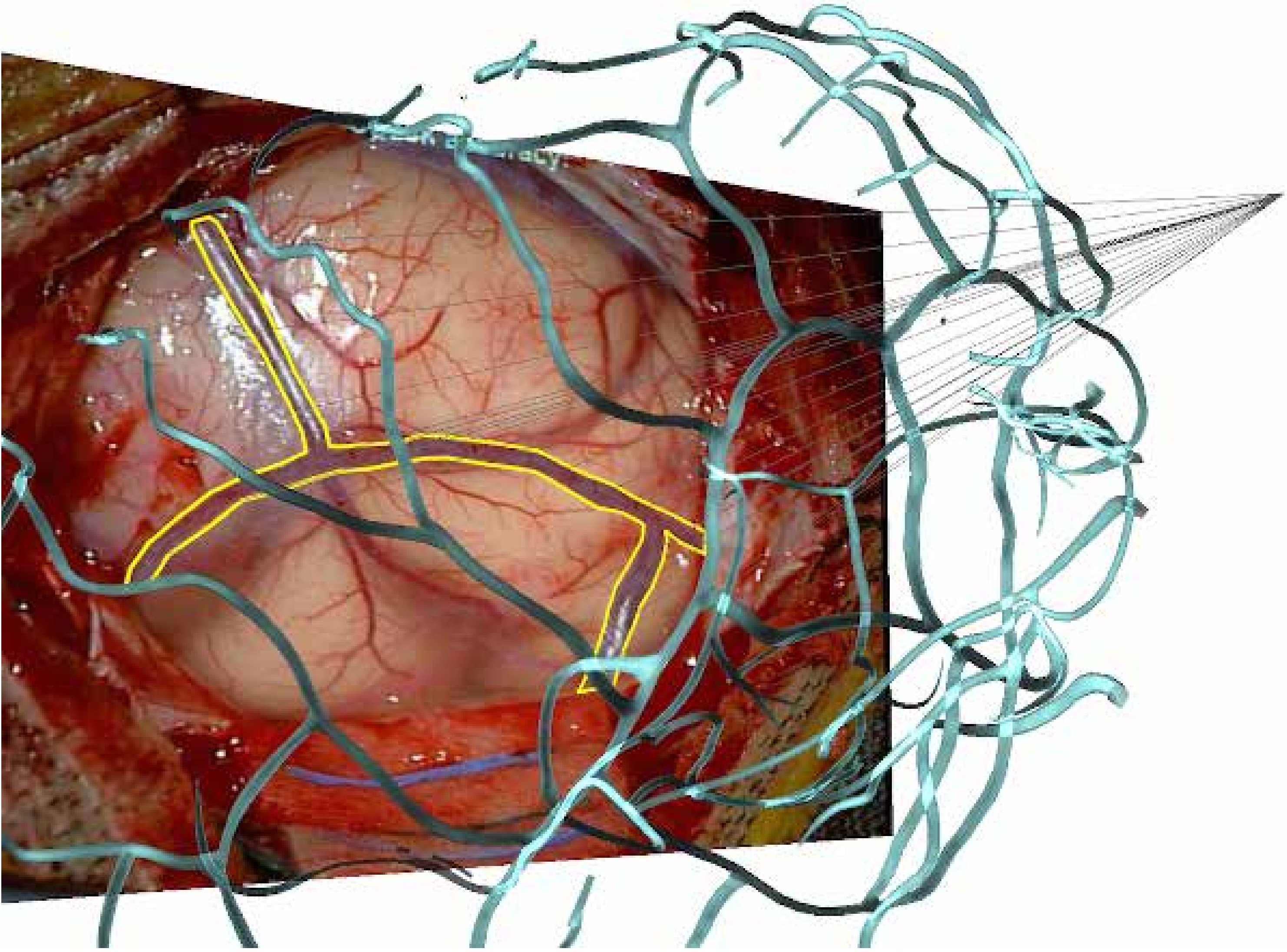


Fig. 1: Problem formulation: we aim at recovering the deformed 3D vessels shape $\delta\mathbf{u}$ from its known reprojection in the image $\mathbf{v}$, the known pre-operative 3D vessels at rest $\mathbf{u}$ and known rigid alignment $\mathbf{c}$, satisfying physical and reprojective constraints $\mathbf{P}$.

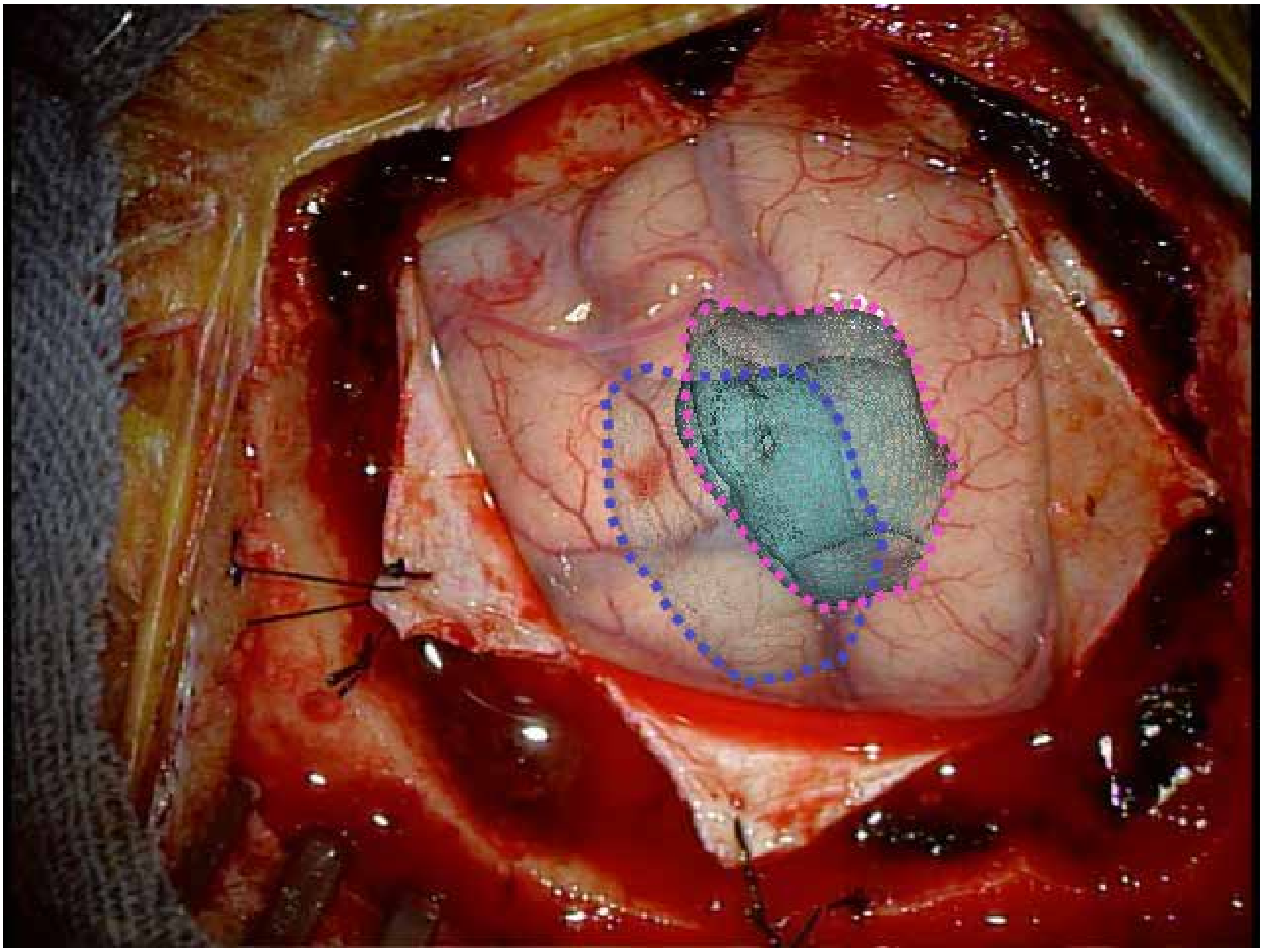


Thinking about next steps, Nazim says their **two-step registration process remains to be solve**d, as neurosurgeons still have to help in the initialization. In the future, he would like to see a one-step end-to-end pipeline that has the registration ready with the brain shift already compensated. They are currently working on this. It will involve more data and a more powerful deep learning model, but it will mean it is **a wholly automated process.**

Originally from **Algeria** in **North Africa**, Nazim spent the last decade in Europe, working in France. We wonder, as one of only two Algerians we have ever interviewed, is he treading a rare path?

"It is more common than you think actually," he responds. *"There are many smart and bright people there doing science, and some do travel to Europe or the US to study and work. It is not a lot, but maybe because I am Algerian too, I know them!"*

Tunisia is next door to Algeria and a far more common destination for holidaymakers. Can he tell us something about Algeria that will encourage us to visit instead?



“My opinion is biased, but I encourage everyone to visit Algeria!” he laughs. *“We are a really nice country. Tourism is not our strength, but this is what makes us special. You will meet people who want to do more, learn more, and contribute more to the world. You will feel and experience the real Algeria. If you go to a touristic place, you only see a part of it. Yes, we all need to relax, but we also need to confront the real things!”*

CADA - Cerebral Aneurysm Detection Challenge

Leonid Goubergrits

Matthias Ivantsits

Leonid Goubergrits is Professor of Cardiovascular Modelling and Simulation at Charité in Berlin. Matthias Ivantsits is a PhD student at Charité under the supervision of Anja Hennemuth. They are among the organizers of the Cerebral Aneurysm Detection (CADA) Challenge at this year's MICCAI and they spoke to us ahead of their event at MICCAI.

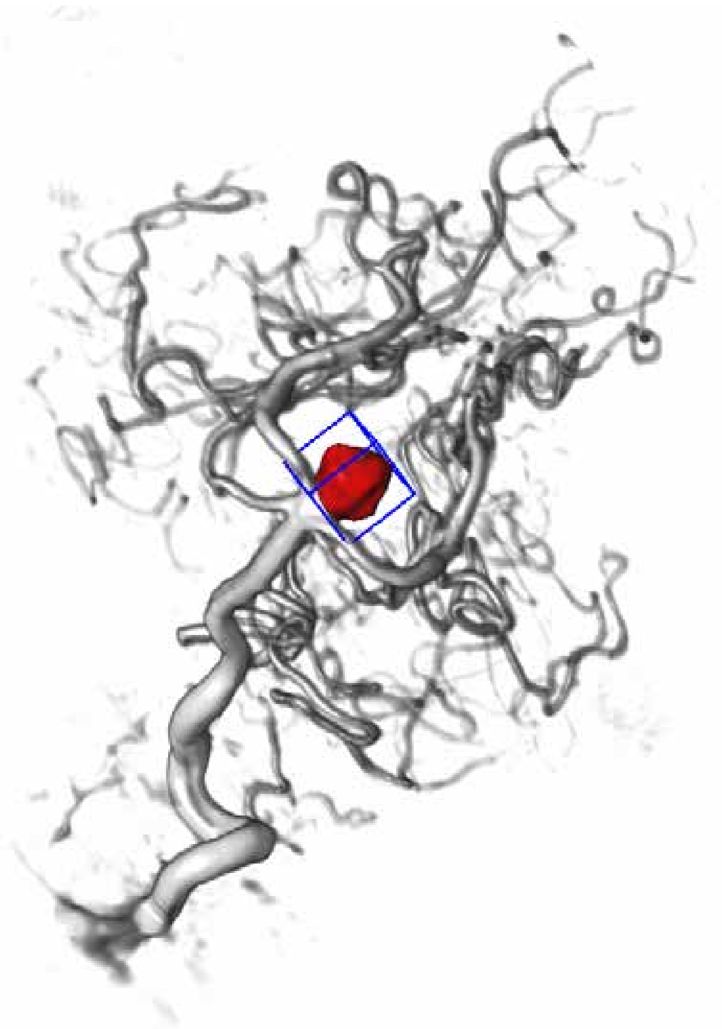
Cerebral aneurysms are a pathological enlargement of blood vessels in the brain. If an aneurysm ruptures, it causes bleeding into the brain, which in most cases is a life-threatening event. Therefore, **early detection is vital**.

These days, with the increasing use of imaging modalities such as MRI, CT, and 3D rotational angiography, small unruptured "incidental" aneurysms are being more frequently detected. Once aneurysm has been identified, a **clinician must decide if it requires treatment**. Studies have shown that only a very small percentage of these aneurysms will rupture, so **supporting clinicians to make the correct treatment decision** is a key motivation for this challenge.

"The first task for a clinician is simply to identify that an aneurysm exists," Leonid explains. *"Next is to detect its location and size. Based on this knowledge, they must decide whether to treat it or not. Then, based on segmentation and reconstruction of its 3D geometry, they will perform either surgical treatment with a clip, endovascular treatment using coils, or will use flow diverters or stents to protect it from rupture."*

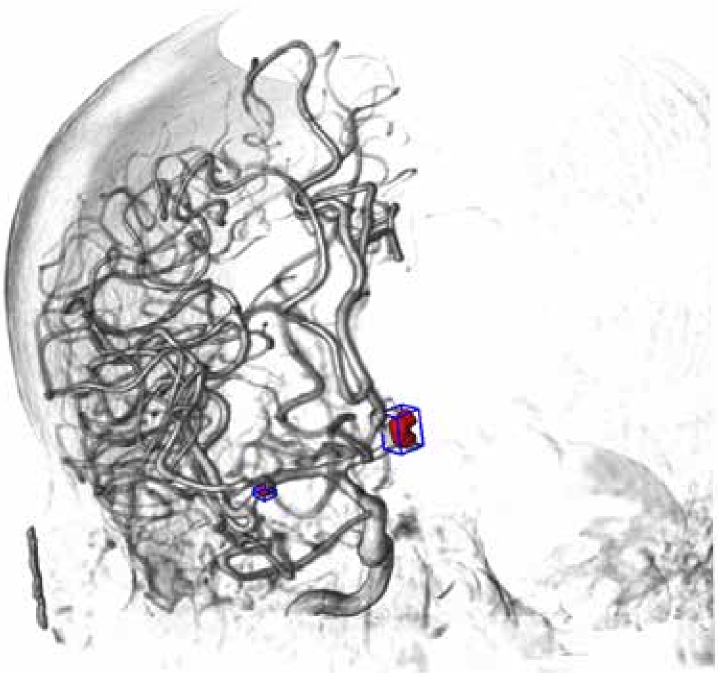
This challenge poses an interdisciplinary problem with **computational scientists, mathematicians, and biomedical engineers working together to solve**

problems in medicine. A key aim is to bring these communities closer by demonstrating to clinicians that they can trust in the science and technology.



Leonid and Matthias hope and expect this to be the first iteration of the challenge, as there are still many questions left to answer and possible future directions to explore.

"This challenge is based on analysis of ***3D rotational angiography-based data****, but clinicians work also with data based on MRI and computed tomography, so we need similar challenges focusing on other imaging modalities,"* Leonid tells us. *"Also, brain circulation is very complex, so for risk analysis it is vital to identify the location of an aneurysm in this complex system of vessels. This is another problem that we could look to solve in a future challenge."*



Finally, they are keen to point out how impressed they have been with the number of participants and the quality of their results.

"They are pretty much state-of-the-art solutions!" Matthias reveals.

Leonid adds: *"We know that participation is not usually grant funded and it is time-consuming, so we thank all participants for their efforts and for their impact on our challenge!"*

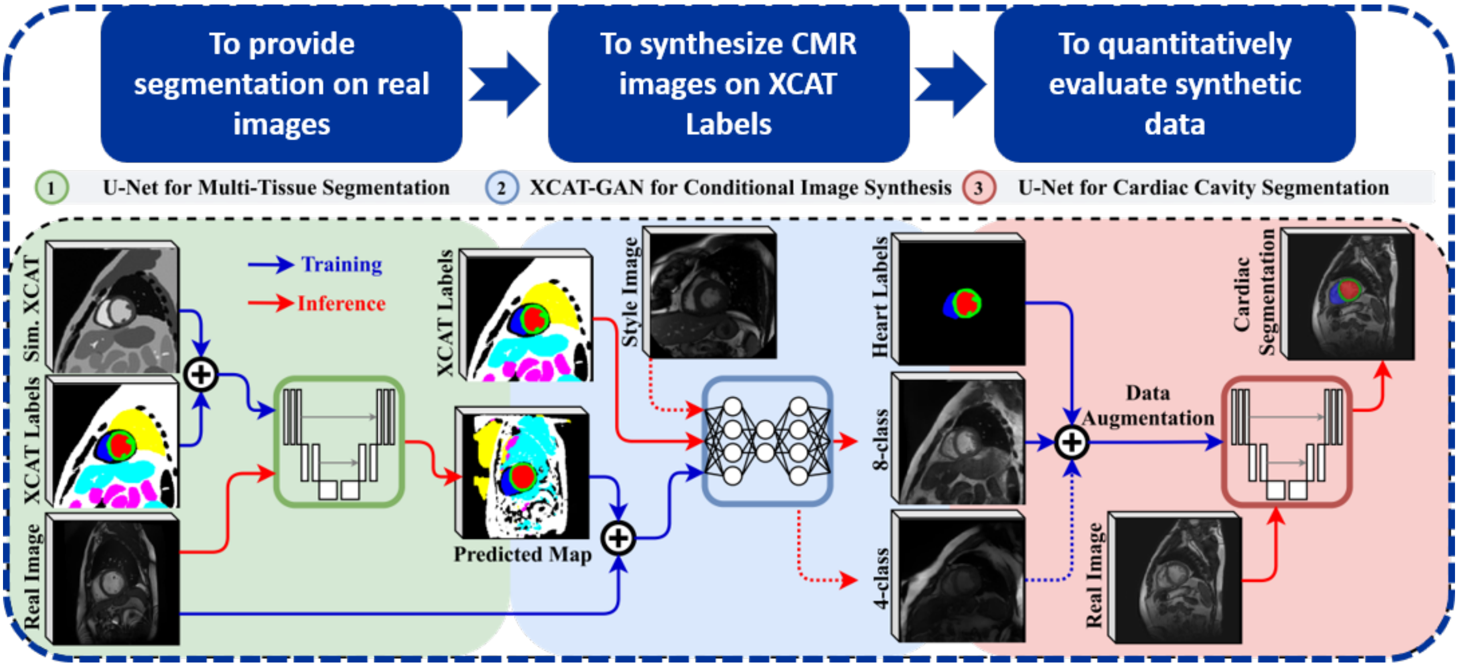
XCAT-GAN for Synthesizing 3D Consistent Labeled Cardiac MR Images on Anatomically Variable XCAT Phantoms

Sina Amirrajab is a second-year PhD student at Eindhoven University of Technology (TU/e) in the Netherlands, where he is a member of Josien Pluim's Medical Image Analysis Group. He is also part of a Marie Curie-funded European project called openGTN, which has many industrial and academic partners. His main supervisor is Marcel Breeuwer, a principal scientist at Philips and a professor at TU/e. His industrial supervisors are Cristian Lorenz and Jürgen Weese, from Philips Research Hamburg. Originally from Iran, this is Sina's first time at MICCAI. He spoke to us ahead of his oral presentation at MICCAI.

Sina Amirrajab

Sina's work is about **generating images with ground truth labels** that can be used for other downstream supervised tasks. His paper contains a number of different elements, including **an anatomical model** based on the XCAT computerized phantom; **a conditional GAN architecture** based on a state-of-the-art image-to-image translation network; **a network** which evaluates these generated or synthesized images; **and a simulation part** to simulate cardiac MRI images to be able to train an initial network with a task of segmenting real images.

"I always distinguish between simulating and synthesizing," Sina explains. *"Simulating is based on the physics of imaging modalities combined with a physiological or anatomical model, with access to parameters that you can vary. Synthesizing is based on using different kinds of techniques, like GANs and other generative models, to synthesize images based on what*



you can learn from already available imaging data."

When you are a researcher in the field of medical image analysis there is always a hunger for clinical data, he tells us, but you do not always have the annotations for that data to train your model. Even if you do, the **annotations vary from expert to expert**. This means the ground truth is changing. The idea of **openGTN** is to try to simulate or synthesize a large number of images with underlying ground truth labels. It uses a model to generate these labels, called true ground truth labels, and tries to simulate or synthesize images that are as realistic as possible to the real imaging data. With these realistic images combined with ground truth labels, **it can train a segmentation, registration, or other network and model to analyze real images**.

Within the openGTN project (https://opengtn.eu), Sina had two options – either to go for **a cardiac application or a brain and spine application**.

"I thought about it and I talked to my previous supervisor because he had experience in cardiac MRI," he recalls. *"In cardiac MRI you have moving objects – the heart, which is beating, and the lung, which is moving around.* ***I found it very challenging to try to deal with this motion and to incorporate it in some way into the simulation****. That is why I decided to focus my attention on cardiac MRI."*

The idea behind this work has been forming since late last year, and earlier this year the team had a paper accepted at **MIDL 2020**. Sina says he is particularly proud of the time he has spent collaborating with his co-authors, Samaneh Abbasi-Sureshjani and Yasmina Al Khalil, brainstorming ideas.

We are keen to know what it is like to be a student working with MICCAI Fellow **Josien Pluim**, the head of the **Medical Image Analysis Group at TU/e?**

"She is very kind and always has vision about the things that we are doing, both now and in the future. She is very supportive. That is a nice feeling that I always receive from her."

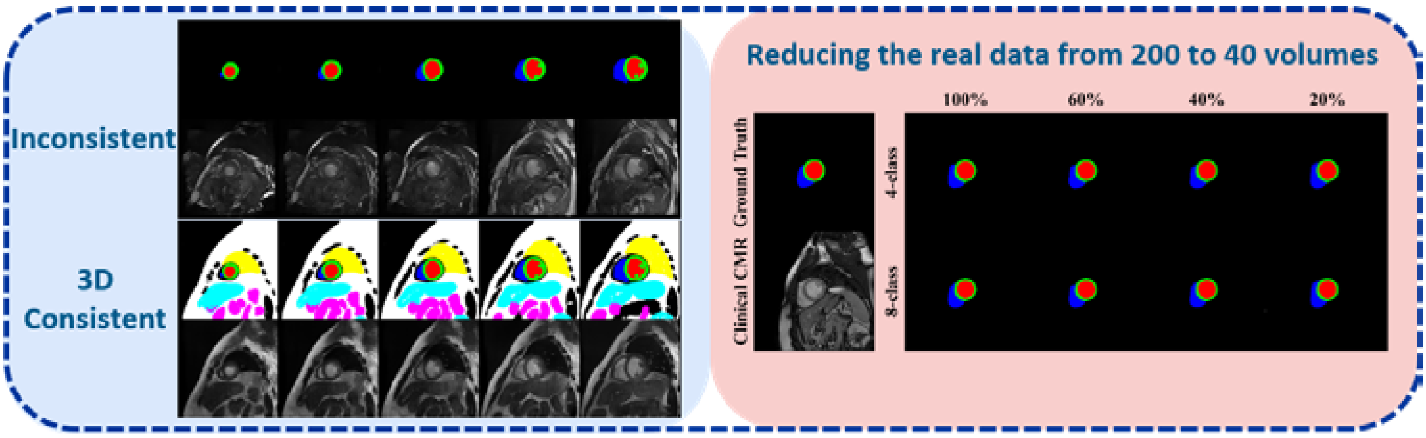
Finally, we ask, what does the future hold for Sina?

"I am opening up my possibilities. I would like to continue working in research – either by doing a postdoc or by doing some very research-oriented work.

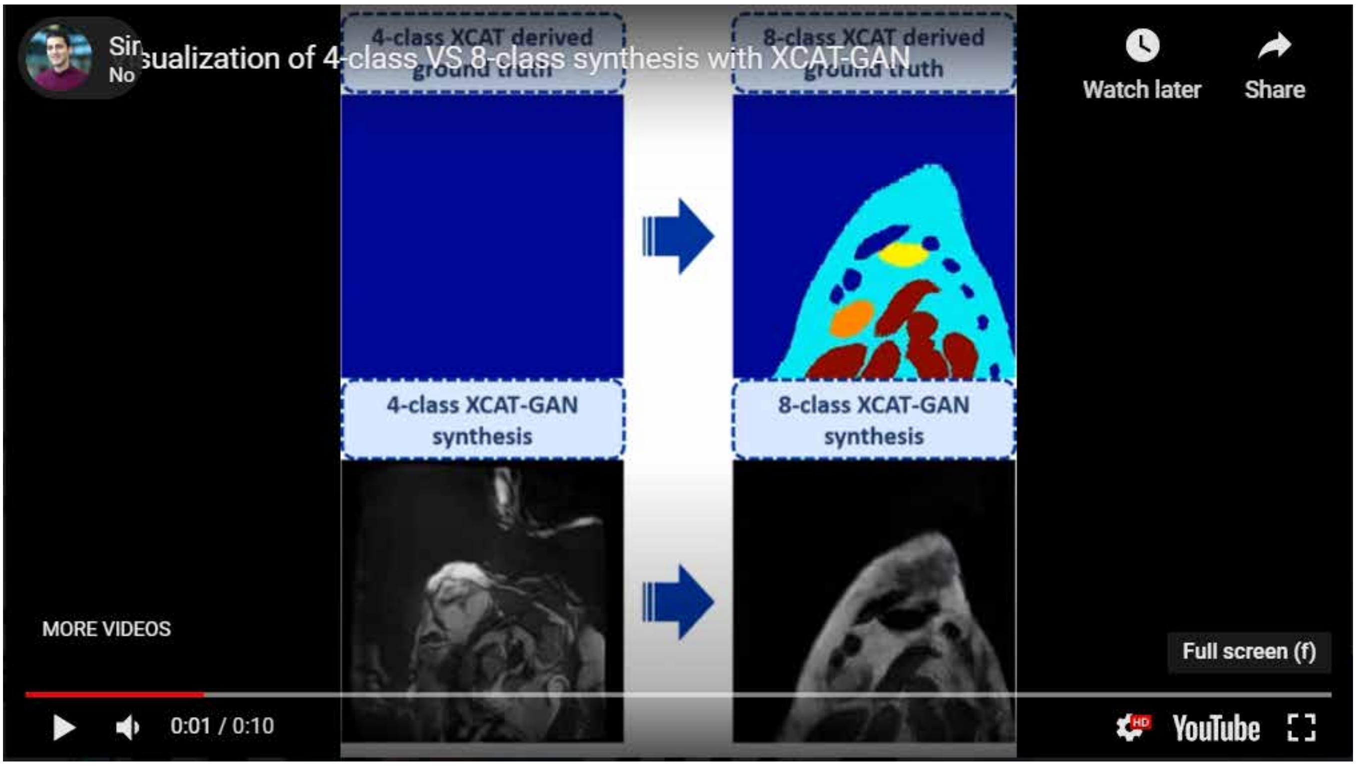
Ultimately, I would love to have one leg in industry and one leg in academia." Maybe in a few years we will be interviewing a student and hear, **'I am supervised by Sina**.'

"The idea of openGTN is to try to simulate or synthesize a large number of images with underlying ground truth labels"

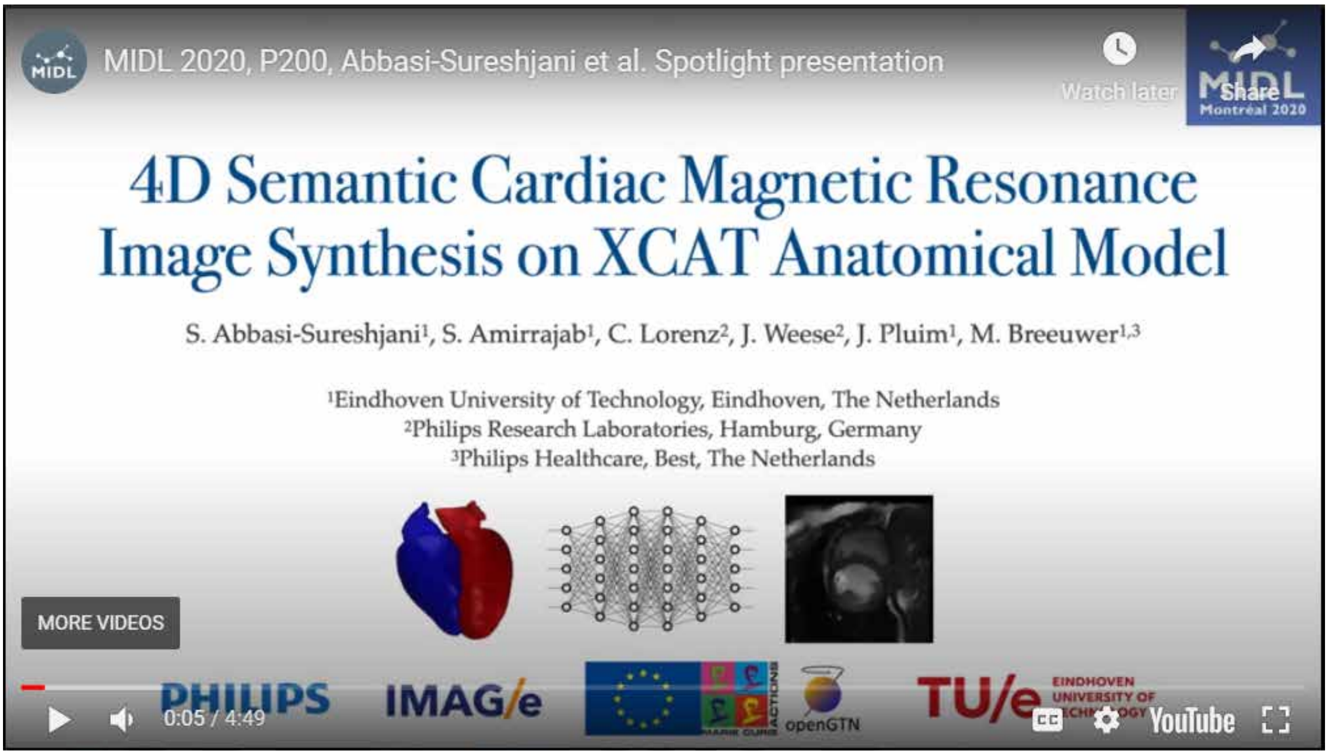
"Yeah, why not?"



Visualization of 4-class VS 8-class synthesis with XCAT-GAN



Paper Presentation on Video



Self-supervised Contrastive Video-Speech Representation Learning for Ultrasound

Jianbo Jiao is a postdoc researcher at the University of Oxford, advised by Professor Alison Noble and Professor Andrew Zisserman. His work is about self-supervised representation learning with multimodal ultrasound data. He spoke to us ahead of his oral presentation at MICCAI.

Jianbo Jiao

This work focuses on both **ultrasound video data** and **speech audio data** from the sonographer. The method was validated on a large-scale clinical ultrasound dataset called **PULSE** – short for **Perception Ultrasound by Learning Sonographic Experience**.

Currently, almost all research areas are using **deep learning tools**, and most of them rely on human annotations to train their models. However, with medical images which require specific expertise, these human annotations are not always easy, or even feasible, to acquire.

This fact motivated Jianbo and his team to address the problem of self-supervised learning, which means learning meaningful representations or knowledge from the data itself without any manual annotations. The challenge here is to define a self-supervision signal to supervise the model so that it can learn some representations.

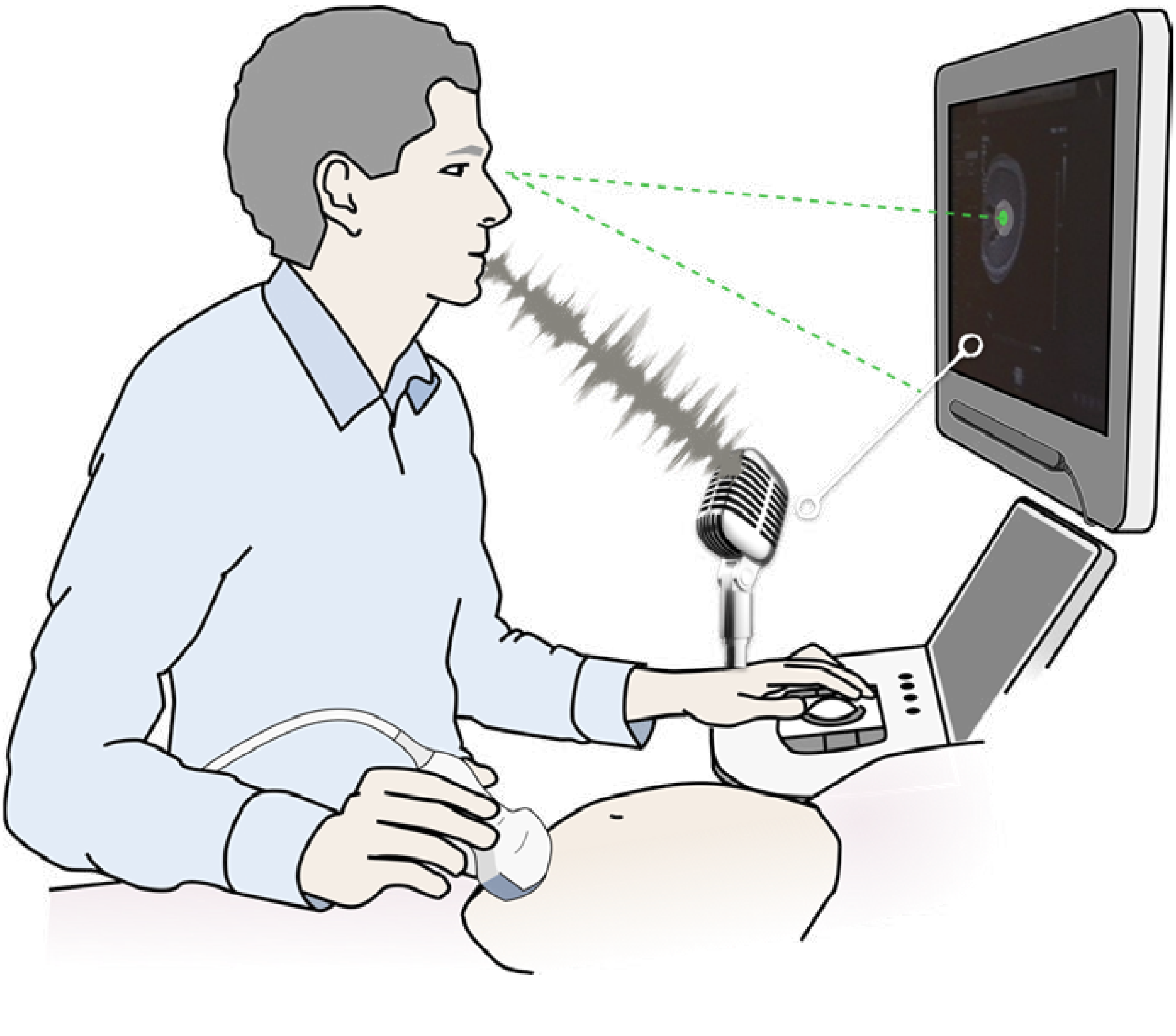
This work starts from a basic model to build the correlations between the video and speech audio data but proposes some **new techniques** to address the specific challenges with **medical images**. Instead of simply using positive and negative pairs for training, it proposes hard-positive and hard-negative pairs to force the model to "learn harder" and learn stronger representations.

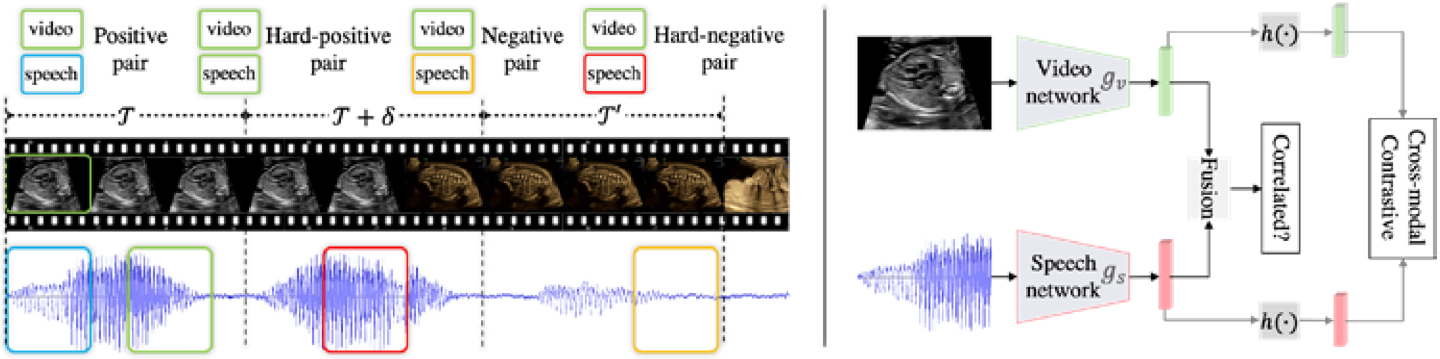
"For natural image video and its corresponding audio, there are very dense correlations," Jianbo explains. *"For example, with someone playing the piano, whenever the actions appear, the sound will appear. However, for medical*

data, the correlations between video and audio are not so dense and strong. The speech audio usually has a very sparse correlation with the corresponding video. In our case, when the sonographer or doctor is scanning the ultrasound data, what appears on the screen is not necessarily related to what the doctor is talking about."

To solve this, the method introduces **cross-modal contrastive learning** to encourage the positive pair to live closer and the negative pair further away in the embedding space. The team found there was a lot of background noise and uncorrelated conversations in the audio data, so further proposed an **affinity-aware self-paced learning scheme** to detect these unrelated signals and adaptively learn the representations accordingly.

Currently, this work only considers the audio and the video data, but the aim





is to **incorporate more modalities to build a multimodal method** to learn the representations. The PULSE dataset includes **eye-gaze tracking data** and **motion data** from the sonographer, for example. The team would also like to explore what the model actually learns for this specific representation and interpret how and why it learns such useful representations.

Finally, we are fascinated to know – what is it like to work with such eminent professors as **Alison Noble** and **Andrew Zisserman**? *"I learn a lot from them and am very grateful and honored to be advised by them,"* Jianbo tells us keenly. *"They are both very supportive of my research and give me sufficient freedom. On the one hand, I get the **computer vision perspective**, and on the other,*

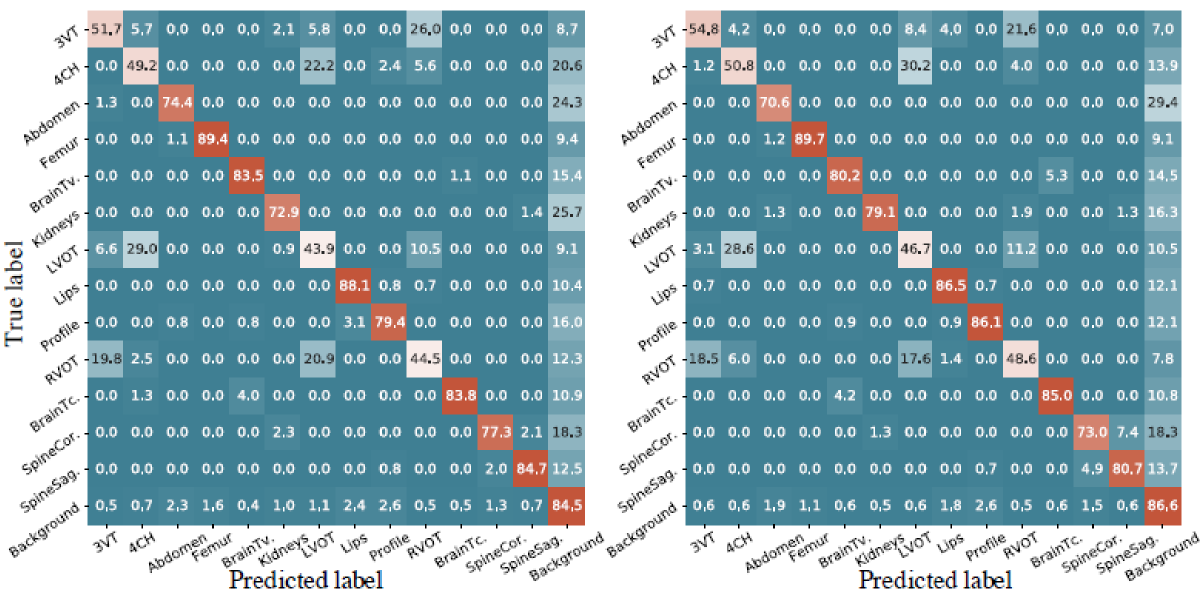


Fig. 3. Confusion matrix on standard plane detection. Left: *Video*. Right: *Ours*. (Best viewed in digital form.)

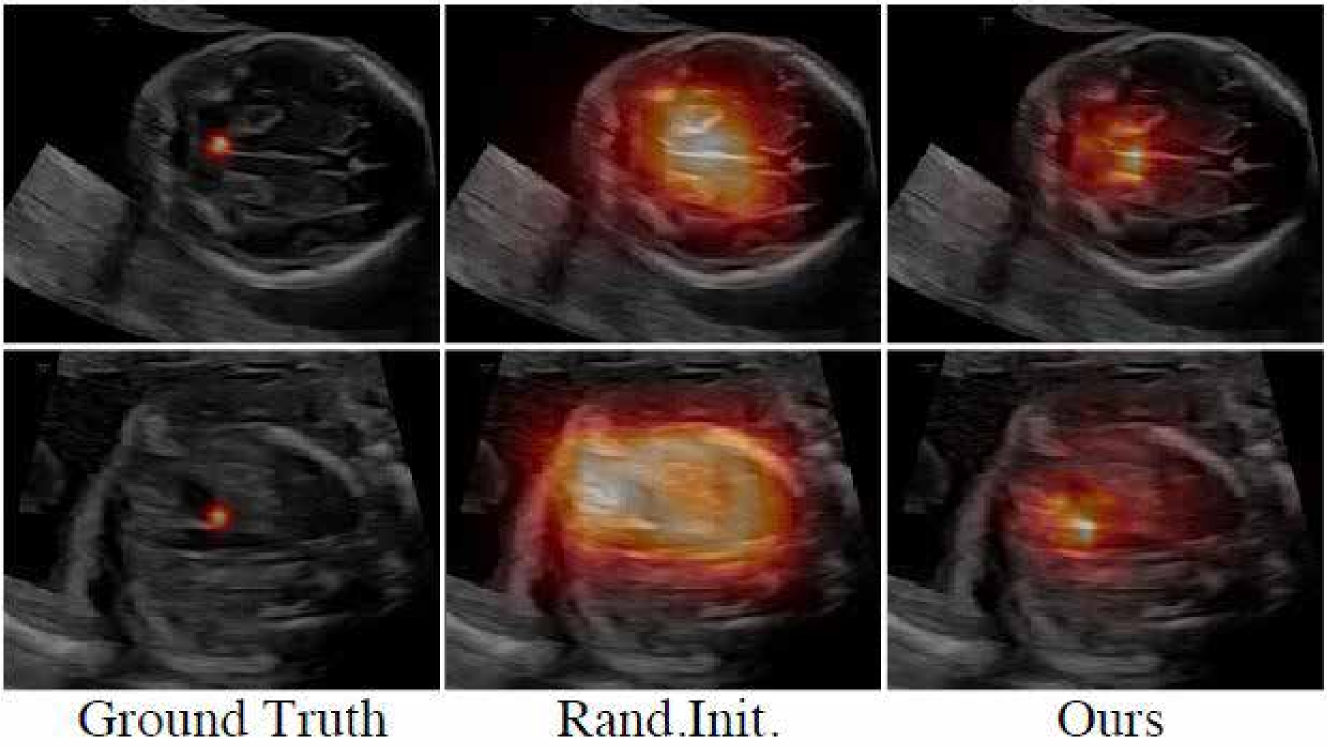
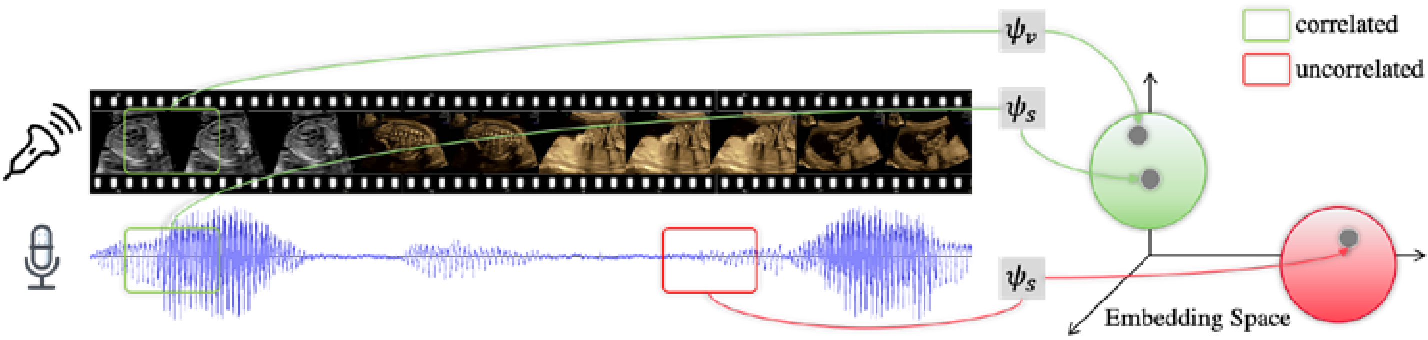


Fig. 4. Qualitative performance on eye-gaze saliency prediction.

*I get insights into **real medical imaging applications**. They share a big common quality for providing detailed and constructive suggestions and advice, which was beyond my expectations before I came to Oxford. They are very energetic scientists working for a long time every day, so I really appreciate their guidance. I benefit a lot from their long-term vision."*



David Menashe is an algorithm team leader at RSIP Vision. He writes about our market-leading work using AI to enhance medical ultrasound applications.

by David Menashe

Ultrasound is an easy-to-use, portable, and inexpensive technology, with many different applications, making it one of the most promising technologies in medical imaging. An excellent current example of the versatility and importance of the technology is the recent application of lung Ultrasound to COVID-19 diagnosis, and specifically to rapidly and effectively identifying severe cases.

However, ultrasound images can be **noisy** and are not always easy to interpret. It takes a lot of skill by the technician to **produce a good image** and even more skill to **interpret images** or videos and derive the correct diagnosis.

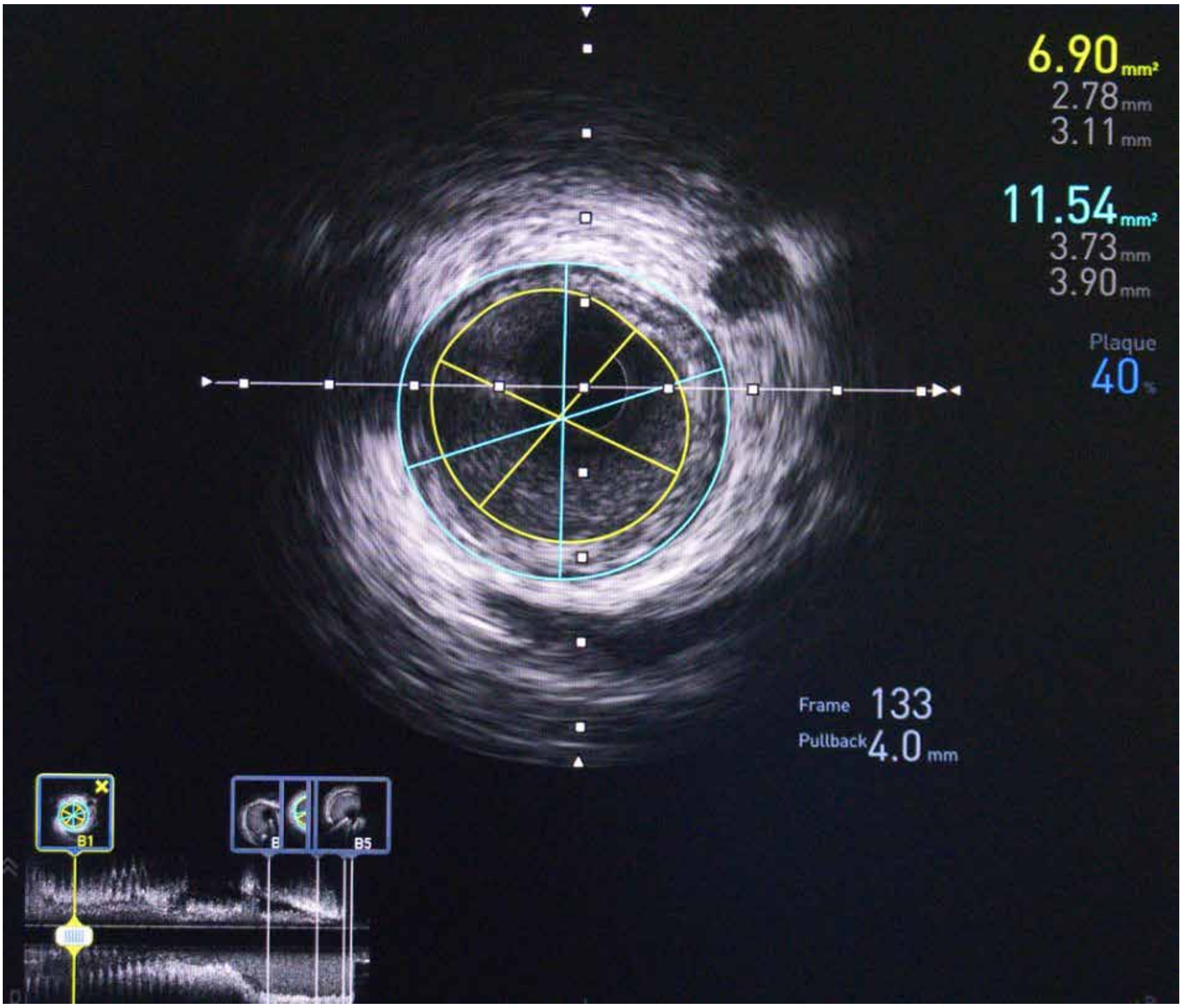
In **heart echo** studies, for example, there may be 60-70 video clips of the heart and nearby blood vessels to analyze, with various calculations and measurements to make to obtain a full picture of what is going on. Cardiologists use these clips to diagnose a multitude of different diseases and conditions, which can be monitored and followed up at regular intervals. It is a **time-consuming process when performed manually**, and often clips can be irrelevant or of poor quality.

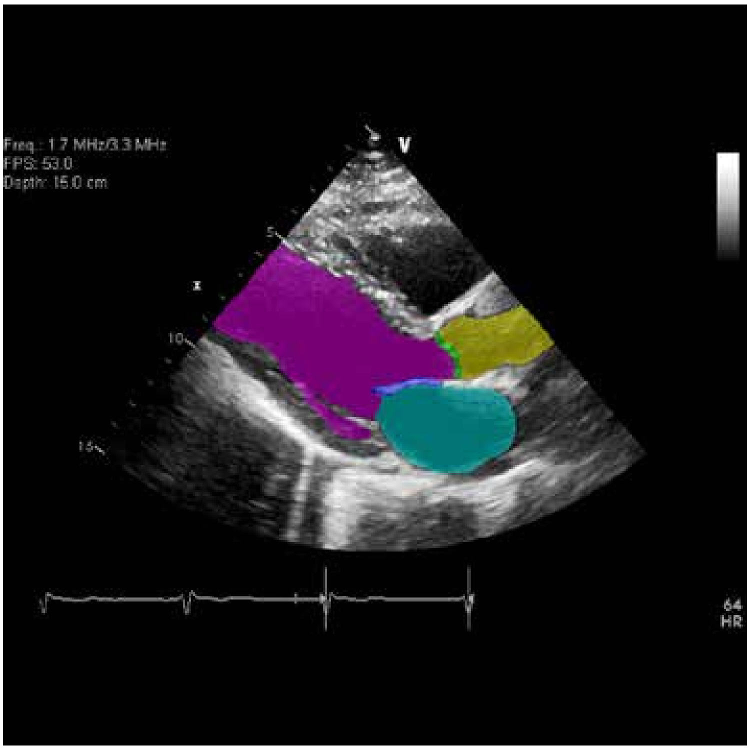
This is where **artificial intelligence** steps up. AI can help experts to sift through a huge amount of data, identify critical images or clips, and eases the workload for medical professionals involved in diagnosis.

At **RSIP Vision** recently, we have used AI to enhance the analysis of heart echo studies, including analyzing the video clips to assist the physician to select only those which are good quality; sorting the clips according to what view and cross-section of the heart they show; detecting features on the clips, such as valves and chambers, and automatically taking measurements, such as one of the key measurements in cardiology, the ejection fraction, which measures how efficiently the heart is pumping the blood out of the left ventricle. At an even higher level, we have begun to develop models which can predict certain pathologies directly from the raw Ultrasound image, thus not only providing tools to the cardiologist, but also assisting in the actual diagnosis.

We have been working on several other applications of **AI in ultrasound**, such as identifying and characterizing lesions in **breast ultrasound**, which is a **simpler, cheaper and safer technology** for detecting breast cancer than mammography. There are **no issues with radiation in ultrasound**, so you are free to have as many as you want! Another example is using transrectal Ultrasound (TRUS) to help guide prostate intervention procedures.

Fetal ultrasound is another example of AI-enhanced imaging. Every parent knows how important fetal ultrasound is in monitoring pregnancy. The latest technological advancement in this field is a **portable ultrasound device** that enables pregnant women to perform at-home scans, with feedback via **telemedicine**. In this scenario, AI tools help guide the user through the process.



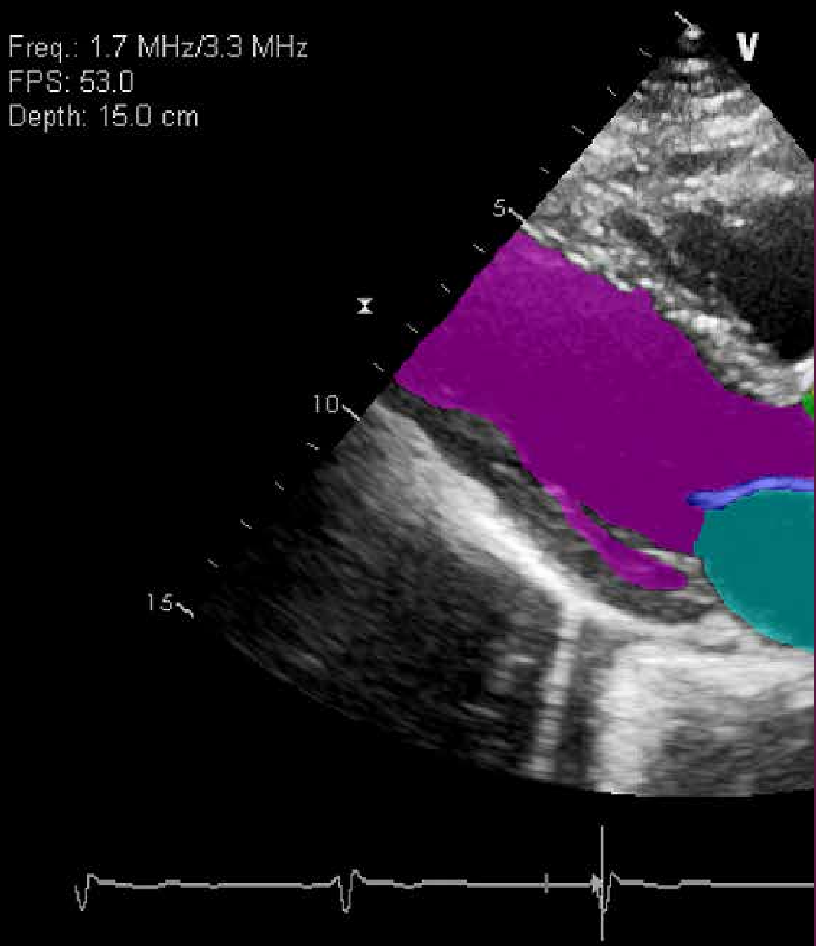


Medical device manufacturers do not always have the level of expertise needed to fully leverage a deep learning model as it is not their main field of work. Their focus is on getting the best hardware at the cheapest possible price. Whilst some may understand the theory of how to train a model or what parameters to use, they do not have a thorough understanding of the data and annotation process.

Before training a deep learning model, it is vital to understand your data. Is it sufficiently variable to cover the application being developed? There are so many factors that come into play here, including the **quality of the ultrasound**, and characteristics of the patient, such as age, gender, and **BMI**, which can all effect the size of the organ being imaged. Once the data is ready, you need to work with your **echo specialist or radiologist** on the best labeling and annotation procedure. It is an iterative process which is **partly about the algorithms and partly about the data**. Try one approach, train a model, and then give the data back to the echo specialist who will look again and may make further suggestions.

At **RSIP Vision**, we have a highly skilled team, with many years of experience between them of developing and managing **artificial intelligence and deep learning projects**. It is what we do every day. We also have access to specialists who know how to interpret data and help us to label and prepare it in such a way that we can get the most out of a deep learning model. It is a complex process requiring extensive experience and, it goes without saying, a great deal of patience!

Do you want to learn more about AI in Ultrasound?



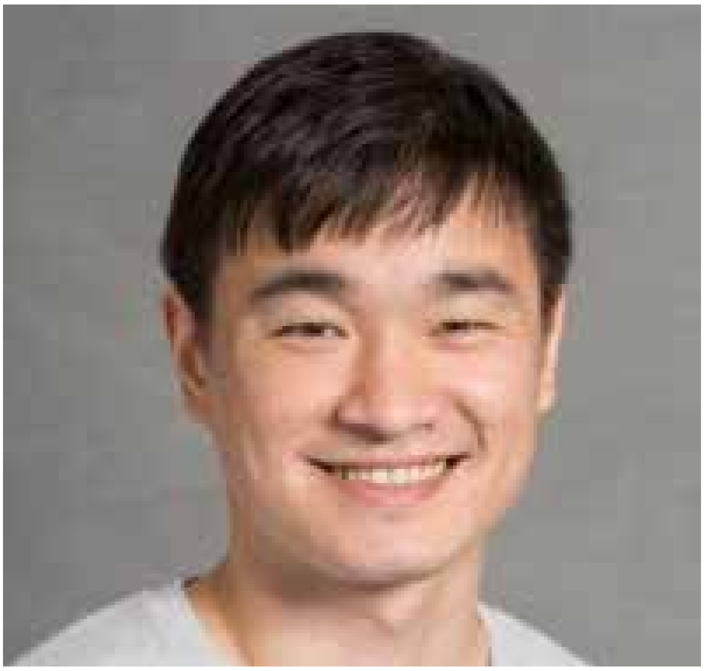
Did you miss our interviews with Yipeng Hu from UCL and with Kristin McLeod from GE Healthcare?

Here they are again, on the following pages!

Yipeng Hu is a lecturer at UCL and a member of the Wellcome / EPSRC Centre for Interventional and Surgical Sciences (WEISS) and the UCL Centre for Medical Image Computing (CMIC). As part of our series of interviews with experts from the MICCAI field, we speak to Yipeng about his career so far and progress and innovation in ultrasound-guided intervention.

Yipeng's medical imaging journey began during a biomedical engineering degree in China. In his third year, he chose to focus on **medical image processing**, the only computing option available, and hasn't looked back since. After his degree, he moved to England and took on a Master's in **biomedical engineering and medical imaging** at UCL. There, he worked on a research project with Professor Dean Barratt exploring **motion modeling for prostate intervention**. He went on to do a PhD with Dean Barratt and his second supervisor, Professor David Hawkes, who founded CMIC in 2005. The PhD topic was **multimodality image registration between MR and ultrasound for prostate cancer patients**.

"One of the most exciting times in my life was going on to develop my PhD work into an image-guided system called SmartTarget for ultrasound-guided interventions for biopsy and

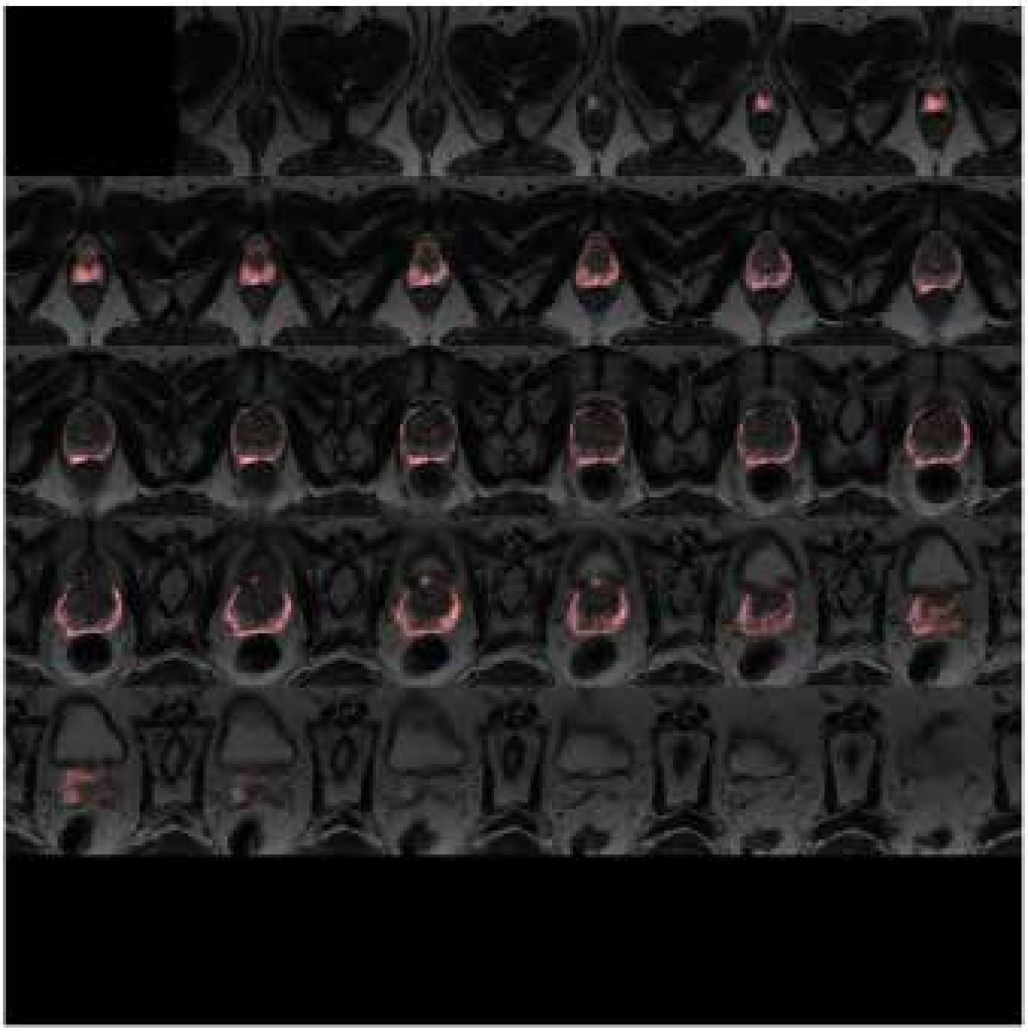


focal therapy," Yipeng beams. *"We had a spinout company and it was eventually FDA-approved and CE-marked. We went from very basic methodology development using machine learning to motion modeling for multimodality image registration to completing a guidance system that can be used in theatre. That experience was so interesting. It's basically what I did in the first decade of my research. The system has just been bought by an American urology company."*

Despite having successfully translated his own work into a real-world product, Yipeng tells us there is ongoing debate about whether translation should be part of academic research. In his view, research is still a vital part of development. WEISS is breaking down the barriers here with its unique set-up that brings engineers and clinicians together under the same roof. Academic work doesn't stop once a paper has been written.

Like many other fields, the application of **deep learning** has revolutionized medical image computing, and it started early. *"I think it's been about five years now,"* Yipeng says. *"One of the reasons is because we are very close to computer vision.* ***Surgery and interventional applications*** *are usually at a later stage of the curve in terms of adopting new methodology, but it's starting now. The advantage is everything can be done very quickly in real-time, which is essential during surgery."*

However, there are challenges. Data is a key requirement for deep learning algorithms to work, but for ultrasound, **data availability** can be limited. Also, most applications involve heavy human interaction, whereas in computer vision, you work with a picture or video. *"I personally see there's a traditional image-guided interventional subfield and a robotic field,"* Yipeng explains. *"The robotic field is assuming robots will be* ***controlling all medical devices*** *in the future. With that as our end goal, we try to make our algorithms as automated as possible. We're still very close, but these two fields will merge at some point."*



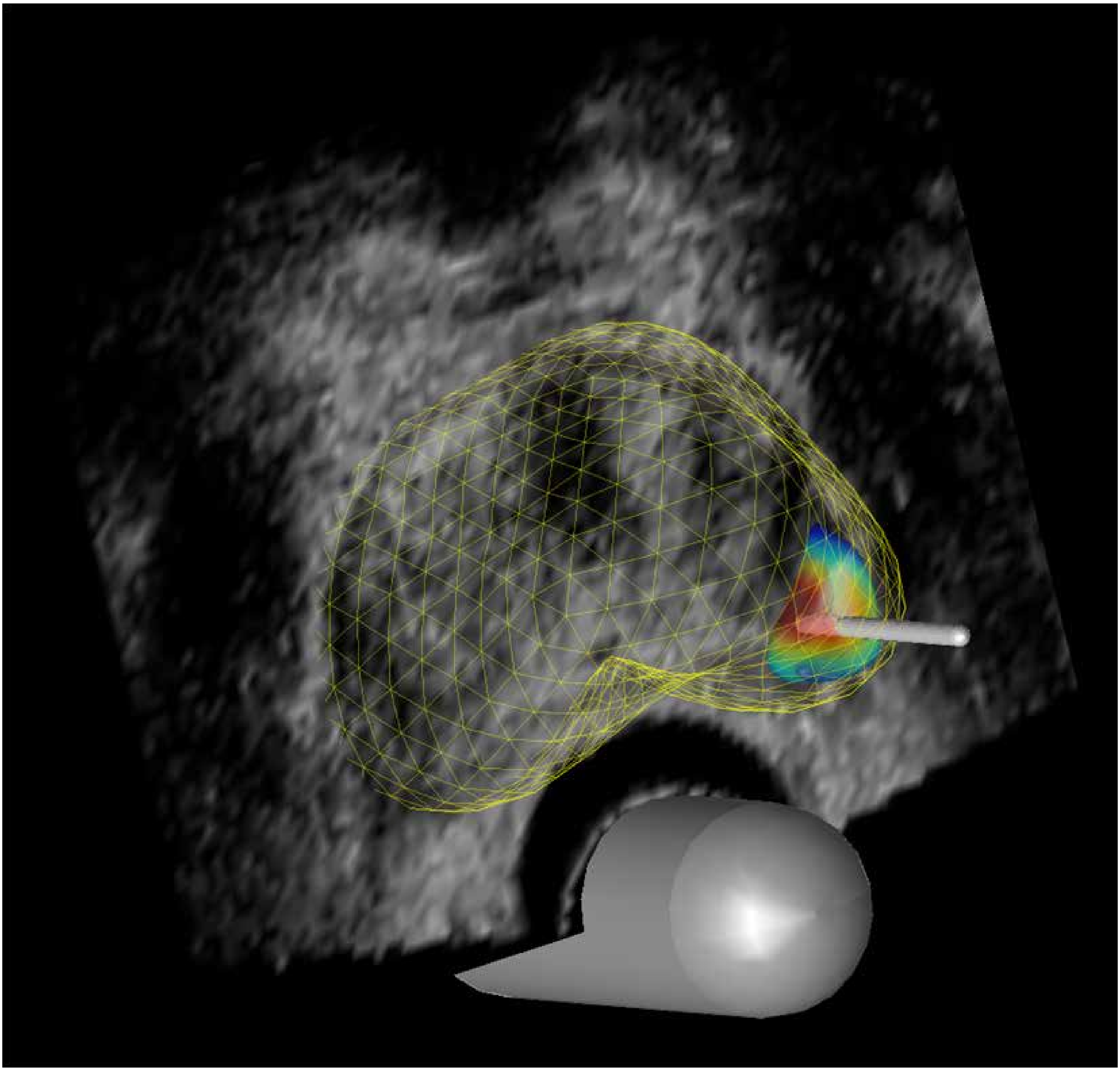
He is encouraged by a growing trend for openness, but laments the fact that not everyone is on board with it yet: *"We can't wait to share our code and can't wait to share our paper and put it on arXiv. I see it as a positive thing, but our field is still lagging behind. Particularly in industry. If you talk to people, they are pretty open, but you have no idea what their five-year plan is. I understand industrial secrets, but as an academic researcher, I would like to know what they are working on so that we don't crossover. It would make the whole thing more collectively efficient. In making contributions to translational stages, that openness and sharing of information is so important."*

He offers a word of caution for people using deep learning technology in medical image computing and surgical interventions. *"Quite a few people are inexperienced when it comes to deep learning. We are all guilty in that department,"* he says modestly. *"Fancy algorithms and promising results are all well and good, but without **proper methodology, design and analysis**, they won't translate easily into clinical impact."* However, he is pleased to see more groups are looking at ways to validate their methods using **statistical principles with real-world data and prospective patient studies**.

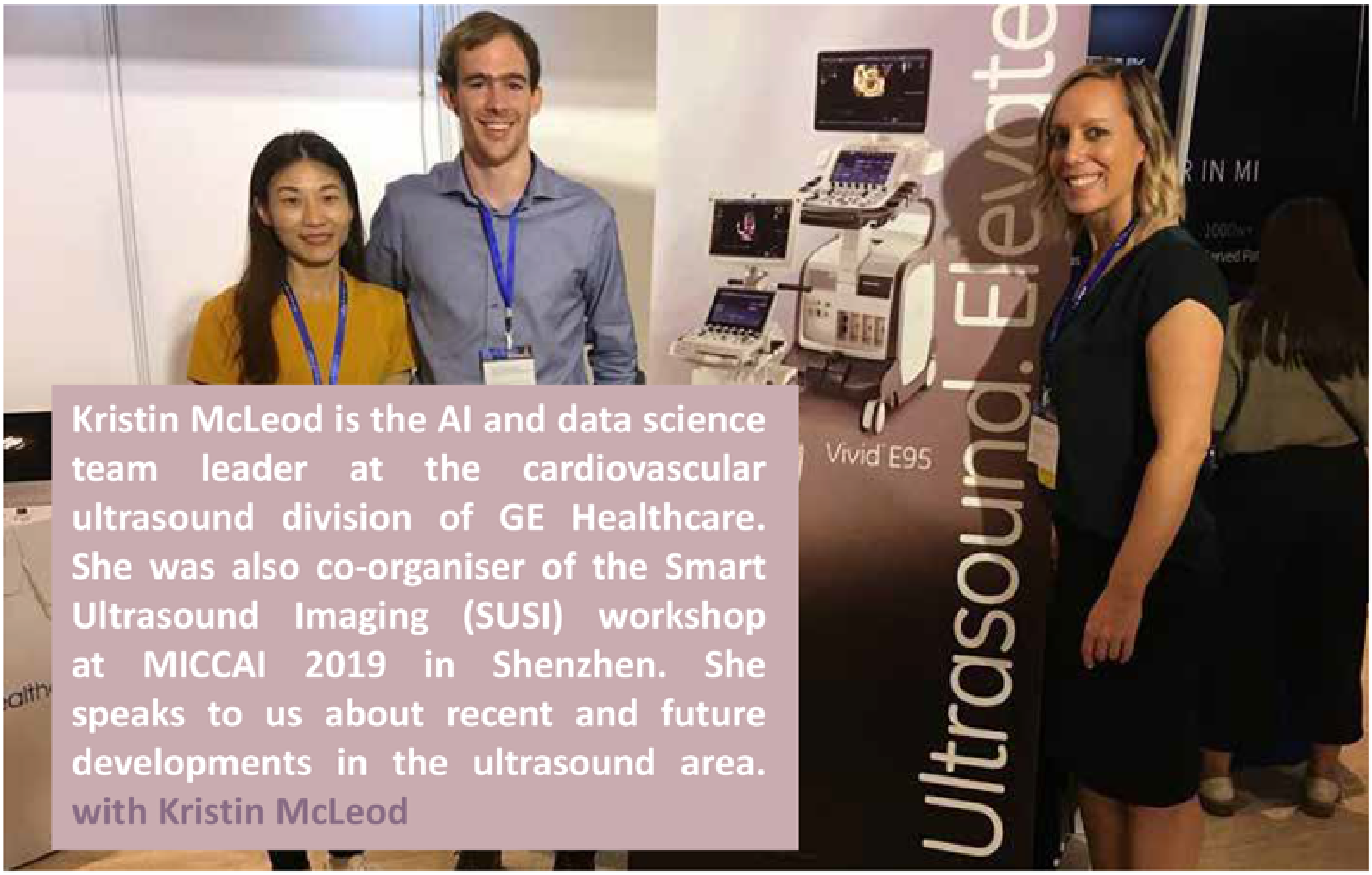
*"**Clinical trials** are still a gold standard for determining how useful something is clinically. There is no question about that,"* he affirms. *"If you have a deep learning algorithm, instead of using retrospective data from other clinical trials, you should start your own clinical trial to validate it. But if you look around the world, how many are driven by a computer scientist? Almost none."* Although Yipeng is optimistic that the tide is starting to change, and he hopes that in the future, academic research will lead the way. *"We are impartial and well placed without any other conflicting interests."*

Yipeng is hopeful that we will all make it to **Lima, Peru** in October for **MICCAI 2020**, where he will be co-organizing the **Advances in Simplifying Medical Ultrasound (ASMUS) workshop** with **Alison Noble** from University of Oxford and **Stephen Aylward** from Kitware. Stephen previously organized the **Point-Of-Care Ultrasound** (POCUS) workshop and it is an extension of that. If international travel is still difficult, he promises us it will be virtual.

Ultrasound is a particularly interesting modality right now, but it has its challenges, so research institutes often opt for using other methods. Yipeng tells us this creates a bias which makes it easier to access well-curated data if you are working with **neuroimaging**. *"That's* *why I think efforts like this are necessary and why we are doing it,"* he explains. *"We're trying to get more people interested in ultrasound. People don't realize its potential. For some reason, in the UK in particular, less and less radiologists are using ultrasound skills. They tend to use CT and MR, which are more expensive, but easier to interpret. They require far less skill and experience. That's not necessarily a bad thing, but we feel very strongly that you have to strike a balance."*



This article was first published on [Computer Vision News of June 2020]



Kristin McLeod is the AI and data science team leader at the cardiovascular ultrasound division of GE Healthcare. She was also co-organiser of the Smart Ultrasound Imaging (SUSI) workshop at MICCAI 2019 in Shenzhen. She speaks to us about recent and future developments in the ultrasound area.
with Kristin McLeod

Smart Ultrasound

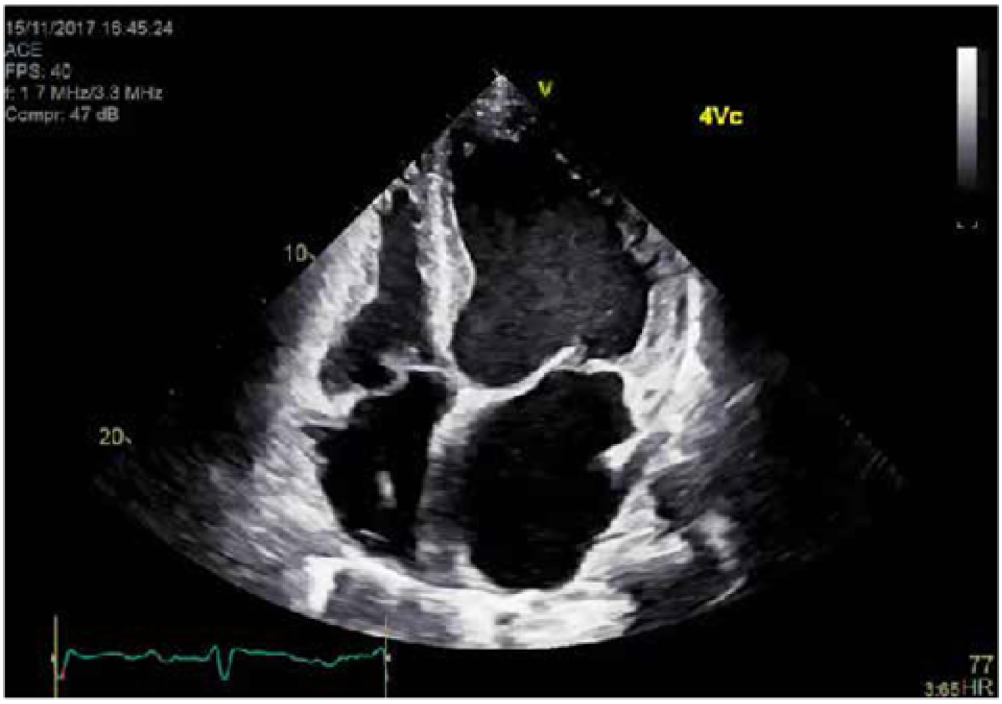
Kristin tells us that the **SUSI workshop** was the first time people had gathered more generally across the ultrasound domain, including general imaging, women's health imaging, hand-held support care imaging and cardiovascular ultrasound. It was a chance to present the latest works in an arena where they hadn't typically had much coverage. Participants could see where different technologies overlapped and leverage the synergies between modalities.

There are pros and cons of ultrasound, but thanks to the latest advancements in medical imaging and image processing technologies, Kristin says we're starting to work towards more of the spatial resolution that is typically seen in CT and MRI. With greater use of AI to augment the imaging, we're seeing much better image quality from ultrasound.

However, there's been a lack of standardisation in ultrasound compared to other imaging modalities. With CT and MRI, there's quite a standard way to acquire the images, but with ultrasound images you have to find an acoustic window between, for example, the ribs or other obstructing objects in the foreground.

Thanks to developments in terms of guidance and standardisation within the acquisition itself, the robustness of ultrasound as an imaging modality is improving. This will no doubt lead to better patient outcomes, but Kristin says the biggest breakthrough she has seen in this area recently is the use of **reinforcement learning** in terms of the algorithms themselves:

"Reinforcement learning has been a big step forward. This is where we have the potential to develop algorithms that can outperform humans. The algorithms that have been used in previous years were designed in such a way that they could reach human performance, but due to the way they were structured, could not outperform it. With this being more and more widely used, the accuracy and the usability of these types of algorithms will increase quite substantially over time."



Great image quality on GE Healthcare's 4Vc ultrasound probe

Thinking ahead to the next 12 months, Kristin would like to see more collaboration **between the academic and industrial sites**, and more clinically driven algorithm development. In the past, universities have developed advanced algorithms to solve problems that are not necessarily suitable within the clinical workflow, so she'd like to see that academic work being translated into **clinically useful tools**.

*"I think the funding bodies are pushing for this kind of collaboration. At GE, we're initiating a **collaboration between universities and hospitals** where there is a requirement to spend half of the time in **industry**. That's a step forward, but it also requires a change in mentality from all sides to think beyond our own priorities to the common goals and how we can reach those together. It requires a bit of compromise on each side, but that's what is needed in order to really make this happen."*

This article was first published on [Computer Vision News of January 2020]

"... that has taken years and a lot of expertise!"

Naureen Mahmood is the CEO of Meshcapade, a tech startup based in Germany. She has been a research software engineer at the Max Planck Institute for Intelligent Systems. [**Find here 100 more interviews with Women in Computer Vision!**]

Naureen, where were you before Max Planck?

I also attended other institutes. I did my Master's at Texas A&M. They have a computer visualization and animation program, which is sort of a cross between computer vision and computer graphics. Then I joined Max Planck as a research engineer. I was there for almost 6 or 7 years. Then we started this company, a tech startup here in Tübingen.

What inspired you to start this company?

That's a very good question! [*laughs*]

Working at Max Planck, there was a lot of research being published and also a lot that was never published. It was useful, but it wasn't really novel, or it wasn't packaged up in a publishable manner. What we found was that at almost every conference, we would be contacted by other researchers, other institutes, and then slowly other industries, asking us questions about how they can use our research in their applications and asking for ways to make it easy for them to use. To go back a little bit and tell you about the technology itself, we are building 3D statistical models to understand how humans move and how humans look. This goes from the surface of skin deformations, how the skin deforms when we are moving, also how the body sizes change across a population. We combine that into this statistical model. It's parametric, which means you can create different human identities. Independent of that, we can also create many different realistic poses. The muscles and the surface of the skin automatically deform for each different pose. This is something that we have learned from the 3D scanned data.

You are not the first to work on that. What is special about your work?

There are many people who are working on bodies from images, bodies from scans. The special part really is that nobody had learned this way, from real 3D scanned data. By the time I left Max Planck, they had multiple large full-body 3D scanners. They were one of the first in the world to have a full body 3D scanning system. Then they also upgraded it later on to get a full body, 4D scan, so that it would be able to capture the dynamics of the soft tissue when you are moving. This helped us create a very large dataset of high-resolution 3D scans. Learning how human identities change across a population was a very hard process. Since there are so many 3D scanning systems, a lot of people can do that. However, the many years that it has taken us to get this, that's a very important ingredient here. A lot of people have the ability to capture a 3D scan. Every 3D scan is sort of

"We were the first ones to do that in a way that was easy to use!"

a mess of points. You have to be able to create coherence across those individual messes of 3D points. That's something that we learned to do automatically. That made it possible to grab 3 scans from any source and take statistics from all of these different 3D scans and create this coherence. We were the first ones to do that in a way that was easy to use.

We also learned the muscle deformations. That was the one thing that nobody else had at that point. Still, nobody was able to do it as well. I'm not taking credit for it! [*laughs*] So many peoples' work has gone into developing this system: ten, twenty different people who have come and gone. It has been years of work. That's why it's very special.

Any other company that has enough resources can collect the data. Curating that data and going through and creating this coherence (which is also creating clean 3D datasets and then creating statistical models out of it), that has taken years and a lot of expertise. Still, there are many people out there who are not using this 3D model system. So we have developed these 3D models, and now there are people who can build on top of that who can create systems that can identify bodies in images. In the background, they still have a 3D model, and that's what people use. They use our model. This is used in methods for automatically cleaning 3D scans, automatically detecting bodies in images, automatically detecting RGBD data. The application grew from there. Now, people are also using this same model to create lots of synthetic data, so that they can train AI systems for autonomous driving. They can create synthetic routes for autonomous driving systems to learn from. People use this model because it can create the most realistic looking people, realistic looking body shapes and deformations from motion. That's why people come to us!

We forgot to mention one really important fact about this venture. You are not doing it alone. You are working with Michael Black! Can you tell us about Michael's influence on this project?

This whole idea of developing a 3D body model was something that Michael and his group were already working on, when I joined Max Planck. They were using an older version of the model that was called the SCAPE model. They created their own version of it using their own data. When I joined, we started developing applications around it using this older version. That's when we started getting a lot of requests: we would go to conferences and find people asking, *"How can I use it for my new application that I'm planning?"* That's when we came up with the idea to make a newer version, which is even easier to use than SCAPE. That's how we started working on this model called SMPL. This is also a play on words, I guess. We wanted this model to be easy to use and simple. That's why we made the acronym SMPL [*laughs*]. Michael is very fond of these cute acronyms.

How is it to work with Michael? Our readers want to know!
Michael has a brilliant mind, but at the same time, he is also very humble. He listens to everyone's ideas. He lets everyone have a chance to speak. We had group meetings at the department, and everybody would have an equal say. Even if they don't have a background in computer vision, and they don't understand how the algorithms are working, he would want feedback, from artists even and the other folks in the group. "*What do you think?*" "*Does it look realistic?*" "*Does it look like something you could use?*" "*Does this make sense to you?*" It's a very important trait, I think: being open to everybody's feedback. At the same time, he is a great advisor. I have been lucky to have worked with a lot of great people in the past. Michael has no doubt been a great example of those. He takes responsibility for anyone that works with him and also their future. He would do regular meetings and sessions with his students and help them all the way through to their graduation, and even further on for getting jobs and for their job interviews. At the same time, for the technical staff and everybody, he has always been very invested in all the people that he has worked with. I'm sure it must be a very tiring position to be in, but he's also very excited about every single thing that he does. That's also a great thing about him that inspires others to get excited about it as well. Going back to your earlier question, he was the one who actually had the idea of starting a company. [*laughs*]

I have two questions: an easy and a toughone. Michael told me about you, that you are a rockstar! Can you see why he thinks so? Then the easier question: If you are a rockstar, what is Michael for you?
So for the first one, yes, that is a difficult

VIZchester United was obviously the VIZ-lab's football team at Texas A&M. We will not report how good they were...

one! I don't know! It's hard for a person to say that for themselves. I do persevere. I've been able to catch up to every environment that I've been in. I absorb and adapt. Even while I was working at Max Planck, even while I was a research engineer, I did get involved… because we were a very small team, we had to do everything. We would need engineers in the scanning room. I wouldn't get scared to get my hands dirty trying to play around with the sensor. I've done engineering work. I've done administration work. I've done financial planning work [*laughs*]… of course software engineering, creating lots and lots of tools for all of the group members and myself.

And you wrote a lot of code?
I wrote a lot of code and paper reviews. I think being able to jump into lots of different things is what helps me be this sort of rockstar… [*laughs*] … as Michael says! That's something I like to do. I like to get involved in everything. There's also the saying: *"Jack of all trades, master of none"*. [*laughs*] I'm also afraid of that! I do like getting involved and finding out how people are doing different things. It's a flaw but also a good thing.

Back to Michael, can we find a name for him?
There's a rockstar and there's a superstar! Michael is an amazing friend and amazing mentor to learn from.

What happened before Max Planck and Texas? You are not from Germany. Where are you from?
I am from Pakistan.

You are the third Pakistani that we have interviewed for this Women in Computer Vision series, after Sophia Bano and Imama Noor. It's an honor to have you. Can you tell us something about growing up in Pakistan as a science oriented and mathematically gifted girl?
When I was in school, in a small town, I was already one of the best math students. Then I went to university, for my Bachelor's, in Lahore. Pakistan has a lot of talent. I am definitely not an anomaly among the people of Pakistan. There are so many people who are very gifted mathematically, in physics. I think it's just a lack of opportunity and being able to explore your talents more. I was lucky that I was able to find a good

The team met this year with Bundeskanzlerin Frau Angela Merkel...

university, to find a good position at the university. Even during my Bachelor's, honestly, I wasn't sure that I would go for computer graphics and computer vision. I had known about computer science, and I started learning a little bit of programming, with Fortran actually.

We have this in common! The only language that I ever studied.
I started only with Fortran, and I thought it was amazing! [*laughs*] Everything was so logical! From there, when I went to do my Bachelor's, there was also C++ and Java, and a whole world opened up!
As a Pakistani, I was definitely one ofthe lucky ones to get into this university. There are not many spots for these kinds of universities. I think there definitely needs to be many more. During my college years in Pakistan, I could see many colleagues who had so much potential, they just never pursued it further.

"It's more about faith in people!"

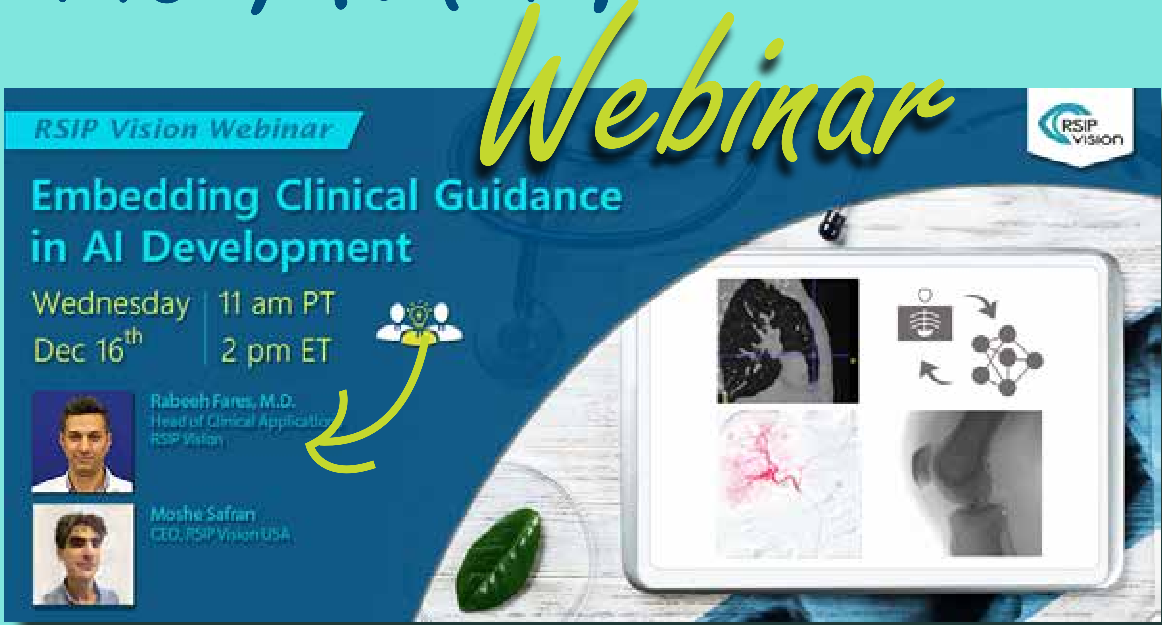
What was it about you that you were able to go abroad and seize opportunities? What was the special fire that you had to do that?
[*laughs*] That's a good question. I don't think that I personally have a special firethat other people don't have. Many have this ambition of doing great things with their skills, but lack the opportunities. I was lucky to be there at the right place, at the right time. I applied to scholarships. From Pakistan to the US, I went on a Fulbright scholarship. That's how I got to go. Without the scholarship, I would not have been able to pay US college prices.

What's your message to mankind?
There is something I can say for women, but it honestly applies to everyone. First thing, don't be afraid to pursue opportunities just because you think you won't make it. That has been one of the reasons why I was able to persevere. Second thing, don't be afraid to ask people for help. Always be willing to break into a group and introduce yourself. This is something that I had to work on myself [*laughs*]. I always assume that I will be able to figure it out. Even before we had Google, I would go to the library to try and figure it out. But people are mostly very happy to join in when they have a similar interest. Based on these things, always aim for the best!

Does faith help you?
Yes! I guess so! Besides faith in general, it's more about faith in people. Sure, there are not very nice people out there. If you look, you will also find people who are willing to help.

Find out 100 more interviews like this with Women in Science!

This Month!


Webinar
RSIP Vision Webinar
Embedding Clinical Guidance in AI Development
Wednesday
Dec 16th
11 am PT
2 pm ET
Rabeeh Fares, M.D.
RSIP Vision
Moshe Safran
CEO, RSIP Vision USA
RSIP Vision


Bay Vision Virtual Meetup
AI-Assisted Surgery Perspectives and Challenges
Tuesday
Dec 15th
10 am PT
1 pm ET
Guest Speaker:
Stefanie Speidel
Professor for Translational Surgical Oncology at the National Center for Tumor Diseases in Dresden.
Hosted By:
Moshe Safran
CEO, RSIP Vision USA
RSIP Vision
NCT
Meetup

COMPUTER VISION EVENTS

IEEE IPAS
Virtual
Dec 9 - 11

ICIGP 2021
Sanya, China
Jan 1 - 3

WACV 2021
All sessions will be virtual
Jan 5 - 9

ICPR 2021
Conference will be Virtual
Jan 10 - 15

Deep Learning 2.0
Virtual Summit
Jan 28 - 29

Ai4 Cybersecurity
Taking Place Digitally
Feb 3 - 4

ICAART 2021
Online Streaming
Feb 4 - 6

ICPRAM 2021
Online Streaming
Feb 4 - 6

VISIGRAPP 2021
Online Streaming
Feb 8 - 6

SPIE Medical Imaging
San Diego, CA
Feb 14 - 18

SUBSCRIBE!

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

FREE SUBSCRIPTION

(click here, its free)

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

Fill the Subscription Form
- it takes less than 1 minute!

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

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