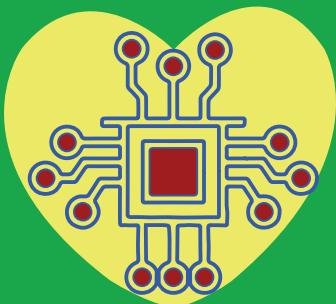


October 2021

Computer Vision News

The Magazine of the Algorithm Community

with the new supplement



Medical Imaging News

page 28



SAFE LANDING AND DROP ZONE ESTIMATION

2 Editorial



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](#)

Copyright: RSIP Vision
All rights reserved
Unauthorized reproduction
is strictly forbidden.

Follow us:



Dear reader,

Over 5 successful years, **Computer Vision News** has been at the forefront of every topic in AI and computer vision, from self-driving vehicles to OCR, precision agriculture to general robotics.

But one field in particular has seen immense progress and enjoyed growing coverage in our magazine: **AI in Medical Imaging**. We haven't failed to notice the stories you consume the most and recognize the huge demand for quality MedTech content from the community.

This month, Computer Vision News is taking a giant leap forward. As a pioneer in medical imaging, it's only natural that RSIP Vision leads the way again in publishing its newest supplement: **Medical Imaging News**.

There is no better time to launch Medical Imaging News than now, as we celebrate the outstanding success of **MICCAI 2021**. Don't miss our **BEST OF MICCAI** stories (page 32) to learn more about the fascinating and innovative work on display at this premier virtual event.

Whether you're interested in general computer vision, medical AI, or both, let us continue to be your online home for **the latest scientific and technological developments** and highlights with a human touch.

Computer Vision News and Medical Vision News will always be published together and accessible from the same link, so please [tell your colleagues and friends and share the link](#) so that they can subscribe for free - the more, the merrier!

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

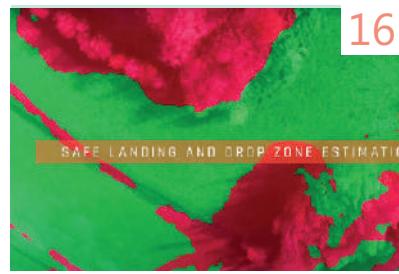
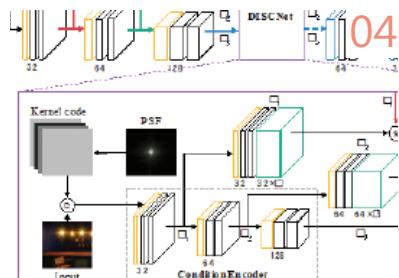
"Thanks, Ralph Anzarouth for continuing to do an excellent job highlighting the activities at MICCAI2021 with your daily newsletter."

Anne Martel
Professor at University of Toronto and
Senior Scientist at Sunnybrook Research Institute





Computer Vision News



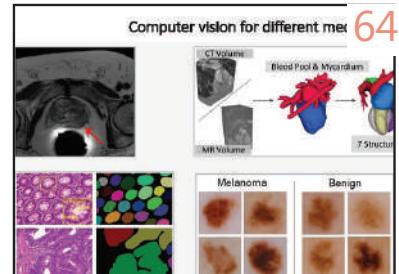
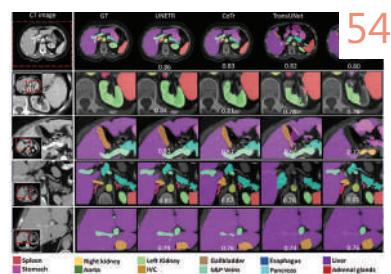
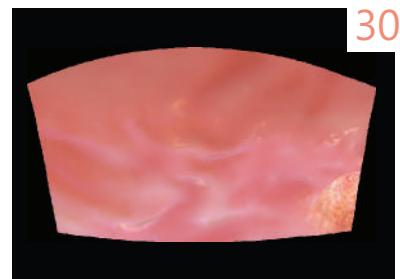
04 Removing Diffraction Image...
Computer Vision Research
by Marica Muffoletto

10 Creating a multi-object tracking...
Coding Workshop
by Ioannis Valasakis

16 Spleenlab
AI Systems for Autonomous Mobility

22 RePAIR Pompeii
AI for Archeology

Medical Imaging News



30 Bladder Panorama Generator
AI for Urology

32 Best of MICCAI
The best features chosen by us

54 Transformers in Medical Imaging
by NVIDIA and MONAI

64 Computer Vision for Automated...
ICCV Workshop Preview

4 Computer Vision Research

Removing Diffraction Image Artifacts in Under-Display Camera via Dynamic Skip Connection Network

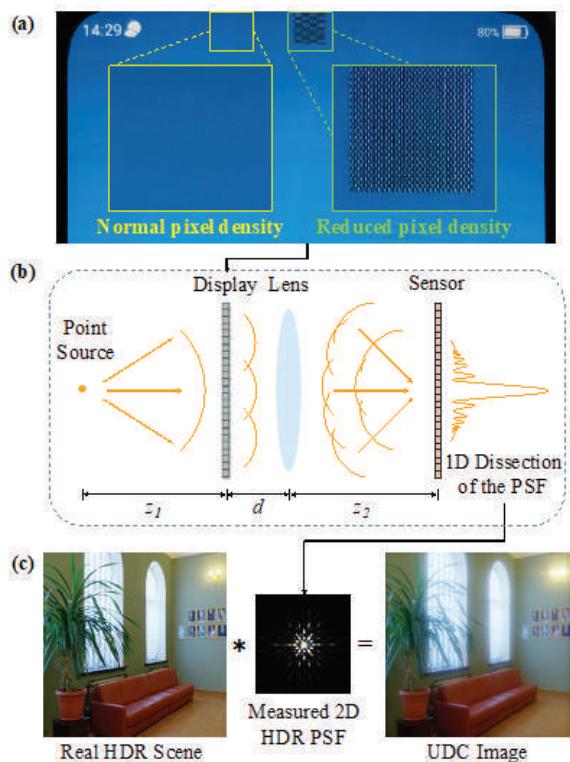


by Marica Muffoletto

Dear readers, welcome back to a new issue of Computer Vision News full of great research!

This month we are reviewing a paper from CVPR 2021, **Removing Diffraction Image Artifacts in Under-Display Camera via Dynamic Skip Connection Network**, written by Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, Jinwei Gu. We are indebted to the authors for allowing us to use their images to illustrate this review. Their paper can be found [here](#).

If you like to play around with the quality of cameras in your tech devices, this might become your favourite CVPR paper of the year. The subject of this research is indeed how to remove image artifacts in the newly defined imaging system called Under-Display Camera (UDC), which is employed in some smartphones, for videoconferencing with UDC TV, laptops, or tablets. UDC introduces a new class of complex image degradation problems (strong flare, haze, blur, and noise), which still need to be satisfactorily dealt with by the computer vision community. A typical UDC system is constituted by a camera module placed underneath and closely attached to the semi-transparent Organic Light-Emitting Diode (OLED) display.



The picture on the left illustrates how this setup affects the light propagation: although the display looks partially transparent, the gaps between the display pixels (a), where the light can pass through, are usually in the micrometer scale and hence the incoming light gets substantially diffracted. This effect is modelled through a Point Spread Function that can be measured to convert a Real HDR Scene into a UDC image.

This paper aims to mathematically describe this phenomenon and investigate the artefacts caused by the diffraction effects in such a system.

The first contribution consists in the formulation of an image formation model for UDC systems which considers dynamic range and saturation and could simulate more complex and realistic degradation compared to the State-of-the-Art.

The degradation model is described as:

$$\hat{y} = \phi[C(x * k + n)]$$

where x represents the real scene irradiance that has a high dynamic range (HDR), k is the known convolution kernel (PSF), ϕ denotes the 2D convolution operator, and n models the camera noise. $C(\cdot)$ is a clipping operation with a set threshold and $\phi(\cdot)$ is a non-linear tone mapping. These two elements add a saturation effect derived from the limited dynamic range of digital sensors and make the model closer to the human perception of the scene.

The second element in the paper consists in defining the PSF. This can be simulated but it was found that the real-measured PSF slightly differs in colours and contrast due to model approximations and manufacturing imperfections.

Hence, the authors measure the real-word PSF by placing a white point light source 1-meter away from the OLED display. The PSF is used as part of a model-based data synthesis pipeline to generate realistic degraded images.

To do this, the objects considered are real scenes with high dynamic range. This is essential because 1) the spike-shaped sidelobes (typical of the PSF) can be amplified to be visible (flares) in the degraded image, and 2) due to the high dynamic range of the input scene, the digital sensor (usually 10-bit) will inevitably get saturated in real applications, resulting in an information loss.

Hence, images captured by UDC systems in real scenes will exhibit structured flares near strong light sources. The previous imaging system, however, cannot model this degradation, because it captures images displayed on an LCD monitor, which commonly has limited dynamic range.

This is shown below, where is demonstrated that the real HDR scene captured by the

6 Computer Vision Research

UDC device (b) shows flare effects near strong light sources, while for the monitor-generated LDR scene (c) with a limited dynamic range, these are no longer visible.



Moreover, the authors introduce a second experiment to show the importance of using real HDR scenes in the image formation model to correctly model the real degradation of a typical UDC system. The image below illustrates a scene clipped from LDR on the left, and HDR on the right, where the flare artifacts caused by diffraction effects are only visible in the latter.



As last contribution, the paper presents a Dynamic Skip Connection Network (DISCNet) that incorporates the domain knowledge of the UDC image formation model to get rid of the diffraction artefacts in the UDC images.

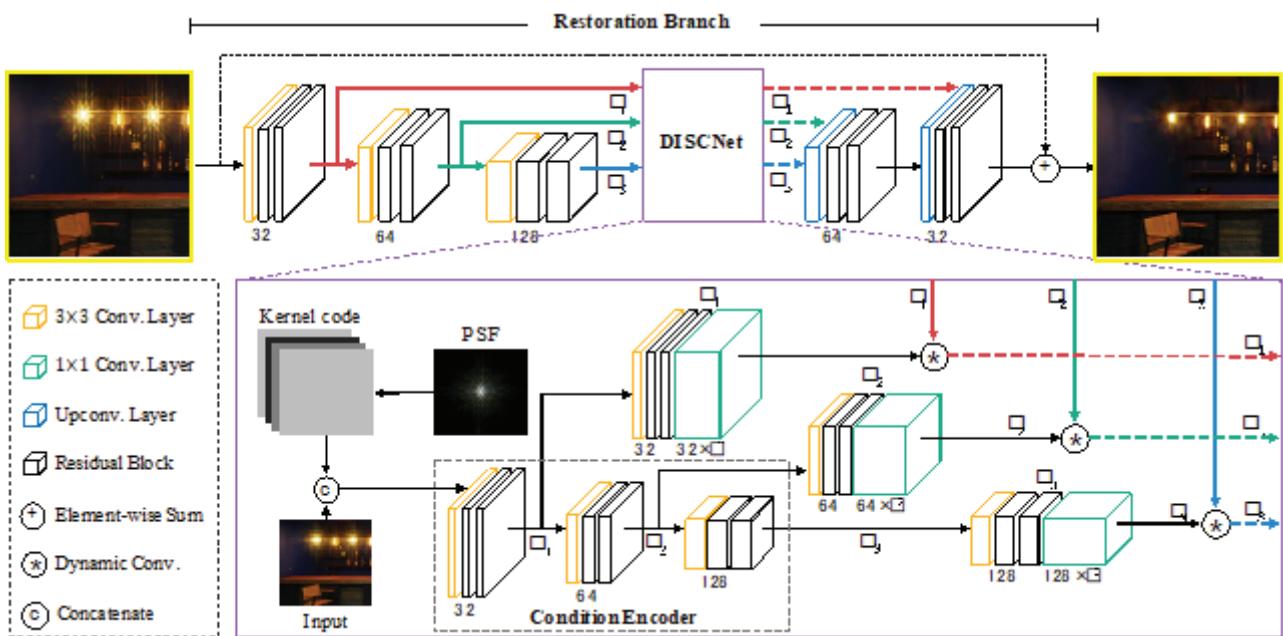
This is found within a main restoration branch which builds upon an encoder-decoder architecture with skip connections to restore the degraded images. The

Removing Diffraction Image Artifacts in ... 7

extracted features from the encoder are fed into **DISCNet** which transforms them into R1, R2, R3. These are then reconstructed back to the final tone-mapped sharp images.

The network is fed with condition maps of size $H \times W \times (b+C)$, where b stands for the kernel code (a b -dimensional vector of the PDF dimensionally reduced by Principal Component Analysis) and H , W , C represent size and channels of the degraded images respectively.

Given the condition maps as input, the **condition encoder** extracts scale-specific feature maps H_1, H_2, H_3 using 3 blocks like the encoder of the restoration branch. This manages to recover saturated information from nearby low-light regions in the degraded images with spatial variability. Then, the extracted features at different scales are fed into their corresponding **filter generators**, where each comprises a 3×3 convolution layer, two residual blocks, and a 1×1 convolution layer to expand feature dimension. The predicted filters are output and passed into a **dynamic convolution** element which finally refines the features and cast them into the main restoration branch.



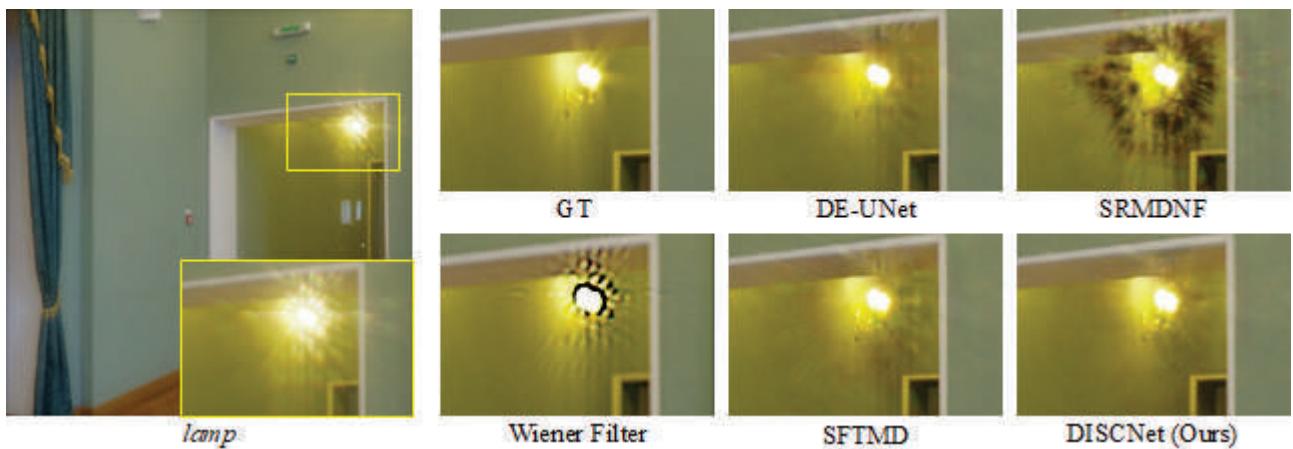
The network is trained on a combination of:

- A synthetic dataset, generated from HDR images with large dynamic ranges, from which a degraded image has been simulated using the degradation model defined above and the calibrated PSF.
 - A real dataset, made of three images of different exposures, taken with a ZTE Axon 20 phone and combined in a unique HDR image.

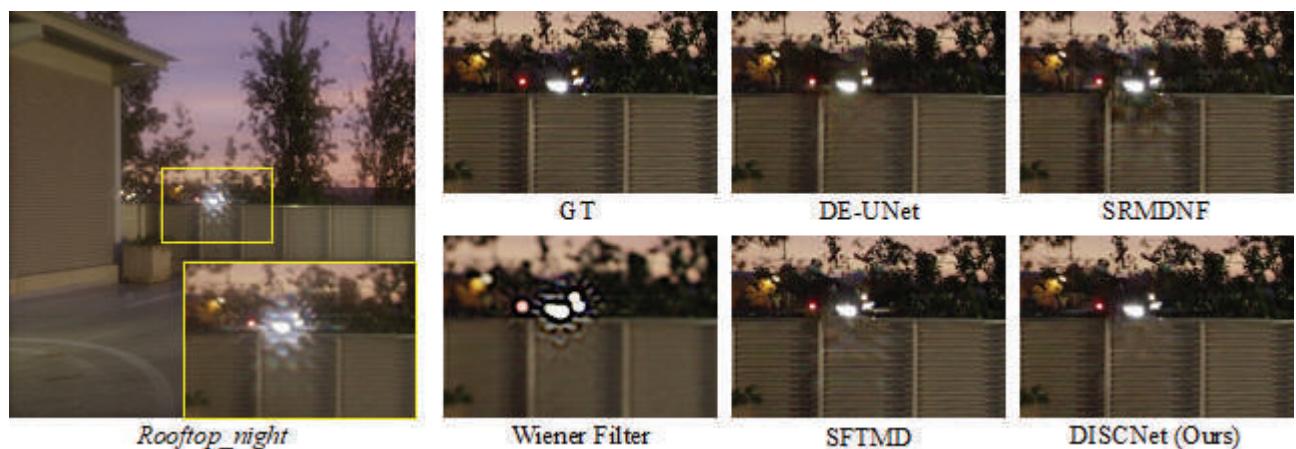
8 Computer Vision Research

Experimental results show that the method is effective for removing diffraction image artifacts in UDC systems.

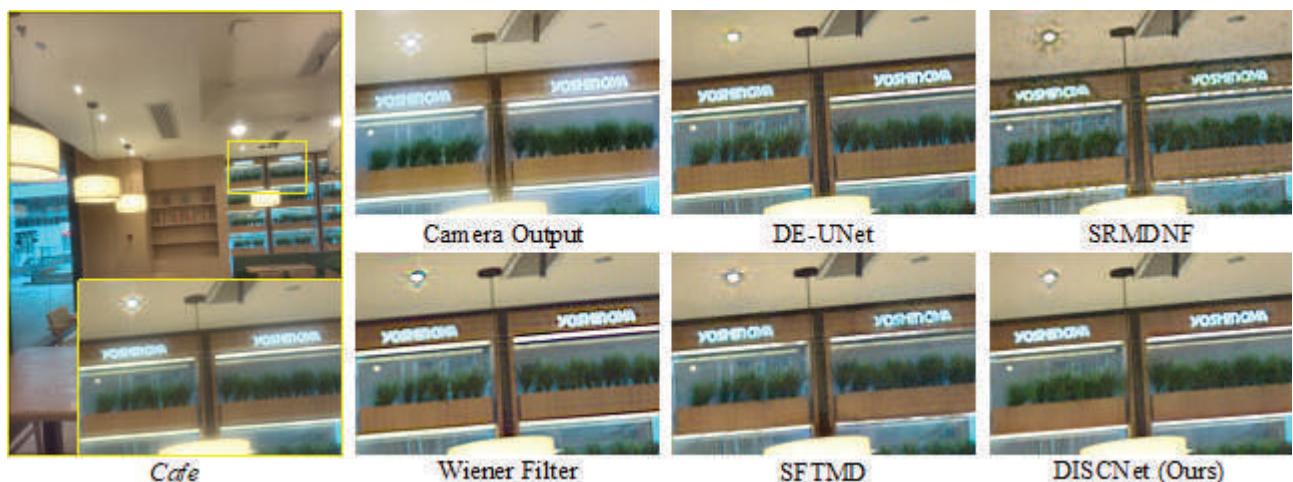
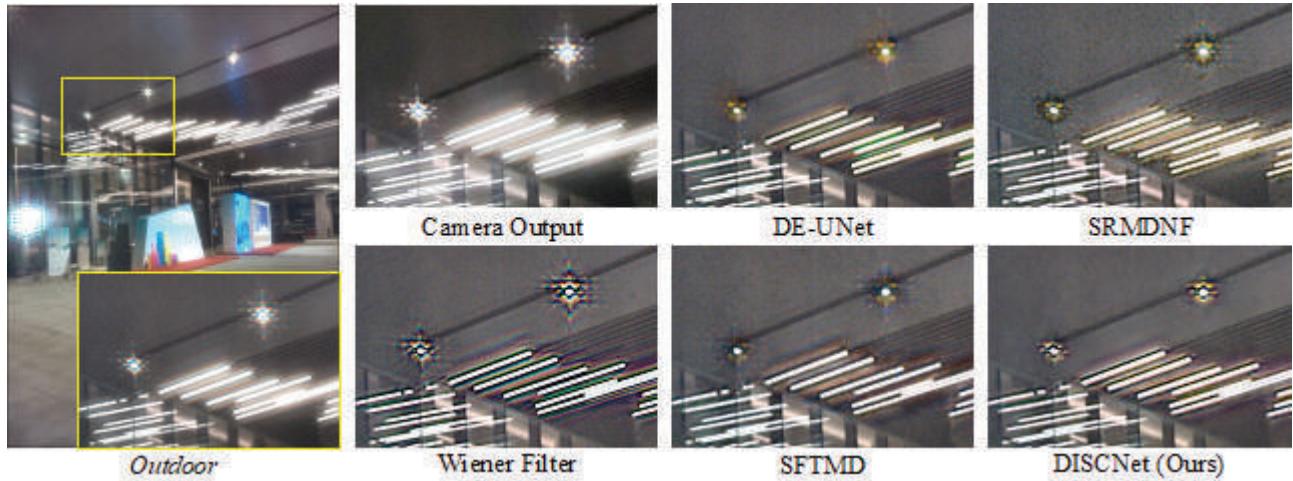
Figures below show how the proposed DISCNet successfully suppresses flare and haze effects around highlights and removes most artifacts in nearby saturated images. The results of the proposed method are compared with 4 State-of-the-Art methods: the **Wiener Filter**, a classical deconvolution algorithm for linear convolution formation which achieves the lowest image quality compared to the other deep learning methods; the **SRMDNF** method which uses a super-resolution network and cannot adapt to degraded regions caused by highlight sources; the **SFTMD** network which iteratively connects the kernel code of degradations and, in this application, also leverages the kernel information to solve the non-blind problem (comparable performance but highest computational cost) and finally the **DE-UNet** method, a Double-Encoder UNet which doesn't explicitly use the kernel information (blind model).



Similarly, comparisons with representative methods are shown for the real images below. The network proposed by the authors manage to remove diffraction image effects, while leaving least artifacts introduced by the camera. Since the ground-truth images are inaccessible, another comparison is made with the camera output of a ZTE phone.



Removing Diffraction 9 Image Artifacts in ...



PyTorch codes and data from this paper are available on [github](#). Best of luck with making your images look great ☺
See you all next month!

10 Coding Workshop

Coding Workshop: Creating a multi-object tracking model using a pre-trained RNN



IOANNIS VALASAKIS, KING'S COLLEGE LONDON



@WIZOFE

Nice to meet you everyone this month! I hope that you enjoyed the last technical coding and you didn't have a great trouble following it. I am always eager to receive more feedback from you 😊 Let me know what computer vision tools are your favorite or if you have trouble implementing a deep learning method on visual computing and I will be happy to cover that in an article!

This month's article is about using a pre-trained network (you can find that implementation on the GitHub MOT Challenge).

MOT Challenge

There is remarkable progress on object detection and association in recent years which are the core components for multi-object tracking. The main focus though has been on improving singular networks. In this example ([see the original paper](#) to read more about the topic), the proposed architecture (pre-trained network) combines two tasks in a single network to improve the inference speed.

Older research has shown degraded results in this combined network, mainly because the association branch is not appropriately learned. Here, after discovering the reasons behind the failure, a simple baseline was presented to address the problems. It was shown that it remarkably outperforms the state-of-the-arts on the MOT challenge datasets at 30 FPS. Hopefully, this baseline could inspire and help evaluate new ideas in this field. Now let's see the implementation of the pre-trained network!

Creating the R-CNN

With the following code you can implement a variation of the pre-trained model and use COCO to download the trained weights, import the R-CNN mask and set a specific image directory where the images can be run for the training.

Creating a multi-object tracking model...

11

```
import os
import sys
import random
import math
import numpy as np
import skimage.io
import matplotlib.pyplot as plt

# Root directory of the project
ROOT_DIR = os.path.abspath("../")

# Import Mask RCNN
sys.path.append(ROOT_DIR)
from mrcnn import utils
from mrcnn import model as mrcnn_model
# Import COCO config
sys.path.append(os.path.join(ROOT_DIR, "samples/coco/"))

import coco
%matplotlib inline

# Directory to save logs and trained model
MODEL_DIR = os.path.join(ROOT_DIR, "logs")

# Local path to trained weights
COCO_MODEL_PATH = os.path.join(ROOT_DIR, "mask_rcnn_coco.h5")
# Download COCO trained weights from Releases if needed
if not os.path.exists(COCO_MODEL_PATH):
    utils.download_trained_weights(COCO_MODEL_PATH)

# Directory of images to run detection on
IMAGE_DIR = os.path.join(ROOT_DIR, "images")
```

Using TensorFlow backend.

Configuration files

We'll be using a model trained on the MS-COCO dataset. The configurations of this model are in the `CocoConfig` class in `coco.py`. For inferencing, modify the configurations a bit to fit the task. To do so, sub-class the `CocoConfig` class and override the attributes you need to change.

12 Coding Workshop

```
class                                     InferenceConfig(coco.CocoConfig):
    # Set batch size to 1 since we'll be running inference on
    # one image at a time. Batch size = GPU_COUNT * IMAGES_PER_GPU
    GPU_COUNT                      = 1
    IMAGES_PER_GPU                  = 1

config = InferenceConfig()
config.display()

Configurations:
BACKBONE_SHAPES           [[256 256]
[128 128]
[ 64  64]
[ 32  32]
[ 16  16]]
BACKBONE_STRIDES          [4, 8, 16, 32, 64]
BATCH_SIZE                 1
BBOX_STD_DEV               [ 0.1  0.1  0.2  0.2]
DETECTION_MAX_INSTANCES   100
DETECTION_MIN_CONFIDENCE  0.5
DETECTION_NMS_THRESHOLD   0.3
GPU_COUNT                  1
IMAGES_PER_GPU              1
IMAGE_MAX_DIM              1024
IMAGE_MIN_DIM              800
IMAGE_PADDING               True
IMAGE_SHAPE                [1024 1024      3]
LEARNING_MOMENTUM          0.9
LEARNING_RATE               0.002
MASK_POOL_SIZE             14
MASK_SHAPE                 [28, 28]
MAX_GT_INSTANCES           100
MEAN_PIXEL                 [ 123.7  116.8  103.9]
MINI_MASK_SHAPE            (56, 56)
NAME                        coco
NUM_CLASSES                81
POOL_SIZE                  7
POST_NMS_ROIS_INFERENCE   1000
POST_NMS_ROIS_TRAINING     2000
ROI_POSITIVE_RATIO         0.33
RPN_ANCHOR_RATIOS          [0.5, 1, 2]
RPN_ANCHOR_SCALES          (32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE          2
RPN_BBOX_STD_DEV            [ 0.1  0.1  0.2  0.2]
RPN_TRAIN_ANCHORS_PER_IMAGE 256
STEPS_PER_EPOCH             1000
TRAIN_ROIS_PER_IMAGE        128
USE_MINI_MASK               True
USE_RPN_ROIS                True
VALIDATION_STEPS            50
WEIGHT_DECAY                0.0001
```

Create Model and Load Trained Weights

```
# Create model object in inference mode.  
model = modellib.MaskRCNN(mode="inference", model_dir=MODEL_DIR, config=config)  
  
# Load weights trained on MS-COCO  
model.load_weights(COCO_MODEL_PATH, by_name=True)
```

Class Names

The model classifies objects and returns class IDs, which are integer value that identify each class. Some datasets assign integer values to their classes and some don't. For example, in the MS-COCO dataset, the 'person' class is 1 and 'teddy bear' is 88. The IDs are often sequential, but not always. The COCO dataset, for example, has classes associated with class IDs 70 and 72, but not 71.

To improve consistency, and to support training on data from multiple sources at the same time, our `Dataset` class assigns its own sequential integer IDs to each class. For example, if you load the COCO dataset using our `Dataset` class, the 'person' class would get class ID = 1 (just like COCO) and the 'teddy bear' class is 78 (different from COCO). Keep that in mind when mapping class IDs to class names.

To get the list of class names, you'd load the dataset and then use the `class_names` property like this.

```
# Load coco dataset  
dataset = coco.CocoDataset()  
dataset.load_coco(COCO_DIR,  
dataset.prepare()  
  
# Print class names  
print(dataset.class_names)
```

You won't need to download the COCO dataset just to run this demo! We are including the list of class names below. The index of the class name in the list represents its ID (first class is 0, second is 1, third is 2, ...etc.)

14 Coding Workshop

```
#           COCO          Class          names
# Index of the class in the list is its ID. For example, to get ID of
# the teddy bear class, use: class_names.index('teddy bear')
class_names = ['BG', 'person', 'bicycle', 'car', 'motorcycle', 'airplane',
               'bus', 'train', 'truck', 'boat', 'traffic light',
               'fire hydrant', 'stop sign', 'parking meter', 'bench', 'bird',
               'cat', 'dog', 'horse', 'sheep', 'cow', 'elephant', 'bear',
               'zebra', 'giraffe', 'backpack', 'umbrella', 'handbag', 'tie',
               'suitcase', 'frisbee', 'skis', 'snowboard', 'sports ball',
               'kite', 'baseball bat', 'baseball glove', 'skateboard',
               'surfboard', 'tennis racket', 'bottle', 'wine glass', 'cup',
               'fork', 'knife', 'spoon', 'bowl', 'banana', 'apple',
               'sandwich', 'orange', 'broccoli', 'carrot', 'hot dog', 'pizza',
               ,
               'donut', 'cake', 'chair', 'couch', 'potted plant', 'bed',
               'dining table', 'toilet', 'tv', 'laptop', 'mouse', 'remote',
               'keyboard', 'cell phone', 'microwave', 'oven', 'toaster',
               'sink', 'refrigerator', 'book', 'clock', 'vase', 'scissors',
               'teddy bear', 'hair drier', 'toothbrush']
```

Run Object Detection

You can run the following code to generate and detect the run object. To visualize the results, we are using a `display_instances` function, inputting the images with their respective region of interest (ROI), masks, the class id as well as the class name and the scores.

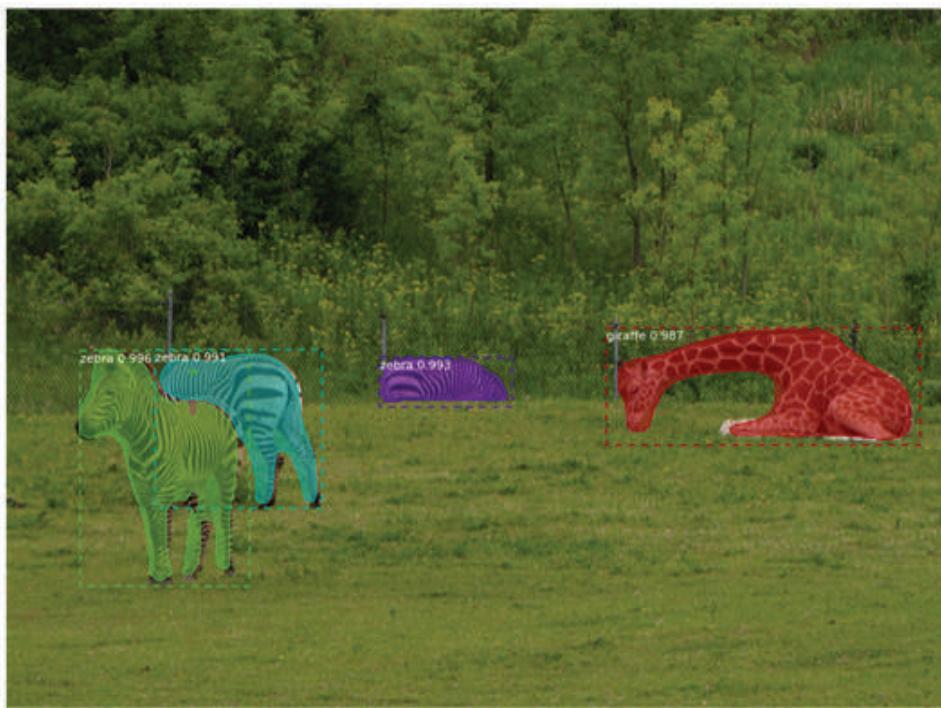
```
# Load a random image from the images folder
file_names = next(os.walk(IMAGE_DIR))[2]
image = skimage.io.imread(os.path.join(IMAGE_DIR, random.choice(file_names)))

# Run model.detect([image])
results = model.detect([image], verbose=1)

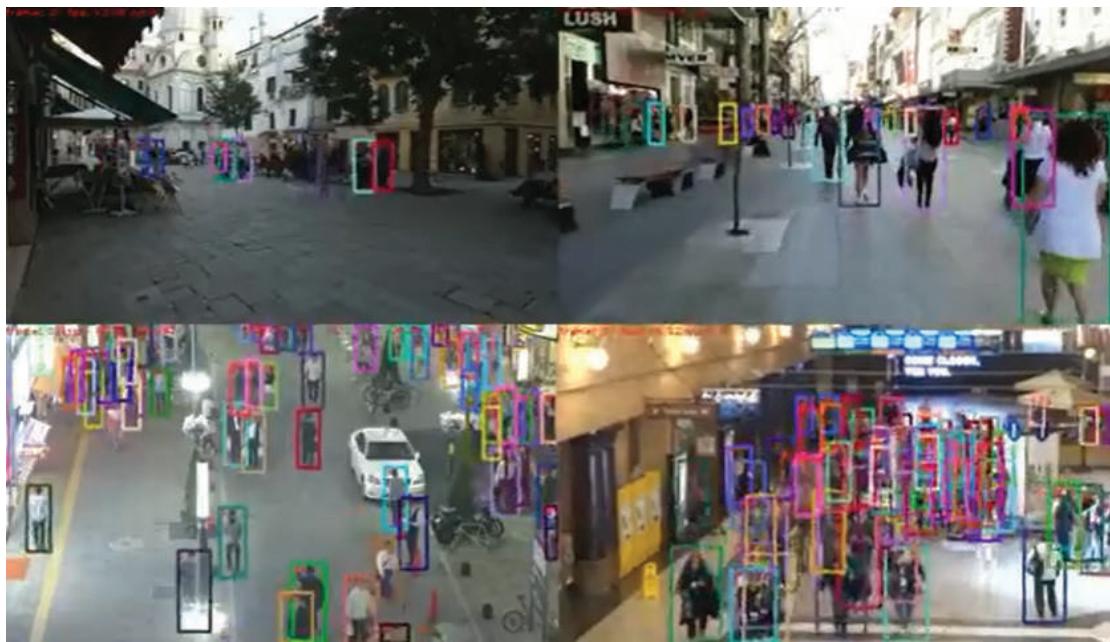
# Visualize results
r = results[0]
visualize.display_instances(image, r['rois'], r['masks'], r['class_ids'],
                            class_names, r['scores'])

Processing images
image      1
shape: (476, 640, 3)    min: 0.00000 max: 2
55.00000
molded_images 20.30000
shape: (1, 1024, 1024, 3) min: -123.70000 max: 1
image_metas 024.00000
shape: (1, 89)           min: 0.00000 max: 1
```

Creating a multi-object tracking model... 15



Here are some videos, visualizing the example network:



Wrapping up!

I hope that I inspired you to look more into the world of pre-trained networks and discover how easy it is to implement a new network, using a pre-trained model! As always, please let me know if you have any questions, or suggestions for the article! It would be great to hear more of you and what tools you would like to be presented; feel free to reach out to me in any of the social media 😊

16 AI Systems for Autonomous Mobility

Spleenlab.ai is a pure AI software company, specializing in safe and intelligent AI systems for autonomous mobility, including unmanned aerial vehicles (UAVs) and autonomous driving. Their vision is to be an AI software supplier for air taxis, which are a composition of both UAVs and autonomous vehicles. Stefan Milz is Head of R&D, Managing Director and Founder of Spleenlab GmbH. He received his PhD in Physics from the Technical University of Munich and has a strong history in professional software development and automotive. Stefan tells us more about Spleenlab.

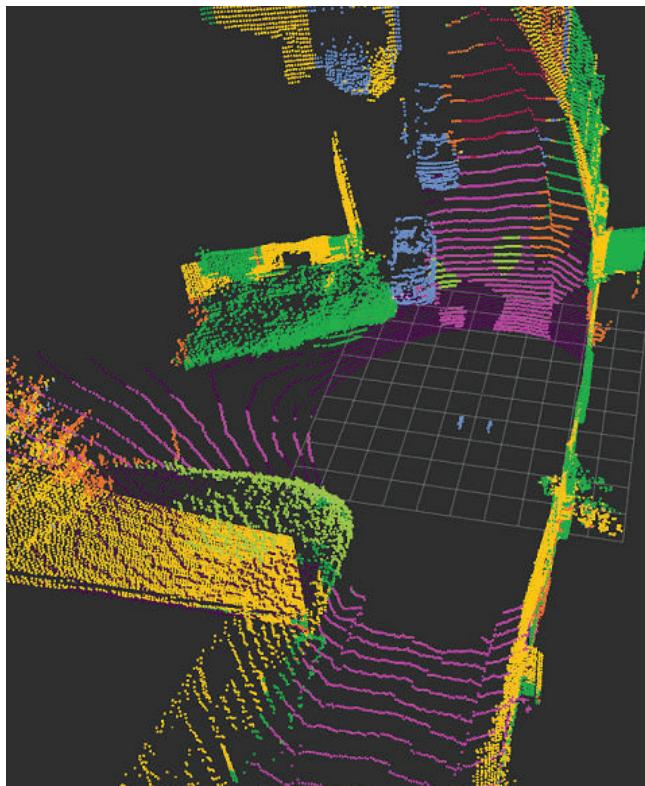
To date, **AI research** has tended to focus on achieving higher accuracy, better models, and more trustworthy

or explainable AI. But for **Spleenlab**, working in the field of autonomous mobility, **safety is king**.

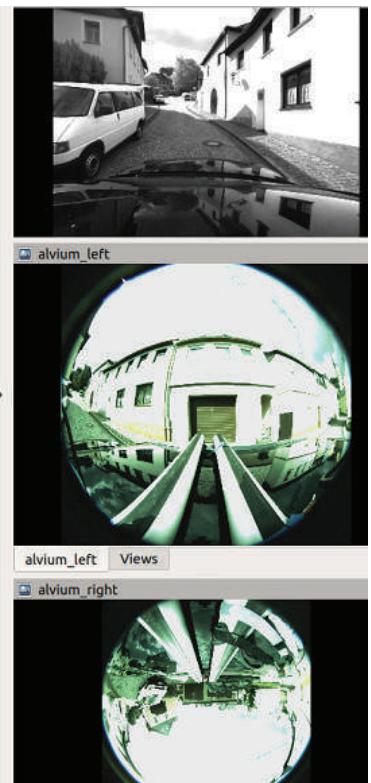
The domains of **UAVs, autonomous driving, and airborne systems** are all regulated by entities such as EASA in Europe and the FAA in the United States. There are important industrial standards such as ISO 26262 in the automotive domain and DO-178 for software in airborne systems, as well as new standards emerging in the UAV domain. Systems are expected to have a fully deterministic behaviour.

"We believe that people need extreme safety," Stefan tells us.

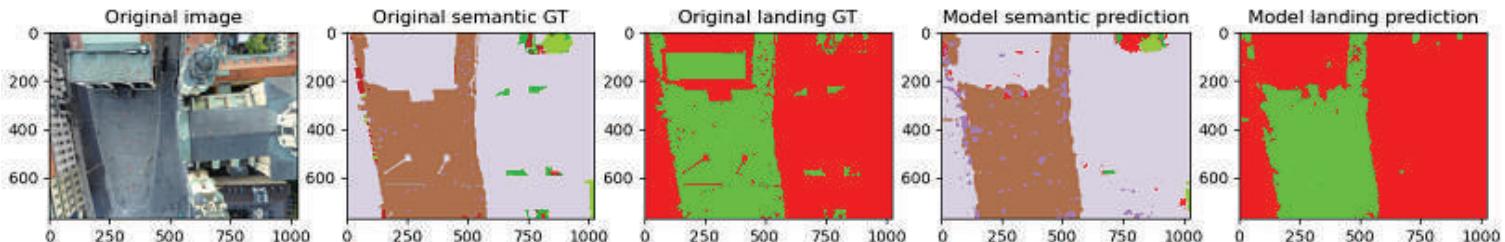
"It's important to foresee what state your system can achieve in an automated scene. We want to have a



Automotive Sensor Fusion



*large if-else tree and foresee every potential part in that tree and assess each state at the end of the tree with probabilities. We say our software could run through each part of the tree. We call that **determinism**. If we have no AI, so-called deep learning, we have a highly statistical system. We can validate our AI, as many people did, but at the end of the day, we have a degree of uncertainty in our system. It's hard to regulate those systems by a standard that requires determinism because you can't foresee the last percentage of your system."*



Landing prediction

Uncertainty is bad. People want to know they are safe in their cars and that drones will not hit other aircraft. But why is there uncertainty in the system? The models are trained on data and validated on data, with millions of parameters in the system that are deterministic, but some methods at inference time, like dropout and pruning make the system statistical. A statistical system always has uncertainty.

"Take autonomous driving, for example," Stefan says.

*"You have a dataset, you train on the dataset, and then you have a validation dataset that you collect in Germany, and it performs at **99 percent accuracy**. There is still a risk of uncertainty because if you put your validation set to a different domain that's collected in France, with a demographic and geographic domain shift, you may get a 98 per cent accuracy, and you do not really know why. There is no direct if-else connection in a deep learning model."*

People are already working on

transparent AI, but Stefan says a new standard that explains its deepest behaviors is still many years away, while models are being sold and deployed in the real world now. That is why Spleenlab is focused first and foremost on safety by design.

Take the example of an automated drone on a package delivery **which encounters an emergency**. Its battery has dropped to a critical level. The system has a rest range of 100 meters. There is no known emergency landing point in the area. What does it do?

18 AI Systems for Autonomous Mobility

"A computer vision model can predict some good landing spots," Stefan explains.

*"The best method to choose is an AI model. Semantic segmentation can semantically parse the scene and find the best spots. We know the model is 99 per cent accurate, but we still don't want to hit anyone on the head, so we take those landing spots and put a safety goal around them. We use a second sensor from either **the thermal imager or a LiDAR sensor** and we look at that predicted spot and validate if it is really safe. With the thermal imager we can validate it in the sense of, are there people, animals, or vehicles there? Those are **the most critical things we want to avoid**. With the LiDAR sensor, we can estimate if the scene is geometrically flat and if there is any dynamic object in sight. With two deterministic passes to validate it, we can then say, okay, this spot is*

safe."

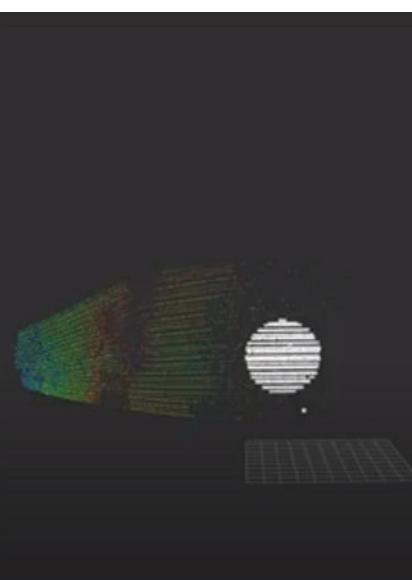
This idea is not fully new in the automated domain, but it has not often been combined with AI. The model must be validated with a large amount of **labeled data**, which is expensive. Deploying a model in a different domain, a different country for example, requires **domain adaptation**, which significantly lowers the cost.

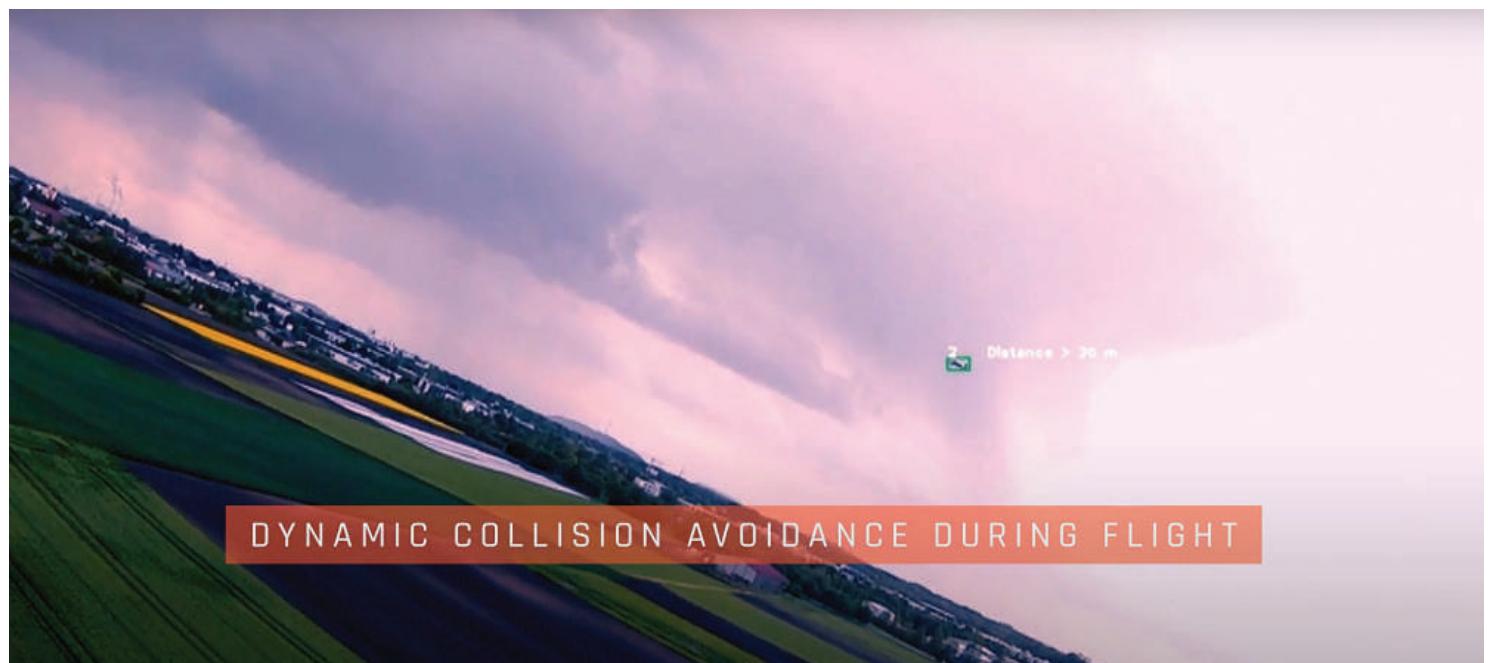
"We have a simulation engine where all the flights and the labels can be generated," Stefan tells us.

*"We also have a lot of labeled data from companies and are working on how we can transfer this data to different domains. We call it **cross-domain AI**. That is our core product. We have been working on a pure software product called **Visionairy perception software**, which has several features for UAVs and aviation."*



Landing Sensor Fusion





DYNAMIC COLLISION AVOIDANCE DURING FLIGHT

Object Detection with distance estimation

Other generic products the team are working on for the UAV market include a **detect and avoid function for detecting manned aircraft**, which heavily needs a vision component, so it is not solved yet. They also have **emergency landing functionality and tracking functions** in the pipeline, and are working on future architecture together with big air taxi manufacturers like **Volocopter** and **DLR**.

There have been some high-profile concerns about the use of AI in everyday life, including fears that the technology could be exploited by bad actors. Does Stefan think this could be an issue for Spleenlab?

"At the moment, we are in a very early stage. We don't see any potential issues

yet, but it's an important question that we have to keep asking ourselves."

Stefan tells us the most difficult part of this work is the **airborne certification** and the proof of safety in terms of flight hours or driving hours. Even with safety by design or a deterministic approach, that is still necessary.

Many companies claim to have the best drone or UAV on the market, and that they will solve problems like package delivery in the near future, but there is one very important thing that is not solved yet.

"It's the positioning system," Stefan points out.

*"You need a **GPS redundant positioning system**, which is safe, but if you want*

20 AI Systems for Autonomous Mobility

*simply to fly near a population then you have to validate your system with a single point of failure, which has the probability of one divided by 10,000. This is the number of flight hours you have to do, and you have to do a strong validation of your system. This is a big opportunity for us because we know there are many companies who want to do that, but it's not solved yet. I believe there is a long way to go before we see package delivery outside of a test field, so we want to solve an easier approach first – **pipeline inspection where no population is nearby**. You still have to show that your system is safe if you want to fly far with the pilot out of the loop. That is called **BVLOS – beyond visual line of sight**. This is not yet solved for certification for most of the use cases in North America and Europe.”*

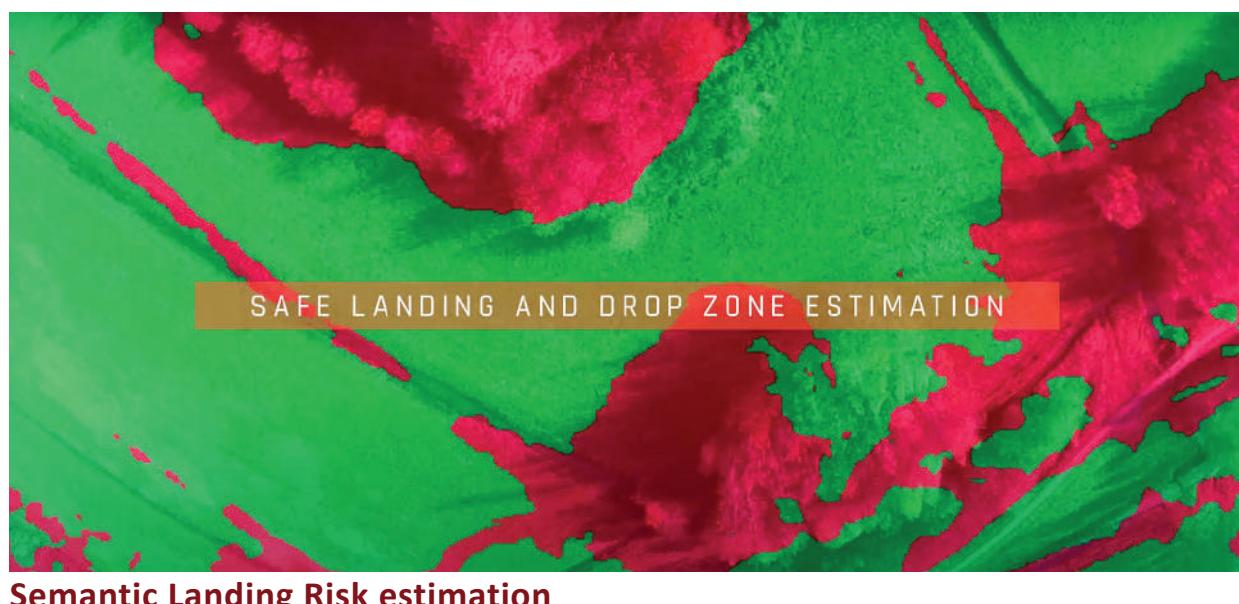
Spleenlab have some very exciting plans for next year and beyond. As a software company, they collaborate

with manufacturers to bring AI to their products, and at the beginning of next year will be launching their simple follow me functions up in the air with drone manufacturer **Quantum-Systems**.

By the end of next year, Stefan says he hopes to see Spleenlab's detect and avoid system in drones. They are also working on **automatic inspection of cell towers**, with the AI looking for the cell tower, flying a drone around it while collecting inspection data, and then bringing it back. This will save money for customers who want to automate the inspection process.

Spleenlab are currently 15 people, and they are hiring.

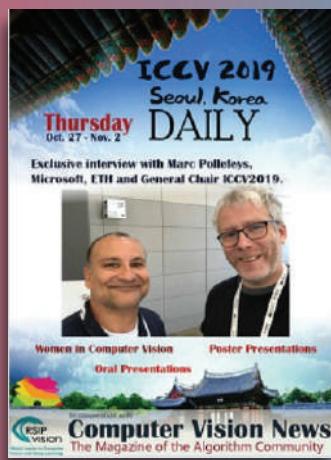
“We are looking for computer vision engineers, deep learning engineers, and PhD applicants with a focus on SLAM perception and sensor processing. Come join us!”



Feel at ICCV as if you were at ICCV

2021 **ICCV** OCTOBER 11-17
VIRTUAL

Get your ICCV Daily news
in your mailbox!



Click here to subscribe
for free to ICCV Daily.



Reconstructing the Past: Artificial Intelligence and Robotics Meet Cultural Heritage (RePAIR)

Marcello Pelillo is Professor of Computer Science and Artificial Intelligence at Ca' Foscari University of Venice. Readers may also remember him as General Chair of ICCV 2017 and as Chief Editor of *Frontiers in Computer Vision*. He is leading an EU-funded project called RePAIR, which is using AI and computer vision to facilitate the reconstruction of some ancient artwork at the archaeological site of Pompeii. Marcello tells us more about this exciting mission.

Around 2,000 years ago, a volcano called **Vesuvius** put an end to life in **Pompeii** (near Naples, Italy), but many artifacts still remain from the time when the Roman city flourished. The **Reconstructing the Past: Artificial Intelligence and Robotics Meet Cultural Heritage** project is working with the archaeological site to physically reconstruct large-scale broken frescoes. The project's acronym is RePAIR and that is exactly what **Marcello Pelillo** and his team are aiming to do.

"Reconstructing a fresco, from a conceptual point of view, is like solving a jigsaw puzzle," he tells us.

"But in reality, it's so much more difficult! These are large-scale frescoes

containing thousands of pieces. Some are very small, but some are pretty big. There's no puzzle box with a nice clear photo of what it will look like when we finish the job. In most cases, it's more of a blind search. Also, it's not like you have all the pieces and it's just a matter of assembling them in the right order. Pieces typically don't match perfectly, and many are missing. It's a really tricky problem."



To solve this, the team use information from the **3D geometry of the pieces**, which are basically stones with one flat side that is colored and decorated. One of the first tasks of the project, which started earlier last month, is to scan all the pieces to get 3D images and create a database containing information about their **shape, 3D geometry, decoration, and appearance**.



One of the team's partners, **Ohad Ben-Shahar from the Ben-Gurion University of the Negev**, has been working on the problem of solving jigsaw puzzles for more than 10 years and is an expert in the field. He and Marcello actually conceived the idea for this project together a while ago. He has proposed **state-of-the-start algorithms and in the project they will use deep learning and related methods** to try to learn the compatibility of two pieces and then use that to guess the position of all the pieces.

However, in the field of computer vision there is not usually the same level of difficulty that is associated with real frescoes, because tiles are often assumed to be square and there are not always missing pieces.

"Given the difficulty of the problem, we don't think that just using a purely data-driven or deep learning-based approach will allow us to completely solve the puzzle," Marcello reveals.

"We're imagining a system in which we incorporate expert knowledge and combine it with machine learning to improve that compatibility function. We're talking about frescoes that were built 2,000 years ago. We have art historians and archaeologists, for example, who will give us precious information concerning the style of painting in those times."

The team will build a system that harnesses the power of deep learning methodology while incorporating expertise in a kind of loop. It will try to propose a solution to the expert who will check it and say, *'I don't think these pieces go together, you need to propose another guess,'* and so on.



24 AI for Archeology

Marcello and his team are not the first to attempt to reconstruct frescoes, but the key differences between this and past approaches are the sheer size of the frescoes, and the fact that ultimately, they are going to physically reconstruct them.

"We will build a robot and use soft robotics," Marcello explains.

"Archaeologists were initially terrified by the idea of a robot holding these pieces because if they break one, it risks losing something very precious. So, once the puzzle has been solved, the information is given as input to the robotic platform and

the robot uses soft-hand technologies to take and manipulate the pieces very delicately. You can control the pressure and make sure that it is just right for the piece. We're going to use the experts in a kind of interactive way, to tell us whether the solution proposed by the robot is plausible or not."

Finally, the robot will put the pieces together, but just next to each other. Expert restorers at Pompeii will take charge from there. It is a very delicate process to put the pieces physically back together using techniques that cannot be incorporated into a simple robotic system.





Marcello and team on site

The team will soon be launching a dedicated website that describes the work. It is a Horizon 2020 project, under the Future and Emerging Technologies (FET) Open program, which is an extremely selective call that gives very few proposals the green light. Marcello and the University of Venice are the project coordinators, with other partners including the Ben-Gurion University of the Negev in Israel, the Italian Institute of Technology in Italy, the University of Bonn in Germany, Instituto Superior Técnico in Portugal, and the Archaeological Park of Pompeii, which is one of the biggest archaeological parks in the world.

"We are introducing a revolutionary technology in archaeology," Marcello says proudly.

"My archaeologist friends tell me that if we succeed, as we hope, then this will be a huge breakthrough in their field. When you have an object that has broken into thousands or even tens of thousands of pieces, it's just hopeless to think that any human team can solve it. Actually, in Pompeii they did try, but had to give up in the end."

Marcello hopes the technology they propose will be able to be used by other museums with broken frescoes, as well as exported to other domains.

"There are other problems, such as reconstructing papyri, vases, or other kinds of broken artifacts," he adds.

"We hope our technology will turn out to be useful when the scale of the problem is unmanageable by humans!"

26 Hidden Stories of the Heart

To all readers of the magazine who live in London, this is a unique call if you are looking for interesting plans for the second weekend of October! Come join research students Marica ([@maricaS8](#), King's College London), Sophie ([www.richtersophie.com](#), Royal College of Arts) and Elizabeth ([@elizabetho157](#), Royal College of Arts), at the Science Museum on the 9th-10th of October to explore the new installation Hidden Stories of the Heart.

Hidden Stories of the Heart invites you to connect with the complexity of our hearts through creativity. Observe handmade papier-maché models, investigate their uniqueness and discover how medical imaging is advancing our understanding of how traumatic life events can impact the shape and function of our hearts.

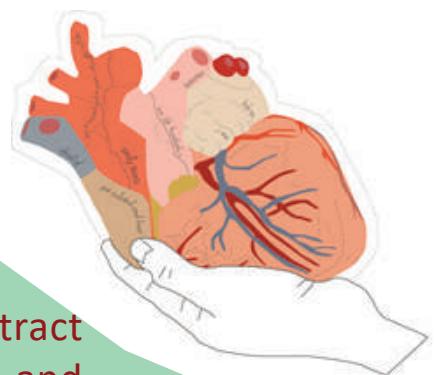
Our hearts are often thought of as our emotional centre. Advances in medical imaging have furthered our understanding of our hearts and studies suggest that trauma from extreme life events can have an impact on the structure and function of the heart.

Drawing on personal accounts of women from the Mountain of Fire and Miracles Ministry and methods of analysing cardiac images, Hidden Stories of the Heart translates experiences of trauma into abstract models and sounds of the heart to explore how stress and disease can affect our vital organs.

The use of papier-maché - which sees distressed, shredded paper transform into a strong structure - echoes the resilience of the women and their hearts.

We invite you to this wonderful exhibition where we will welcome you to listen and share your hidden stories of the heart! Read more at <http://www.artxscience.co.uk/art-x-science-2021/>

Art x Science 2021
Saturday 9 and Sunday 10 October
11.00-16.00
Event location details
Medicine Galleries
Science Museum, London UK



COMPUTER VISION EVENTS

NLP SUMMIT

Online Event

5-7 Oct

ICCV 2021

Virtual

11-17 Oct

AI in Pharma

Virtual

13 Oct

CHEST

Live and On Demand

17-20 Oct

IMVC

Tel Aviv, Israel

26 Oct

Innovation

Summit SF

Hyatt Regency SF

Airport, CA

1-3 Nov

TCT

Virtual Orlando, FL
and online

4-6 Nov

SIPAIM 2021

Campinas, Brazil

17-19 Nov

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.

AM Medical Days

Berlin (Germany) and
Virtual

22-23 Nov

FREE SUBSCRIPTION

(click here, its free)

Did you enjoy reading
Computer Vision
News?

Would you like to
receive it every
month?

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

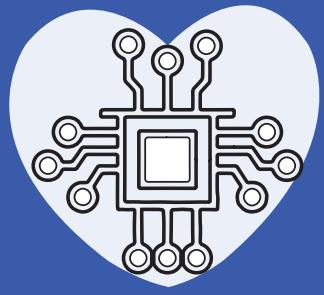
[Fill the Subscription Form](#)
- it takes less than 1 minute!

BMVC 2021

Online

22-25 Nov

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!



MEDICAL IMAGING NEWS

**The new supplement to
Computer Vision News**

October 2021



**The MICCAI community asked
for a magazine fully dedicated to
medical imaging and MedTech.....**

... RSIP Vision did it !!!

**Starting today, Computer Vision
News will be circulated with a new
supplement: Medical Imaging News**



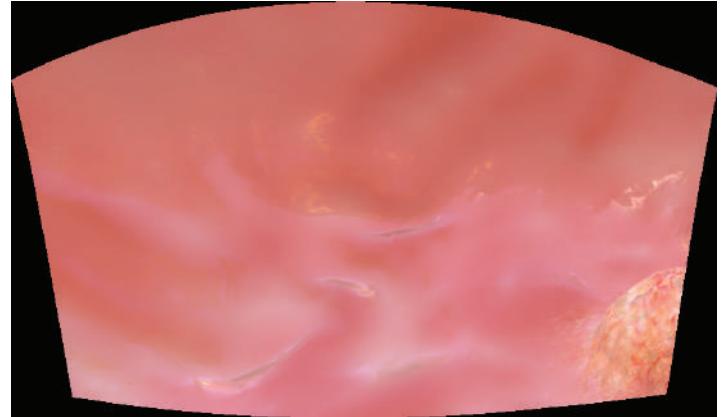
**Subscribe to Computer Vision
News now and get
two magazines in one!**

Bladder Panorama Generator and Sparse Reconstruction Tool

Bladder cancer is a relatively common disease, affecting approximately 1 in 27 men (less common in women). When detected early, survival rates are high. Therefore, early and precise detection is crucial for bladder cancer healthcare. Typical symptoms include blood in urine, painful urination, and even back pain. When these appear, common practice dictates that a cystoscopy is performed – a scope is inserted into the patient's urinary tract until the bladder, which is then scanned in search for lesions and tumors. Additionally, bladder cancer has high recurrence rate, requiring periodical examinations for patients in remission.

The main challenge in bladder cystoscopy is navigation within the bladder. It is vital that the bladder is scanned in its entirety, without leaving unexamined tissue, ruling out missed lesions. Currently, the physician manually scans the bladder repeatedly until they are certain it was completely viewed, and all suspicious areas were examined.

RSIP Vision has recently implemented “Shape-from-Motion” algorithms and created a solution for the aforementioned challenge. This set of algorithms extracts key points from the cystoscopy video used for two tools:

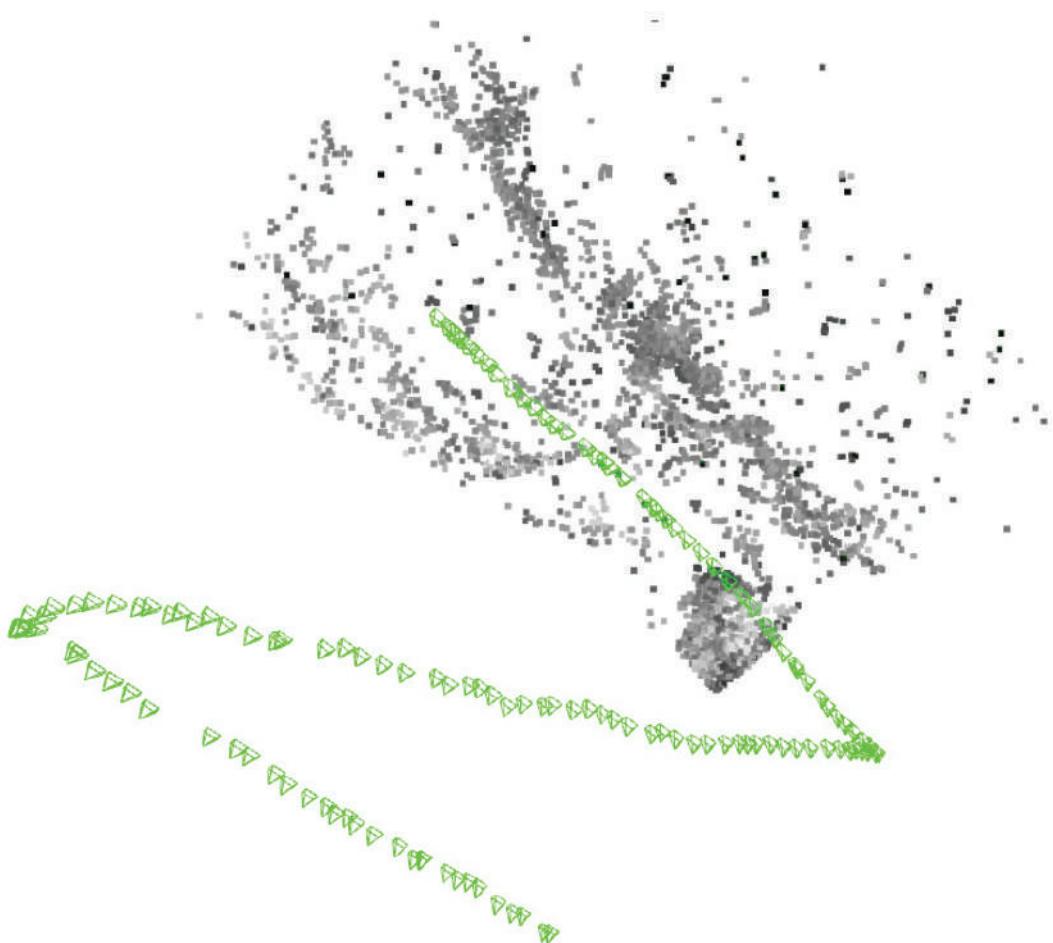
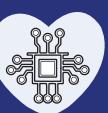


Panorama of a bladder model

For both tools:

- Panorama – Several key points coincide in two or more images. Using these points, the images are stitched into a large, clear, panoramic image of the bladder wall by finding the mapping between a reference image to another
- Sparse Reconstruction – the key points are used in a shape-from-motion framework to get the 3D location of each key point and image camera. This point cloud is a sparse 3D model of the bladder. This model is essentially a map of the bladder. The physician can use it to verify complete coverage of the scan, or as a tool to re-visit suspicious areas.

Additionally, as key points are shared in both tools, matched information is available to the surgeon.



Point cloud of key points on a bladder model and video cameras route

This unique tool can significantly shorten the cystoscopy, as well as increase levels of confidence in the procedural findings. It makes the repetitive bladder scans redundant and provides an image with better accuracy which makes lesion detection easier, as it increases the field of view compared to the narrow cystoscopy view.

Another benefit of this module is that it does not require long run-time, allowing fast and easy implementation in the clinic.

This module was designed to suit the cystoscopy procedure. It can be tailored to any other procedure which uses a scope to scan inner lumens in the body, e.g. gastroscopy, with some alterations. It can also be combined with other computer vision tools such as automatic lesion detection to improve procedural outcomes.

Overall, this is a tool which utilizes advanced computer vision techniques to significantly improve bladder cancer healthcare. More projects in AI for urology [here](#).

32 BEST OF MICCAI 2021



It is always great to see people recognized, especially when it's great friends and folks that you admire very much. Kudos to [Parvin Mousavi](#), [Marleen de Bruijne](#), [Camila González](#), [Nassir Navab](#), [Alejandro Frangi](#) and all award holders from RSIP Vision and all the community!

New MICCAI Fellows 2021



Parvin Mousavi
For sustained High-quality contributions to ultrasound-guided interventions and computer-assisted surgery



Marleen de Bruijne
For strong contributions to medical image analysis and advanced developments of machine learning techniques.



Alejandro Frangi
For outstanding contributions to computational medical imaging

MICCAI Enduring Impact Award



24th INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING & COMPUTER ASSISTED INTERVENTION
September 27 - October 1, 2021 • Strasbourg, FRANCE



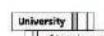
Nassir Navab



- First degrees in maths and physics, Nice and Compiegne
- PhD with highest distinction at INRIA Orsay, 1993
- Post-doctoral fellow at MIT
- Siemens Corporate Research, 1994-2003
- Chair for Computer Aided Medical Procedures & Augmented Reality, Technical University of Munich, 2003-
- Adjunct Professor at Johns Hopkins, 2012-2021



24th INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING & COMPUTER ASSISTED INTERVENTION
September 27 - October 1, 2021 • Strasbourg, FRANCE



MICCAI Young Scientists Award



Changyeop Shin
KAIST



Kun Yuan
University of Ottawa



Camila Gonzalez
Technical University of Darmstadt



Gia Ngo
Cornell University



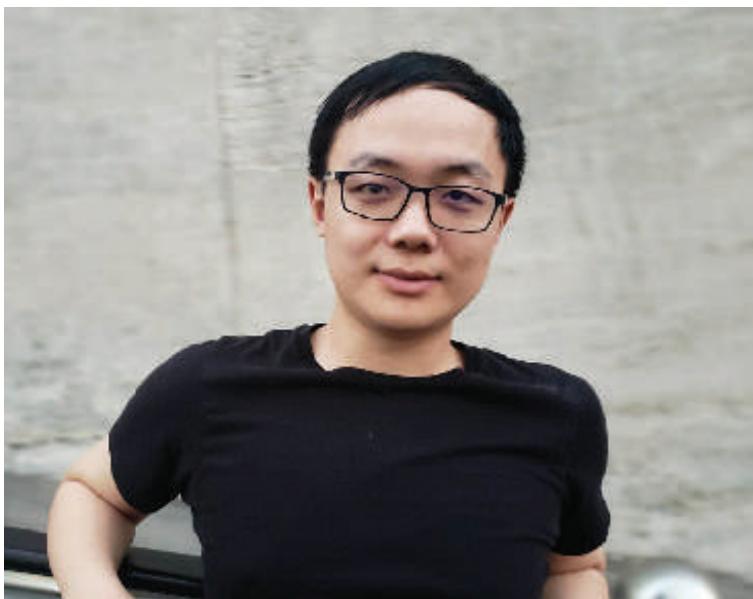
Lisa Kausch
German Cancer Research Center (DKFZ)

- RLP-Net: Recursive Light Propagation Network for 3-D Virtual Refocusing. **Changyeop Shin et al.**
- Surgical Workflow Anticipation using Instrument Interaction. **Kun Yuan et al.**
- Detecting when pre-trained nnU-Net models fail silently for Covid-19 lung lesion segmentation **Camila Gonzalez et al.**
- Text2Brain: Synthesis of Brain Activation Maps from Free-form Text Query. **Gia H Ngo et al.**
- C-arm positioning for spinal standard projections in different intra-operative setting **Lisa Kausch et al.**

Congratulations also to all other MICCAI 2021 Award Recipients, and thank you to all Award Committees!

34 Oral Presentation

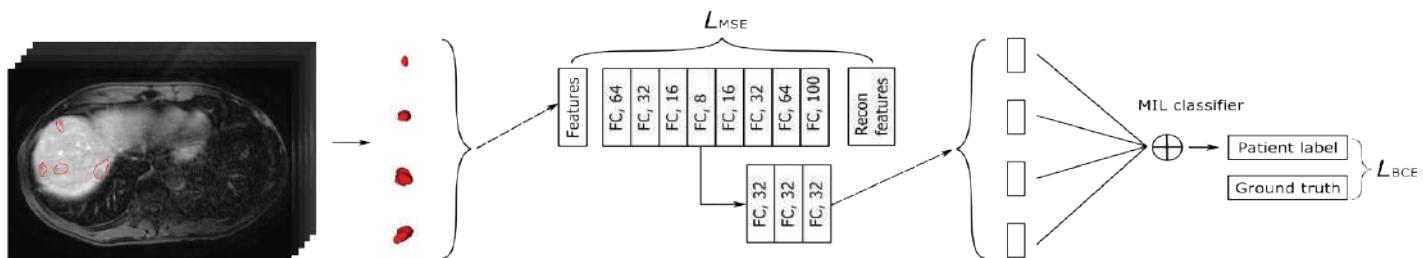
AMINN: Autoencoder-based Multiple Instance Neural Network Improves Outcome Prediction of Multifocal Liver Metastases



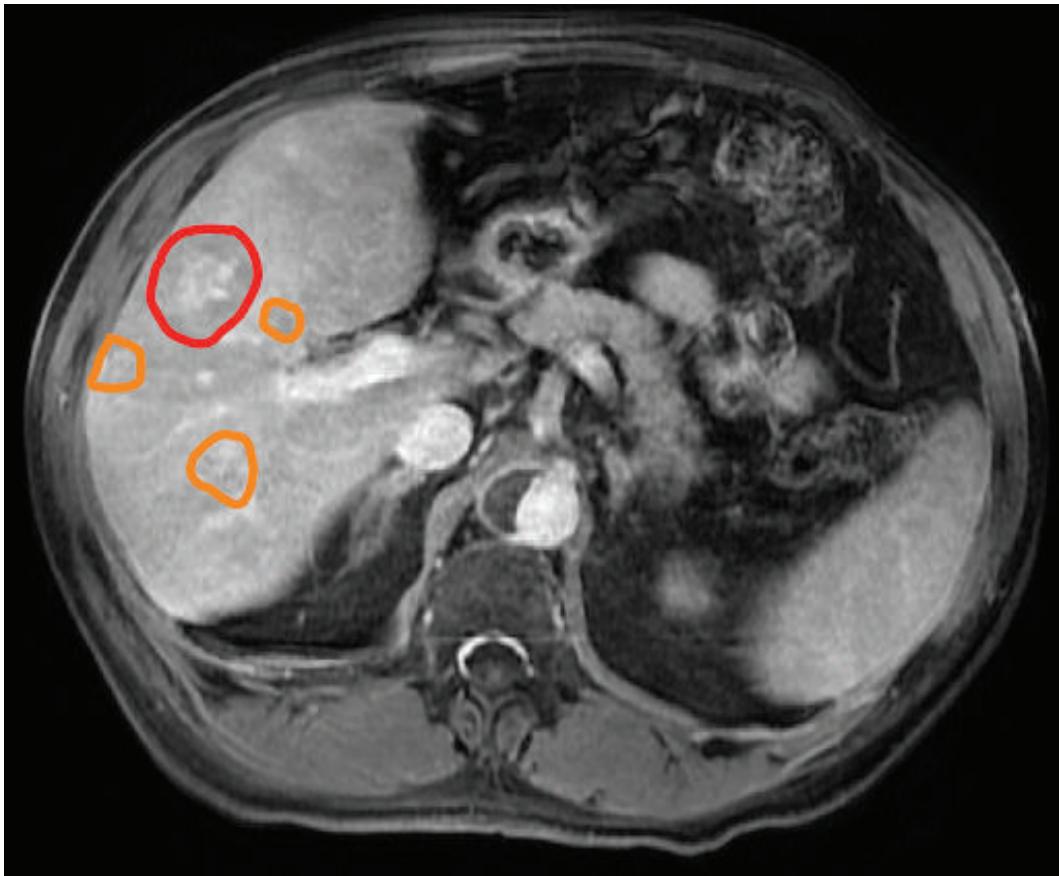
Jianan Chen is a fourth-year PhD student at the University of Toronto under the supervision of Dr Anne Martel. He has developed an autoencoder-based multiple instance neural network for the prediction of survival rates of multifocal cancer patients. He spoke to us ahead of his oral presentation and poster session.

An **autoencoder-based multiple instance neural network** sounds complicated, but it is actually fairly simple. Radiomic features are extracted from MRI scans, an autoencoder selects the features, and then a multiple instance neural network makes predictions based on those selected features. It is all connected and trained end to end.

To the best of Jianan's knowledge, this is the first work that has been designed specifically for **predicting the outcome of cancer patients with multiple tumors**. He spotted a gap because a large proportion of patients have multiple tumors, but existing techniques only look at the largest one or two lesions. Exploring all the tumors can lead to better treatment and improvements in prognosis.



An overview of the proposed autoencoder-based multiple instance learning network. The network structure between the curly brackets is shared for each tumor of the same patient.



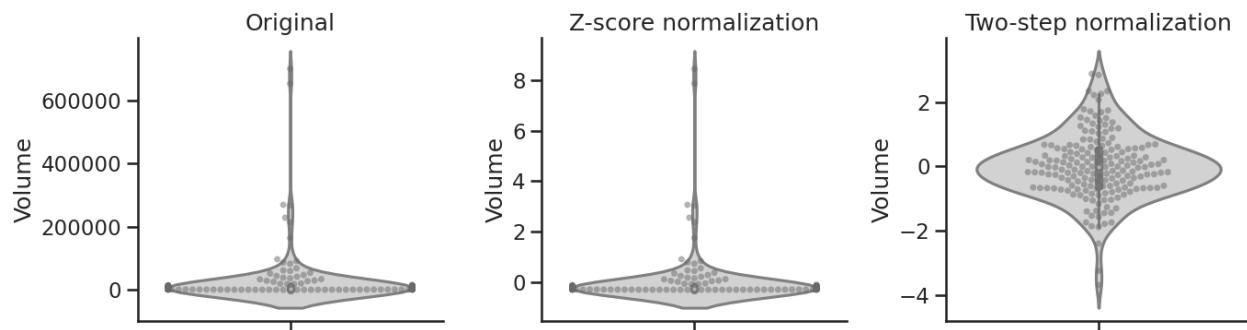
Example MRI scan of a CRLM patient who had multifocal cancer. Existing biomarkers only use information from the largest lesion (marked in red) while ignoring information from other lesions (marked in orange).

The multiple instance learning technique used in this work comes from **digital pathology**, where you have very large whole-slide images that are too big for the computer to deal with, so you crop them into smaller tiles and learn the features. Then at the end you aggregate the information from all tiles to get an understanding of the whole-slide image. This perfectly fits the problem here, which is to predict a patient's survival based on a number of tumors.

Jianan tells us one of the biggest challenges has been **collecting data**.

"When we look at survival of patients, we collect patient data that has had some kind of treatment," he tells us. "You might be surprised, but 80 per cent of colorectal cancer liver metastasis patients, which I studied for my paper, can't receive liver resection and that is the only curative treatment for them. That's why we don't have a lot of data on them. Existing datasets include mostly unifocal patients because we know how to treat them, but it's difficult to collect a large database with multifocal cancer patients. We don't know how multiple tumors affect patient survival or how aggressive a tumor is."

36 Oral Presentation



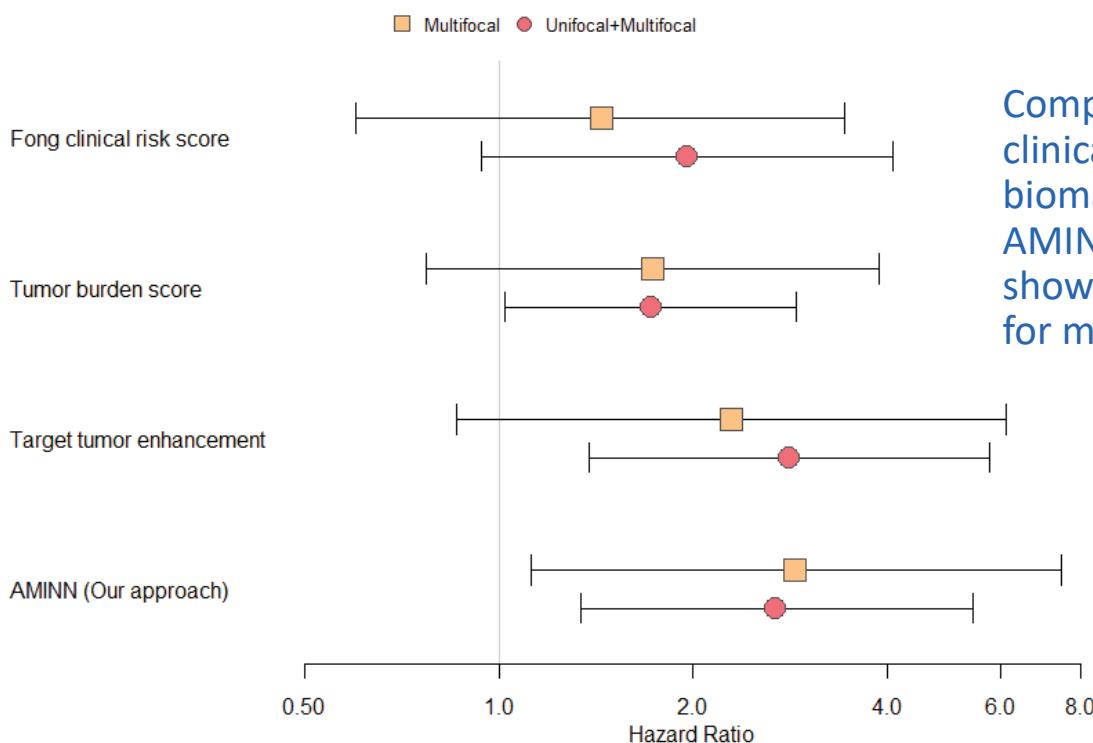
Volume (tumor size) of 181 lesions from 50 patients with Z-score normalization and two-step normalization. Right-skewness of radiomic features can be corrected with the proposed two-step normalization.

Thinking about next steps, Jianan says there are many ways he hopes to take this work forward, including by validating its findings using **independent datasets**, and extending it to other diseases and imaging modalities.

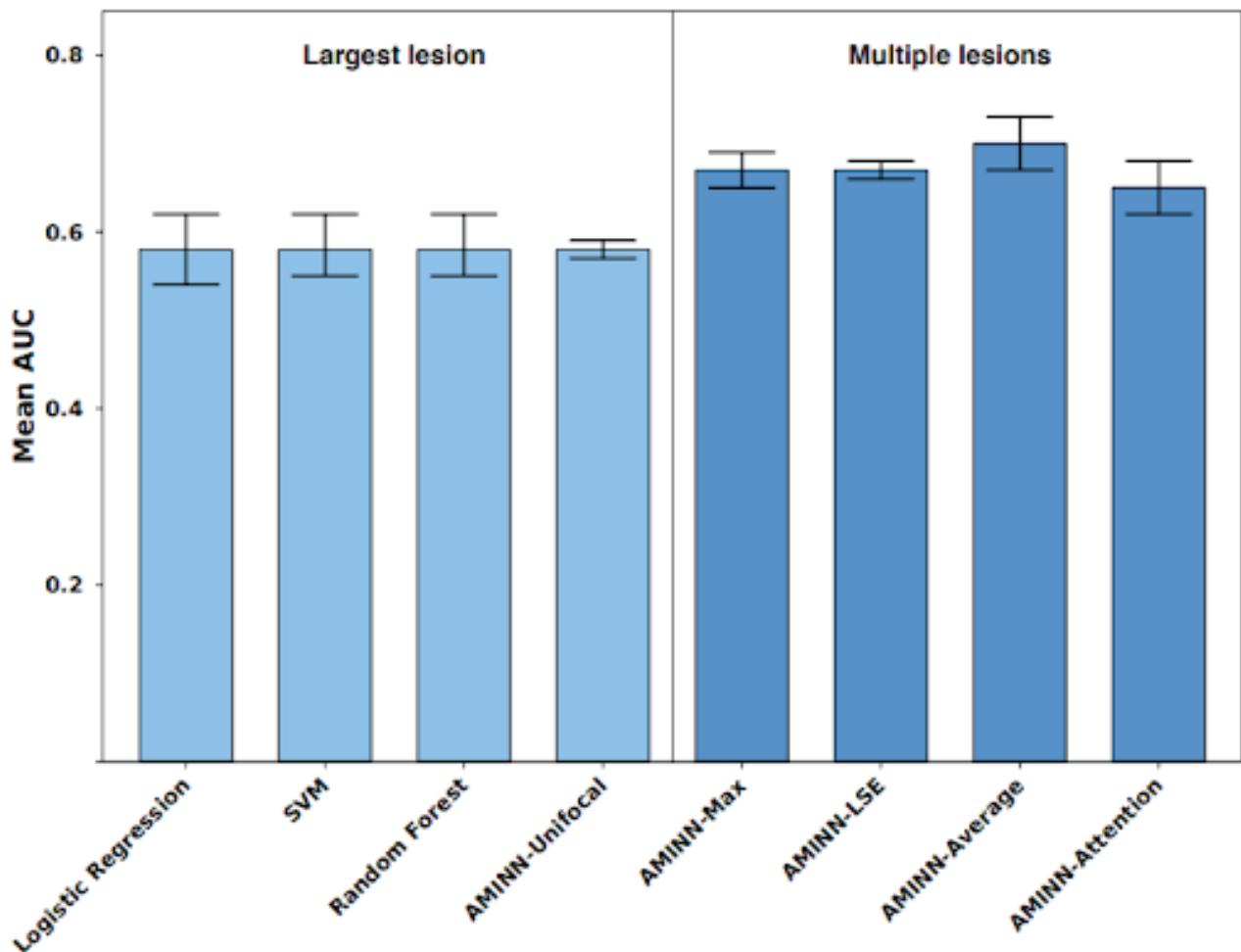
Jianan received his Bachelor's and Master's degree in communications engineering and artificial intelligence, respectively, before switching to **medical imaging with machine learning** for his PhD.

"I wanted to do something meaningful with my research that could affect real people," he reveals. *"I think cancer research is very important."*

Finally, we are keen to know how Jianan has found working with **Anne Martel**, a good friend of this magazine.



Comparison with other clinical and imaging biomarkers for CRLM. AMINN is the only one showing predictive value for multifocal patients.



Comparison of machine learning models for predicting multifocal CRLM patient outcome in 10 repeated runs of 3-fold cross validation. Taking multiple lesions into account improves outcome prediction.

"Anne is a super nice supervisor!" he smiles. "She gives me full freedom to explore what I want to study and always encourages me and gives me advice when I am stuck. She taught me that it's important to not only focus on your research project itself, but also to help out in the community by becoming a reviewer and teaching. This paper and my previous MICCAI paper were all inspired by something I found while reviewing or teaching, so it is great advice!"

38 Poster Presentation

On the relationship between calibrated predictors and unbiased volume estimation



Teodora Popordanoska is a PhD student at KU Leuven in Belgium, under the supervision of Matthew Blaschko.

MICCAI this year is an extra special occasion for Teodora as she has had her first accepted conference paper!

Her work investigates the relationship between model calibration and unbiased volume estimation. She spoke to us ahead of her poster session.

Model calibration is especially important in medical applications, because the output score of the model should reflect its own **trustworthiness**. In other words, the model should be calibrated. A model is said to be calibrated if the score for an input corresponds to the probability of the prediction being correct.

"For example, from all the samples that are predicted with a probably of 0.9, we want 90 per cent to be accurately predicted," Teodora explains. *"We want this to hold for all probability scores. The problem is that modern neural networks are poorly calibrated, as shown in the 2017 paper 'On Calibration of Modern Neural Networks'. They tend to give over-confident predictions. That's why there is a growing interest in this field of model calibration. For model calibration we care about a quantity called calibration error."*

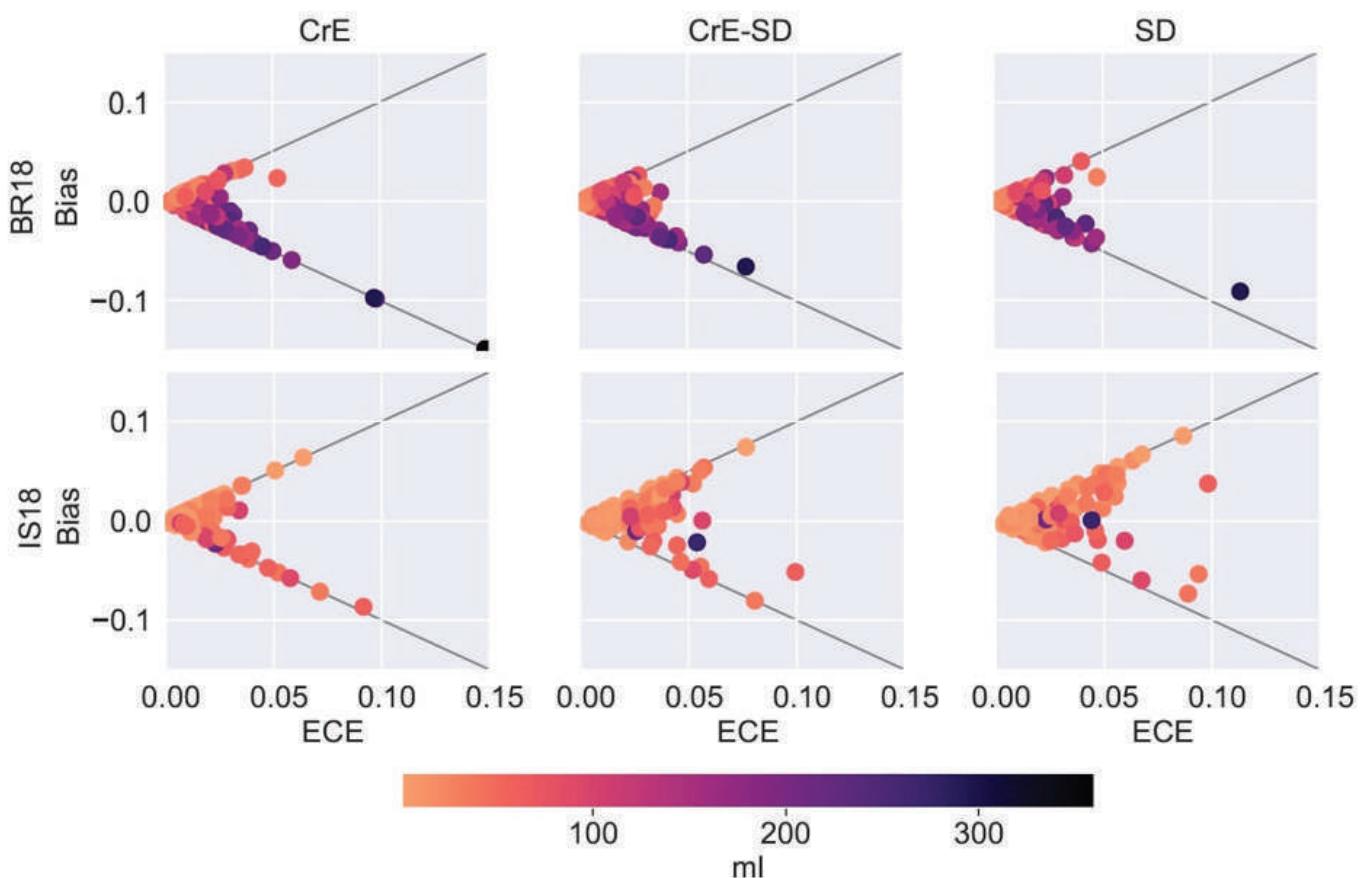


Fig. 1: Scatter plots on BraTS (top row) and ISLES (bottom row), color-coded by tumor/lesion size (in ml). Every point in the plot represents an image.

The second part of the title refers to **volume estimation**. In medical image analysis, the segmentation of an image is usually performed with a **neural network** and is mostly used to calculate certain biomarkers. In the medical domain, the volume of a tumor, organ, or lesion are important biomarkers.

From a segmentation, one can obtain the volume by summing up **the probability scores for each voxel**. An important quantity in this case is the bias of volume estimate. Ideally, to have the true volume, there would be zero bias.

"The main theoretical result from this paper is that the absolute value of the bias is upper bounded by the calibration error," Teodora clarifies.

"If we're optimizing calibration, we're also simultaneously reducing the bias of the volume estimate. If the calibration error is zero, then it means that

40 Poster Presentation

we had an unbiased volume estimate, and we get the true volume. In this sense, the result is not specific on the architecture of the model or the type of organ or tumor or whatever we're measuring the volume of."

The result is a fundamental mathematical result that has been empirically validated on two datasets and 18 models, trained with several loss functions and calibrated with multiple calibration strategies.

Designing a calibrated model by itself is not a trivial task, so the team are currently working on developing **a new method of training well-calibrated models.**

Teodora is the third Macedonian that we have interviewed over the years, following her friend [Ivana Najdenkoska](#) just yesterday and [Jelena Frtunikj](#) at CVPR19, as part of our Women in Science series.

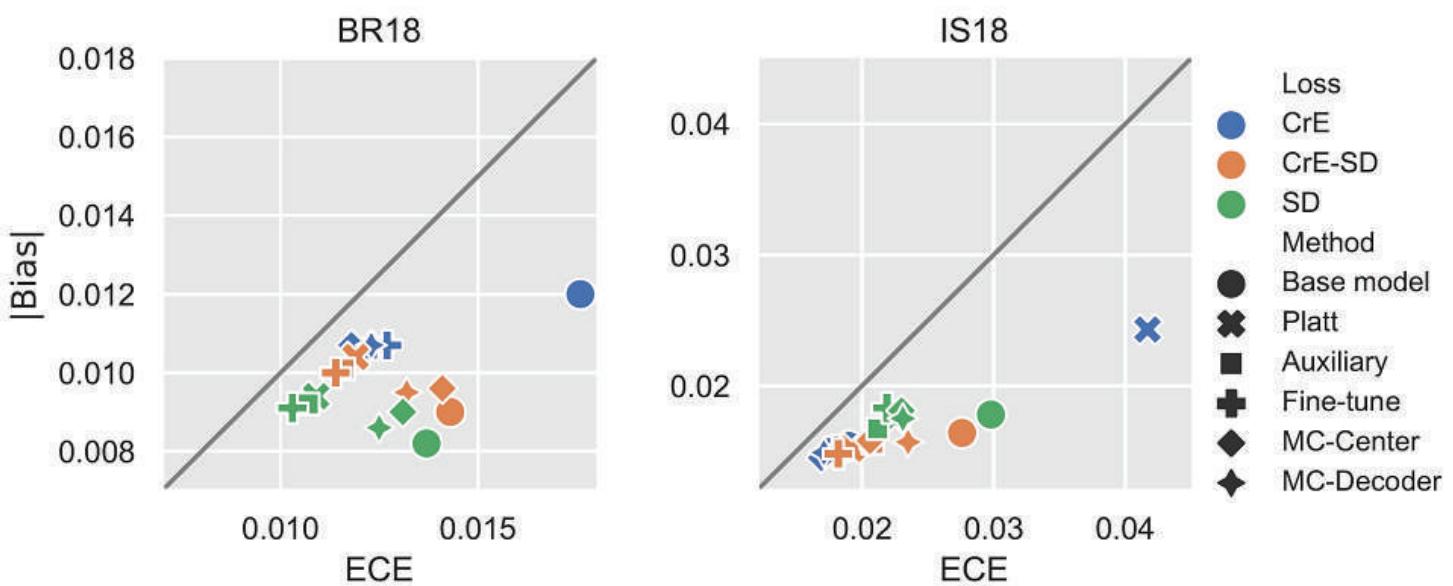


Fig. 2: Scatter plots of ECE versus $| \text{Bias} |$ for all combinations of loss functions and calibration methods. The Pearson correlation between $| \text{Bias} |$ (computed as mean of absolute per-volume biases) and ECE (mean per-volume ECE) is 0.30 ± 0.21 for BR18 and 0.91 ± 0.04 for IS18.

“Macedonia is a very small but beautiful country,” she tells us. “The people are warm and nice. I’m always glad to go back there!”

Teodora reveals the most challenging part of this work for her was discussing the clinical relevance of the theoretical results.

“This is my first time working with medical data, but thankfully all my co-authors have plenty of experience in this area and in the medical domain, so it worked out great,” she adds.

“In medical applications, we want the models to be trustworthy and we want to have unbiased volume estimates. With our work, we are showing that focusing on model calibration is sufficient and calibration error is in fact a superior model selection criterion also with respect to volume bias.”

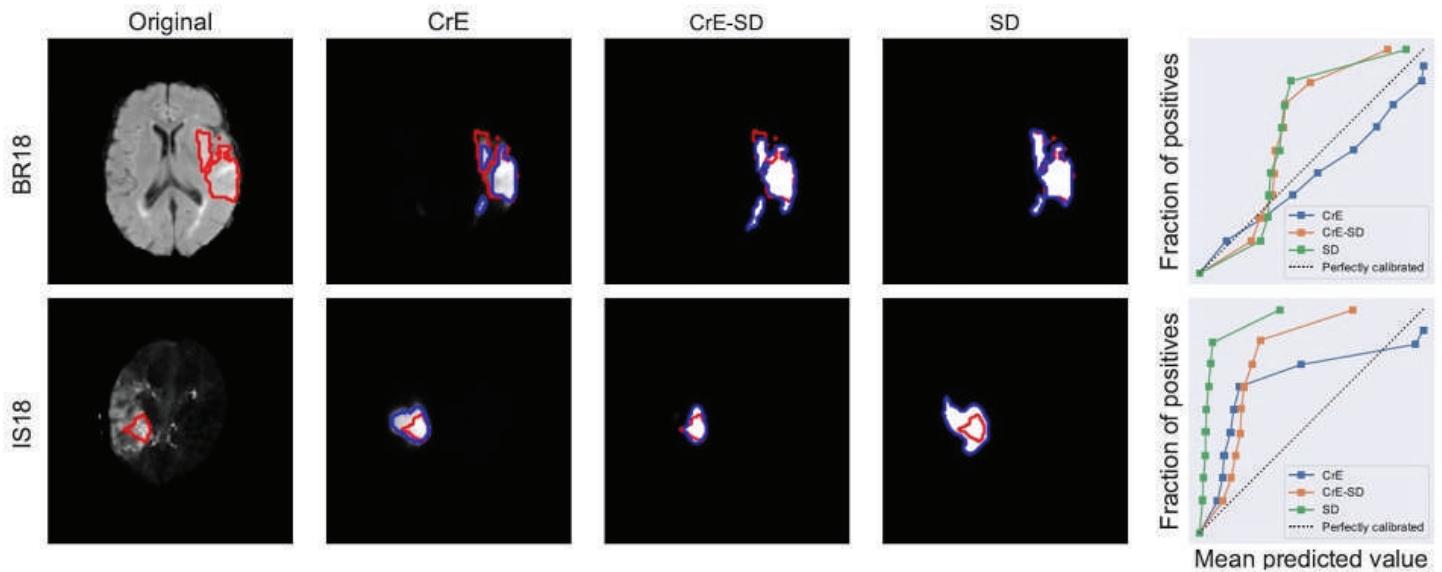


Fig. 3: Qualitative examples of the predictions from a base model on BraTS (top row) and ISLES (bottom row). The red line represents the delineation of the ground truth. The predicted delineations after thresholding at 0.5 are overlayed in blue. The last column shows the calibration curves for the volume.

42 Interview



Mert Sabuncu is an Associate Professor in Electrical and Computer Engineering at Cornell University and Cornell Tech. He also has a secondary appointment at Weill Cornell Medicine in radiology. He speaks to us about his career to date and his hopes for the future of medical imaging.

LAST MINUTE NEWS: Mert has just been selected as winner of the Michael Tien '72 Sustained Excellence & Innovation in Engineering Education Award, the highest award for teaching at @CornellEng... Congrats, Mert!

Mert, can you tell us about what you do in the Sabuncu Lab at Cornell?

Our research group focuses on developing computational methods, i.e., algorithms, for medical imaging problems, including anything from MR acquisition to downstream applications using medical imaging data in the context of clinical workflows.

How did you come to work in the field of medical imaging?

I did my PhD at Princeton University in electrical engineering. Princeton doesn't have a medical school and when I first started my PhD, I wasn't sure what I wanted to work on. I knew that I wanted to do something that involved image processing, and in the early 2000s, there was some excitement around machine learning, so I started learning about

that. While kicking around and thinking about it, I ran into a bunch of neuroscientists at Princeton who did functional MRI research. One of the group, James Haxby – who is now at Dartmouth College – introduced me to some image analysis problems in fMRI. That was my first experience with medical imaging. Then I had the opportunity during grad school to intern at Siemens Corporate Research - Siemens Healthineers now - which is also located in Princeton, New Jersey. I worked there over the summer and during the year I would collaborate with them. They have a heavy focus on medical imaging.

What convinced you that this focus was the right one for you?

I'm a strong believer that the research we do should have a real-world impact. I'm very pragmatic in

that sense. I view the healthcare and biomedicine domain as a very interesting area where we can make a strong impact and hopefully have a positive influence on many people's lives. In my family, I have been surrounded by doctors and medical researchers from a very young age, and I guess that's what first attracted me to biomedical research.

Do you feel that the work that you and your team are doing, and the work that the MICCAI community is doing, is having that impact?

I think that as a community what we do at MICCAI has a big impact. I also think that the research that comes out of my group has an impact in sometimes very non-obvious ways. It's important to remind oneself that what we do can feel a little bit like basic research. A little bit removed from real-world applications. There are still some open questions that we need to work on before we can take our technologies and apply them to real-world problems. That said, as I move along my career trajectory, I'm hoping I will increase my emphasis on real-world translation. That's why nowadays I'm focusing on real clinical collaborations and understanding clinical workflows. I'm hoping to put more effort into translating our technologies into the clinic, either by commercialization efforts, or at least trying to implement things in an academic hospital setting.

Why does it take so long sometimes to translate MICCAI research into the clinic?

There is always a gap. Whatever field you're in, especially in the academic setting, there's going to be a gap between the research and the way it impacts people's lives in the real world. The healthcare world has a lot of stakeholders, including patients, insurance companies, regulatory bodies, hospitals, doctors, and researchers. Having them all aligned and getting through all their needs and requirements is a big challenge. Taking an idea from a concept to a real product that can be used in a clinical workflow on patients, there are a lot of obstacles along the way.

The other challenge is as researchers on the algorithmic side, we often focus on toy problems. That is a good starting point but can be distracting in terms of what matters in the real world. We spend a lot of our bandwidth on these "artificial problems" and make good progress, but to take those breakthroughs and translate them into the real world, there are other challenges that we aren't focusing on. There are obviously exceptions, but there's not a lot of incentive, at least in the academic world, to move along those steps. A lot of those incentives are in the more basic research.

Do you think today's CLINNICAL, MICCAI's first clinical day, is a step in the right direction?

I think these types of ideas and especially communities like MICCAI make a difference. They enable different groups of people to communicate with each other to

44 Interview

understand and align their motivations. We speak different languages, so just coming up with a common language and understanding the different sets of problems people are working on is going to go a long way. There is often a gap between these groups because of a lack of communication, which societies like MICCAI will help us get around.

This is our second virtual MICCAI, and although it is still a fantastic event, many of us long for the days when meeting up was so much easier. How do you feel about this?

I definitely see the pluses of having virtual meetings. Back in the day when everything was in person and we didn't have virtual posters and virtual talks, it was harder for people from certain countries to travel. You couldn't afford it, or it was hard to get a visa, or you had family responsibilities and it was hard to leave your home. There were all sorts of constraints. From that point of view, the virtual format is a good thing. It enables broader participation. But I also think that it's a much more limited experience. It results in a situation where we don't get to randomly run into people and have those social networking opportunities. There's also something about traveling to a location and setting aside five days of your time to focus on the science. You get to understand properly what's going on, listen to talks, and meet different people. When you're not doing that, it's hard to set aside that bandwidth. From that point of view, I think it's a negative to have

everything virtual. Going forward, maybe we can take the positives of both - in-person meetings for people who can make it but supporting virtual talks and posters for people who can't. We need better technologies for socializing too. Gather.town seems to work well in my experience, but there are others and I'm sure the technology is going to improve very quickly here.

Do you have any words for first-time participants at MICCAI about how best to take advantage of everything that is on offer?

I vividly remember my first MICCAI. It was in 2007 - not too long ago! What I remember about those early conferences was it felt quite intimidating. I felt a lot of anxiety about being in a roomful of people trying to give a talk or trying to introduce myself to established professors. It took me a while to get over that and build up the confidence to make the most of it. I would strongly encourage people to acknowledge that it's challenging. When you're at your first conference, a lot of the stuff is going to be brand new to you. You're not going to understand everything, you're inexperienced, you don't know many people, so you need to take it easy. I guess the advantage of the virtual format is you can be more proactive and absorb much more than you possibly could if you attended in person. But the potential risk when you're attending from home is that you try to multi-task or there's other stuff going on around you, so you can't focus. The more bandwidth you



can allocate to the conference experience, the more you will get out of it.

Also, don't be shy. I assumed a lot of the more senior people weren't approachable and it took me a while to connect with them. But now I know that we're all human, so it just takes saying hi or introducing yourself or texting or emailing them. People are usually responsive. Sometimes they are busy, you just need to understand that, but otherwise people are very responsive.

Can you assure participants that if they feel awkward the first or second time, it will change eventually?

100 per cent. This is true for anything, but it is definitely true for a conference like MICCAI. The more you go to these events, the more time you spend amongst people, the more talks you listen to, the more papers you read, the more papers you write, the more talks you give,

the more confident you're going to be in general. There's a big difference between how comfortable you are at the end of your PhD versus the beginning.

At the beginning, everything is new. At first, I didn't even know how to look at a poster and absorb that information! Now, it takes me 10 seconds to go through a poster, but it's taken me years to get here. Not everyone will stick with this field and come to this conference as many times as I have, but if you're going through grad school that's a span of 3-5 years at least, so during that time, the more you attend these conferences, the more experience you'll get and the more confident you'll become. It's also important to realize that a lot of it is psychological. Imposter syndrome is real and even people at very senior levels suffer with it. It's important to acknowledge that and try to build up your confidence proactively.

How do you build it?

By little steps and by communicating. You need to have good mentors and talk to people. I try to instil this in my own students. We talk about how someone feels in their first presentation or their first poster. Just acknowledging that everybody goes through this. None of us are born as Nobel laureates or full professors. It takes time to learn things and build up that confidence. For some people it doesn't come very naturally. It's not even strongly correlated with what you know or what you have experienced, it's just

46 Interview



something you have to be aware of and nurture.

Thinking about your students, when they get their first paper accepted, is your excitement comparable with the excitement you felt in the past when the same thing happened to you?

My students are much more successful than me! They work harder, are more motivated, more productive. It's obviously a function of the competitive environment they're in. It's also a function of the resources they have. But it is true that having your paper accepted at something like MICCAI is very valuable in an objective way. It's going to be an important line in their CV. It can make or break a career sometimes. If you have a few solid papers during your PhD, that might make a big difference in terms of what kinds of career prospects you have versus if you don't. Sometimes a lot of that is random. I tell my students there's a lot of noise in the review process. It's not perfect the way we review papers and decide

their fate. It's important to understand that noise and understand human psychology, how reviews happen and how reviewers respond to certain things. You have to roll the dice multiple times and be persistent because if you declare a loss that means you've given up on an idea or a paper and there's nothing more to do. But the key is to take what you've learned and make your paper better and try again. Learn from your mistakes and from the feedback. Also, sometimes acknowledge that the feedback really is just noise, and you can ignore it. It's important for mentors to point out what's noise and what's not.

Do you have any wow moments from your students that you would like to share with us?

That's very hard for me! I have many good students and a lot of respect for all of them. I can highlight one though, Meenakshi Khosla, who just graduated and is now going on to be a postdoc at MIT with Nancy Kanwisher in neuroscience. Meenakshi has done some very interesting work where she used deep learning models to characterize how the brain responds to visual and auditory stimuli. There's a lot of excitement around the idea of attention in machine learning – whether you're looking at images, text, or language data, if you build models that can incorporate an idea of attention, it will yield better solutions. Meenakshi was building a model where the input was a picture, and the output was how your brain

would respond to that picture as measured by functional MRI data. She built an attention mechanism into that model. You show the model a picture and it figures out where to pay attention to in that picture, then encodes it and somehow figures out how the brain will respond to that. It worked really well and predicted how the brain responds.

What's interesting is when you look at what the model is paying attention to, it is very strongly correlated with eye-gaze data, which the model was not trained on at all. We had some eye-tracking information and we looked at where the model was paying attention to and where humans are paying attention to and there was an incredibly strong correlation. I was blown away by this! To me that was one of the most interesting themes. I get excited about finding similarities between these artificial neural networks and human biology. But for some reason, that element of the work didn't attract the appreciation I expected it to. You know when you write a paper that you're excited about, it gets accepted, it gets published, but then people don't pick up on the thing that you're excited about? I find that frustrating, but there's nothing you can do about it. Maybe 10 years from now somebody will pick up on it and it will have a bigger impact!

I am sure we will have many more surprises to come in the future from your students.

I hope so, yes. I'm looking forward to

it!

Do you have a final message for the MICCAI community?

I have lots of messages! One thing I'd like to see the community focus on more is translation. The clinical day is important in that sense. In academia, a lot of the incentives are on getting your paper out, getting a good publication, obtaining funding, and these are not strongly correlated with the effort to translate your ideas into the real world. That last mile is not rewarded in academia. Maybe we need to come up with a way to incentivize that in the MICCAI Society. It could have an emphasis on clinical translation as an area of focus so that we can award papers in that area and recognize that effort more directly. MICCAI and other societies could influence grant-giving entities to focus on translation and not just basic research because at the end of the day the stuff that we work on only matters to the extent that we can turn it into something in the real world. I don't mean that we shouldn't be doing any basic research, but we should value that last mile effort much more than we do right now. For anyone new to the field or struggling to figure out what they want to work on, I think that's the big opportunity. We've made a lot of interesting developments in basic science or basic research, but we're still lacking on the translation.

You have launched that message out there now, so let's see if your colleagues will catch it!

48 Women in Science

Catarina Barata is an Assistant Professor at Instituto Superior Técnico in Lisbon as well as a Research Fellow at Institute for Systems and Robotics.

[More than 100 inspiring interviews with successful Women in Science in our archive](#)

Catarina, you live in Lisbon and you are Portuguese.

Yes.

You work mainly on skin cancer.

Yes, I started on skin cancer during my Master's thesis, so it's a while ago. My main focus in the beginning was to develop artificial intelligence systems for diagnosing skin cancer, in particular melanoma, which is a huge problem here in Portugal and it's becoming worse! In recent years, I've tried to move a little bit beyond that. More than the diagnosis, I want to understand this disease, and how people treated for this disease react to the different therapies.

Why is this problem becoming worse?

It is related to the way we behave. If you think about our grandparents, when they went to the beach, they'd be fully dressed. In the present, people are more aware of the dangers of the sun and getting sunburnt, so they are starting to wear protection. But the skin remembers! It's not like you get a sunburn in the summer, and you get melanoma next year. There is a big

gap between the sunburn and the disease actually appearing. That's why, in some countries like Australia, where people are very fair skinned, people are in direct contact with the sun and get a lot of sun burns. Then the disease appears after like twenty-something years. That's why it is a big problem. Sun burns that people got in the 90's are coming back to haunt them now.

Would it be better to go to the beach in full clothes like one hundred years ago?

No, but you should respect the hours.



“This can make beautiful things happen!”



Doctors say you should go during the morning and avoid it from noon to 4 PM, but most people don't do that. Here, people spend the entire day at the beach.

We like the beach!

We do! We are Portuguese!

I love Albufeira. What brought you to this field?

That's a very interesting question. I have a biomedical engineering degree. I studied many engineering topics, from chemistry and physics to mathematics. I didn't know what I wanted to do, I just wanted to go into research. In the last year of my degree, I did a course on machine learning, and I fell in love with it. It was just perfect for me! Then I was looking for a Master's thesis on this topic. This thesis about skin cancer just popped out. It was perfect! I

combined the biomedical part with machine learning. Let's give it a try! Then I realized that this is very challenging. Skin imaging is in between real imaging, that you find in different computer vision tasks, and medical imaging, which are 2-D images. You have challenges from the two sides. That combined with machine learning just convinced me to keep working on this!

Did you meet with some doctors before you started?

Yes, the co-supervisor of my Master's thesis was a dermatologist. I met doctors and actually I still work with doctors. We need them to keep working on these things, not only for the data but for everything, for feedback, to understand if we are going in the right direction, or if we need to change a little bit.

50 Women in Science

Sometimes we have different ideas.

You told us only half of the story - the half of the story about the research you are doing. I know that you are also teaching. Can you tell us a little bit about the second part of the story?

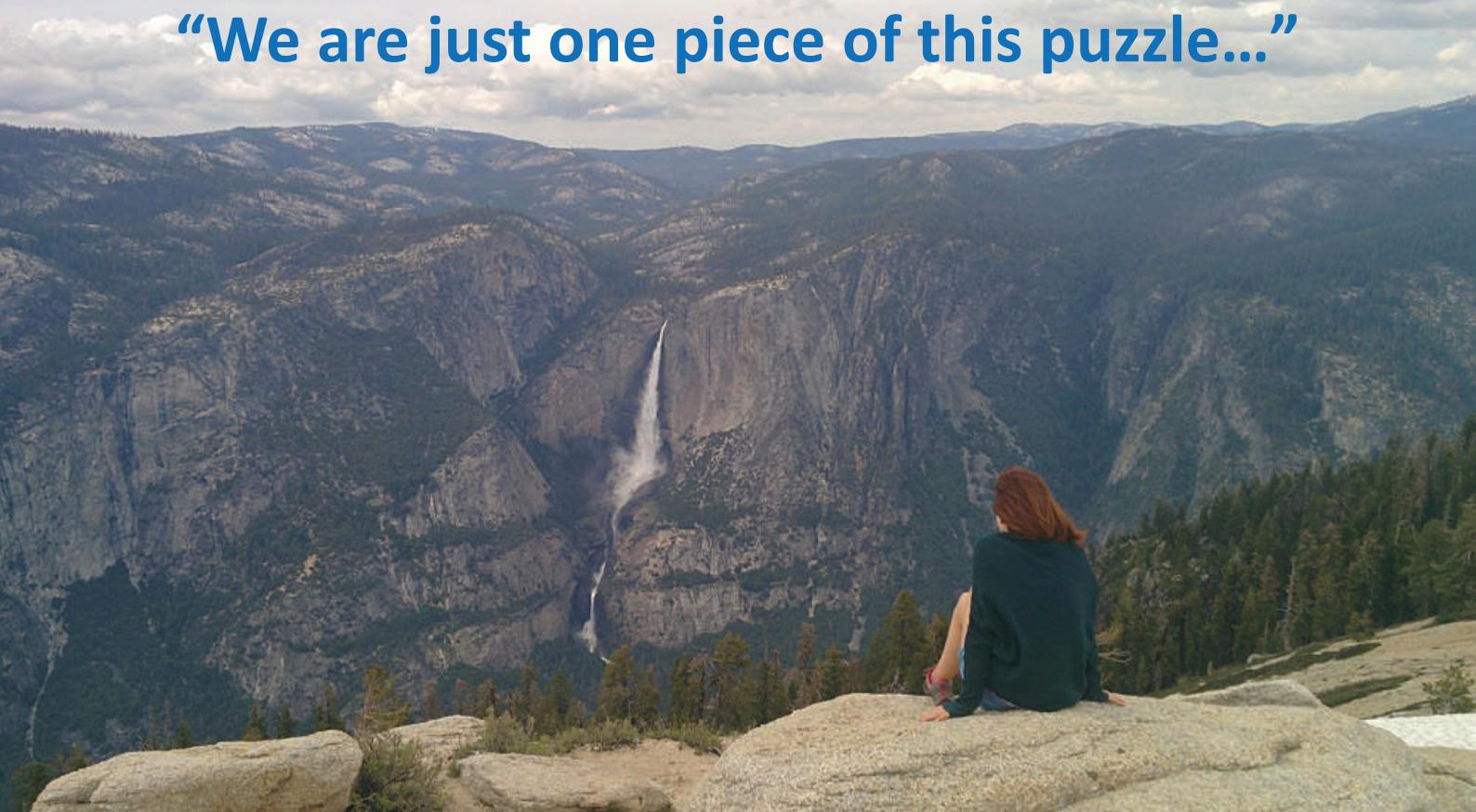
It's funny because I just came out of a meeting to prepare for the next course on machine learning that starts in two weeks. At the moment, we have more than 300 students in this course. As you can imagine, it has changed from 50 students ten years ago to 300 now. A lot of people are interested in these topics. We have to let them know, for example, about these recent deep learning models that everyone is hearing about. We have to contextualize these for the students. That's mainly what I do. I teach machine learning at the university. I

also supervise students. I supervise several Master's students in different computer vision topics and in robotics as well, mostly using machine learning. We try to challenge them to use recent methods.

You have supervised more female students than male students. How did this come about?

To be honest, I don't know. I don't know if it's because I am a female supervisor. Sometimes I know that it plays a role, although my PhD and Master's supervisor was male. We share the students sometimes. What is changing is the way women approach engineering. When I took the machine learning course 10 years ago... no 11 years ago, there were six female students out of about 60 students. It's less than one tenth of the students. Now, we have

“We are just one piece of this puzzle...”



higher numbers. In my class, it's not half-half, but things are definitely changing. I don't know if it's because I'm a female supervisor. I work on other medical applications as well. We also do security, surveillance, and fire detection as well. Maybe that engages the female students a little bit more.

Do they choose you or do you choose them?

They apply, and then I choose. I don't choose based on gender. That's not something I want for me as well. I choose based on who is best for the position.

Do you ever see students making mistakes that you made when you were in their shoes?

[laughs] Yes! Often! One of the biggest things that I see that happens often in class is when they try to program something for the first time. It works, and all of a sudden, they just change a little bit and everything stops working. I remember myself when I was doing my Master's thesis. It worked, and then I changed the resolution of the image just a little bit. It stopped working, and I was like, "Oh my God! It doesn't work anymore!" Students are exactly the same. People get very scared! [laughs]

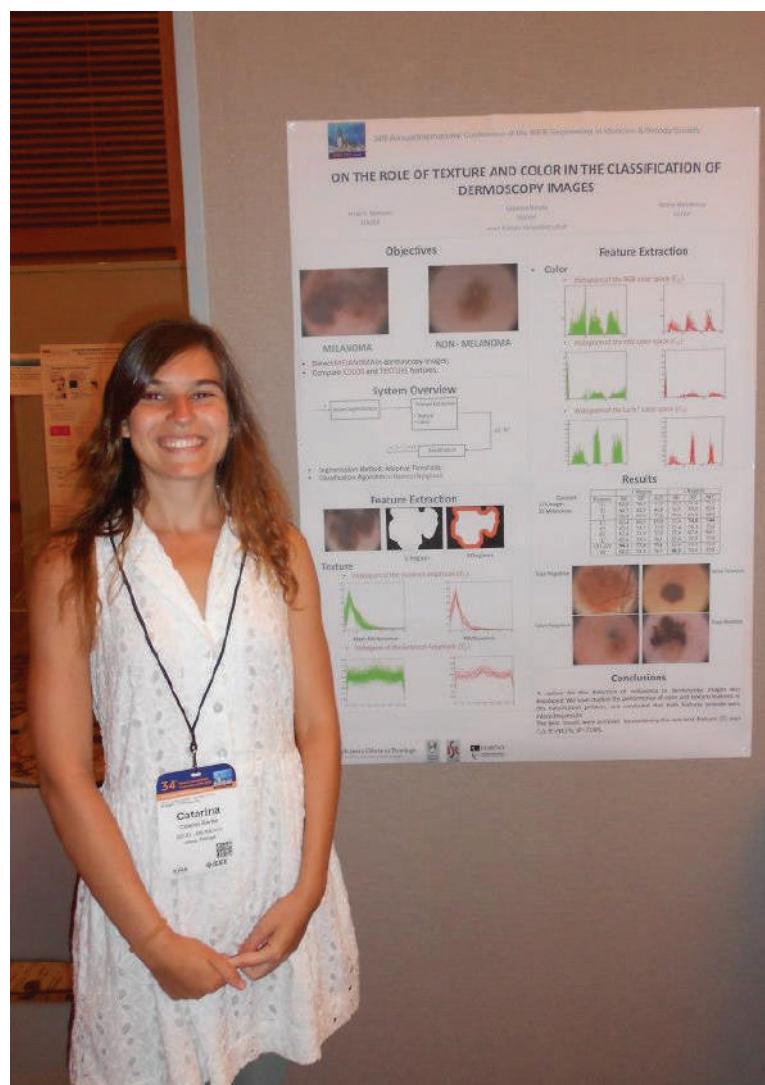
What is the largest class that you have?

We have theoretical classes where the number of students is higher. Then we have small lab classes,

where the connection with the students is closer. In those classes, we have like 30 students.

How do you find yourself speaking to an audience of 30 people? Does it come naturally?

I'm very talkative! People used to tell me that I'm a good communicator. I don't know if that's true. When I'm in the classroom with 30 people, I feel that it's closer so I can approach almost each one of them individually. That's something that I like. I teach these classes in shifts, and each shift will be different. The students will have different capabilities and different backgrounds.



52 Women in Science

Let's say that I'm your little brother, and I want to start teaching. What advice would you give me?

The first thing you need to pay attention to is first of all, you need to prepare the classes. It's a mistake to imagine that you just go in there. If you have a background in machine learning, you can't just go there and rely on that background. You have to prepare. The second thing that you definitely need to do is guide the students. That will make a big difference in the way you teach. As I try to engage with the students, to make them talk to me, tell me their difficulties and if they understand. That allows me to conduct the class.

How do you engage them? It's difficult to engage people!

Yes! Especially in this Covid time.

Everyone is just taking classes online.

Share your secret with us please!

[laughs] In these 30-student classrooms, at a certain point, you start to know their names. You can try to call them by their names. That's something I did for the Zoom classes. I remember some students were more talkative than others. I remember saying: "*I want to hear a different voice.*" I asked them to try to talk to me. Just looking at their faces, I don't understand if they are following what I'm saying or not. If you are solving a problem, try to make them solve a problem with you. It's hard. Like I said, it depends on the shift. Some students are very proactive, some aren't. Some are afraid to fail.

“Opening up our work to others!”



I'm sure that it is much more fun to do in-person teaching.

[Laughs] Yes, definitely!

Can you tell us a story that happened in class?

I look younger than I am. Whenever I go to the classroom, usually the students think that I'm another student! [laughs] I enter the classroom. I put everything on the table, and they remain outside. They just remain outside! I go to the door and tell them, "You can come in!" They are like, "Are you the professor?!"

I can relate: as a very young graduate, I interviewed candidates for my university. One of them tried to explain to me the difference between "*people of his generation*" and mine... he didn't realize I was much younger than he was! [Both laugh]

When they get comfortable with me, they say things like, "So teacher, when did you finish your degree?" I tell them, "I finished my degree way before you started yours!" [laughs] It's not impolite. They are just dying to ask! [laughs]

Is there anything you'd like to tell the world?

There's something I want to share about my field of skin cancer. We need to bring people together. We are researchers in computer vision and machine learning. We are just one piece of this puzzle. For MICCAI, if we are working in medicine, we should bring in doctors. We should bring researchers from other fields. This can make beautiful things happen! Not just looking at the things we do, but opening up our work to others!

[**More than 100 inspiring interviews with successful Women in Science!**](#)



54 Medical Imaging Technology



Ali Hatamizadeh is a research scientist at NVIDIA. He received his PhD and MSc in Computer Science from the University of California Los Angeles.



Prerna Dogra is a Senior Product Manager for Healthcare at NVIDIA, where she leads the Clara Application Framework and the collaborative open-source initiative Project MONAI.

Introduction

Recently, transformer-based models have gained a lot of traction in natural language processing and computer vision due to their capability of learning pre-text tasks, scalability, better modeling of long-range dependencies in the sequences of input data. In computer vision, vision transformers and their variants have achieved state-of-the-art performance by large-scale pretraining and fine-tuning on downstream tasks such as classification, detection and segmentation. Specifically, input images are encoded as a sequence of 1D patch embeddings and utilize self-attention modules to learn a weighted sum of values that are calculated from hidden layers. As a result, this flexible formulation allows us to effectively learn long-range information. This warrants the question, what is the potential of Transformer-based networks in Medical Imaging for 3D segmentation ?

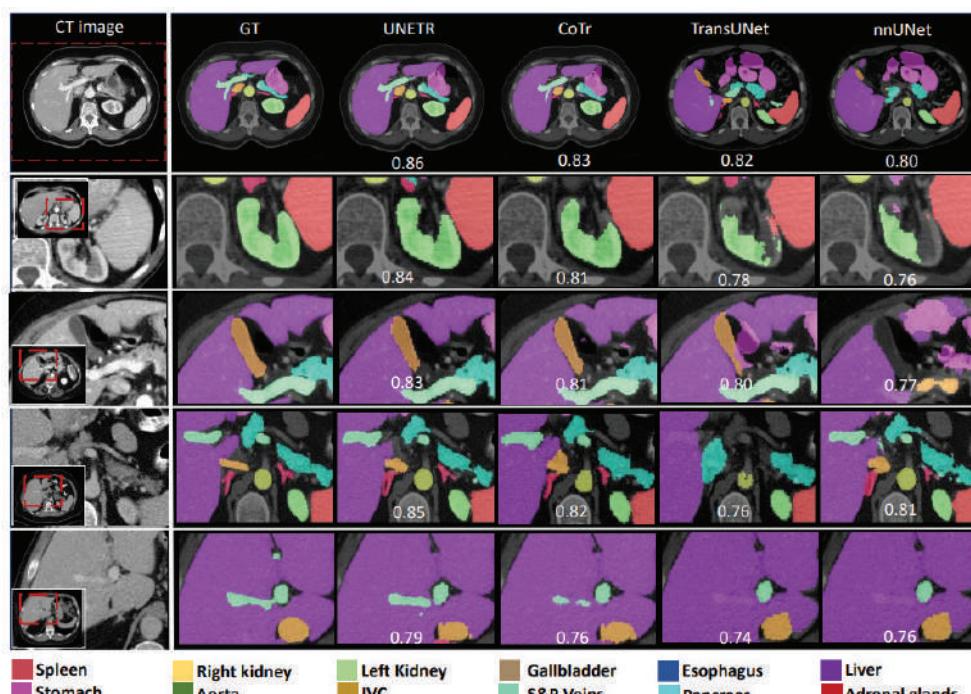
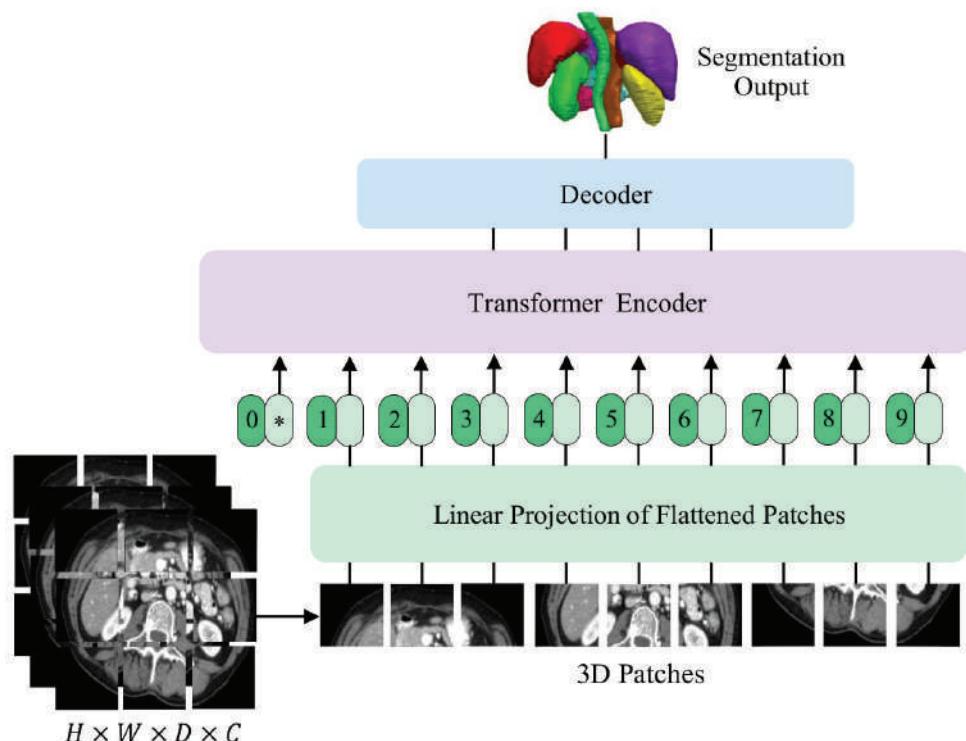
Novel proposed methodologies that leverage transformer-based or hybrid (CNN+transformer) approaches have demonstrated promising results in medical image segmentation for different applications. In this article, we will deep dive into one such network architecture (UNETR) and will also evaluate other transformer based approaches in medical imaging (TransUNET & CoTr).

1. UNETR

NVIDIA researchers have proposed to leverage the power of transformers for volumetric (3D) medical image segmentation and introduce a novel architecture dubbed as UNEt TRansformers (UNETR). UNETR employs a pure vision transformer as the encoder to learn sequence representations of the input volume and effectively capture the global multi-scale information, while also following the successful U-shaped network design for the encoder and decoder.

Why UNETR: Although Convolutional Neural Networks (CNN)-based approaches have powerful representation learning capabilities, their performance in learning long-range dependencies is limited to their localized receptive fields. As a result, such a deficiency in capturing multi-scale contextual information leads to sub-optimal segmentation of structures with various shapes and scales.

Overview of UNETR. Our proposed model consists of a transformer encoder that directly utilizes 3D patches and is connected to a CNN-based decoder via skip connection.



Qualitative comparison of different baselines.

UNETR has a significantly better segmentation accuracy for left and right adrenal glands, and UNETR is the only model to correctly detect branches of the adrenal glands.

UNETR proposes to use a patch-based approach with a transformer-based encoder to increase the model's capability for learning long-range dependencies and effectively capturing global contextual representation at multiple scales. For instance, in the multi-organ segmentation task, UNETR can accurately segment organs with complex shapes (e.g. adrenal glands) and low contrast (e.g. portal veins)

while CNN-based approaches fail to accurately segment these organs. See the figure above for more qualitative comparisons between UNETR and other CNN-based and transformer-based segmentation models.

UNETR has shown promising performance on various volumetric medical image segmentation tasks such as multi-organ segmentation using Multi Atlas Labeling Beyond The

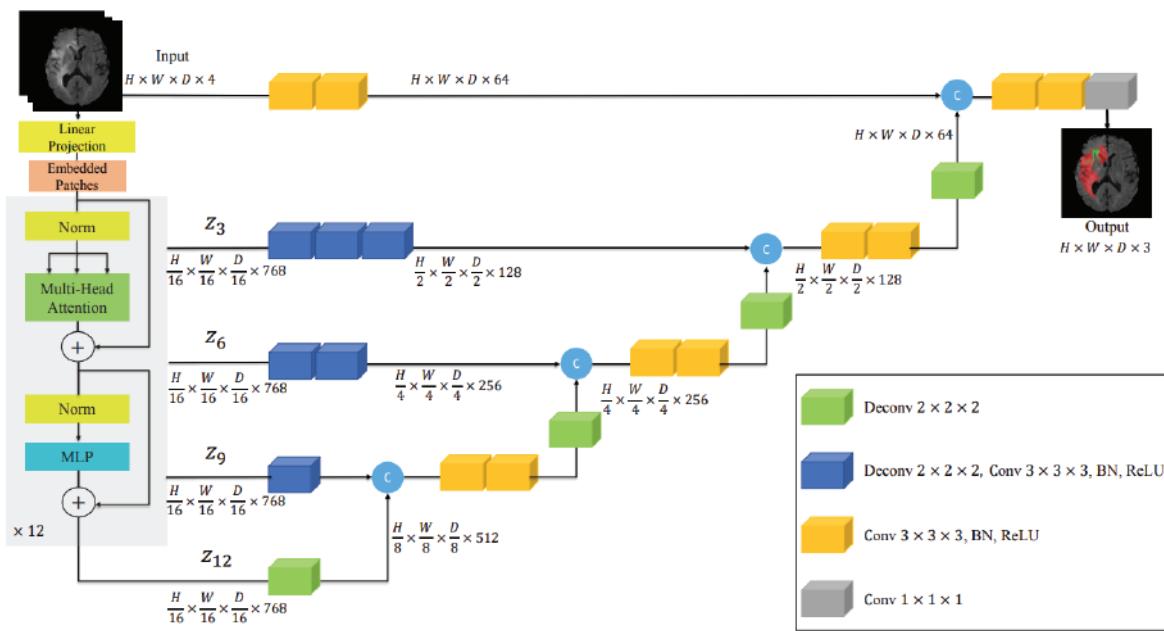
56 Medical Imaging Technology

Cranial Vault (BTCV) dataset and spleen and brain tumor segmentation using Medical Segmentation Decathlon (MSD) dataset. On the BTCV dataset, UNETR is currently the state-of-the-art methodology on both Standard (only training with challenge data) and Free Competition (training with additional data) public [leaderboards](#).

In addition, UNETR so far has shown to be more efficient in comparison to other transformer-based models (e.g. TransUNet) and CNN-based baselines in terms of number of FLOPs and inference time. See Table 1 for comparison of number of parameters, FLOPs and averaged inference time for various models in BTCV experiments.

Comparison of number of parameters, FLOPs and averaged inference time for various models in BTCV using a sliding window approach.

Models	#Params (M)	FLOPs (G)	Inference Time (s)
nnUNet [21]	19.07	412.65	10.28
CoTr [48]	46.51	399.21	19.21
TransUNet [7]	96.07	48.34	26.97
ASPP [11]	47.92	44.87	25.47
SETR [53]	86.03	43.49	24.86
UNETR	92.58	41.19	12.08



Overview of UNETR architecture. A 3D input volume is divided into a sequence of uniform non-overlapping patches and projected into an embedding space using a linear layer. The sequence is added with a position embedding and used as an input to a transformer model.

The encoded representations of different layers in the transformer are extracted and merged with a decoder via skip connections to predict the final segmentation.

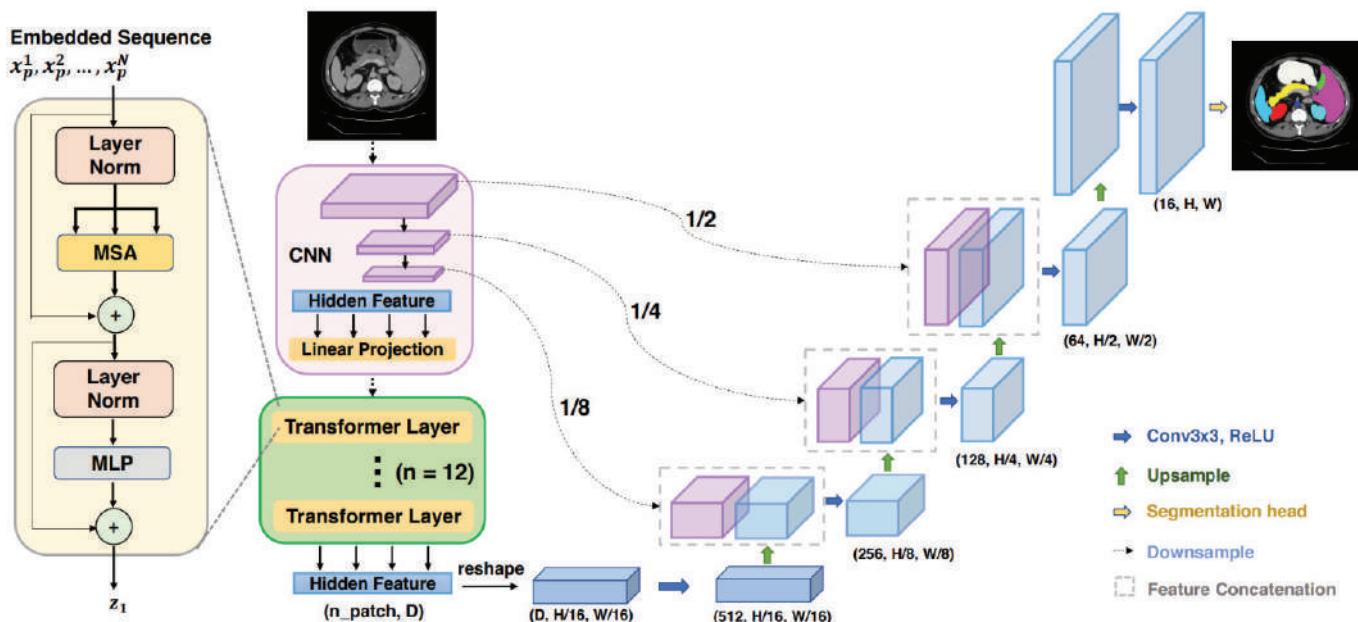
In the spirit of open innovation and to accelerate the research in this emerging field, NVIDIA has open-sourced UNETR via [MONAI Github public repository](#). In addition, a standalone UNETR repository is available in [MONAI research contributions repository](#). Furthermore, two UNETR tutorials ([pure MONAI](#) and [MONAI + PyTorch Lightning](#)) for multi-organ segmentation using BTCV datasets are available on [MONAI tutorials](#) for researchers to further explore this methodology in practice.

Two notable approaches that have leveraged transformers for medical image segmentation are TransUNet and CoTr. These approaches will be discussed in detail in the following sections.

2. TransUNet

TransUNet is a 2D hybrid CNN-Transformer segmentation model that leverages a vision transformer (ViT) as a standalone layer into the encoder of UNet architecture. Specifically, TransUNet uses a CNN as a feature extractor to generate feature maps as input of the ViT model in the bottleneck of the architecture. The ViT model uses self-attention layers to effectively process the extracted feature maps that are fed into the decoder for computing the final segmentation output. TransUNet has achieved comparable performance on the tasks of multi-organ segmentation using BTCV dataset as well as Automated Cardiac Diagnosis Challenge (ACDC) for automated cardiac segmentation.

Here is the [paper](#) explaining the architecture and the approach in further details, while the code and models are available [here](#).



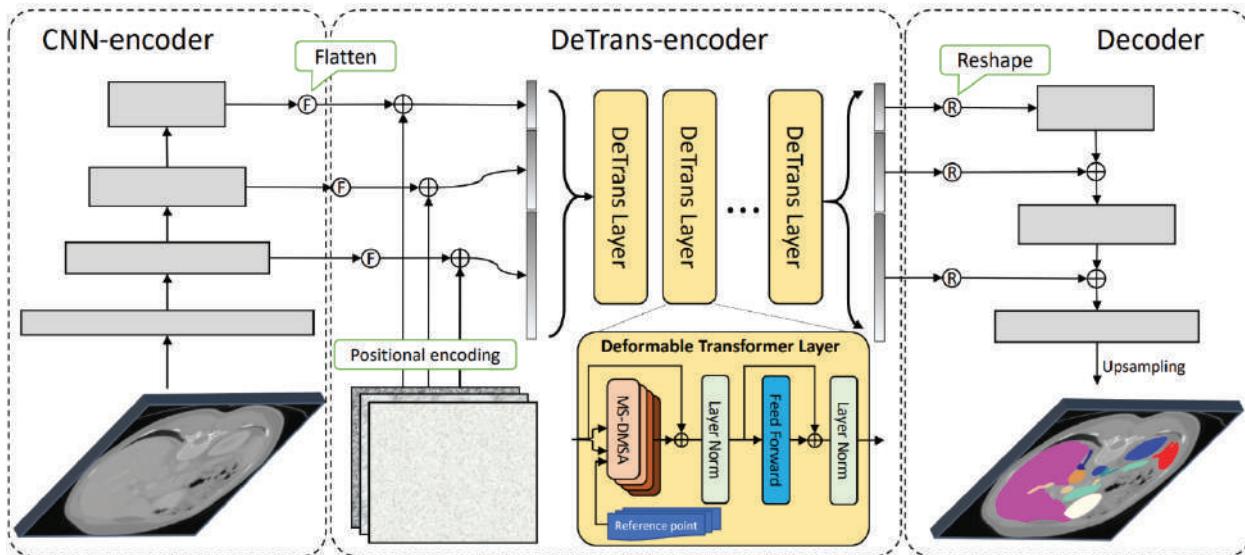
Overview of TransUNet architecture and schematic of the transformer layer.

CoTr: CoTr proposes a 3D framework that efficiently bridges CNNs with transformers for medical image segmentation. For this purpose, it introduces a deformable Transformer (DeTrans) to capture long-range dependencies in the extracted feature maps. The deformable self-attention mechanism in DeTrans allows for selectively paying more attention to a small set of key positions in extracted image embeddings. CoTr was tested and trained on BTCV multi-organ segmentation dataset and achieved competitive performance in this task.

Here is the [paper](#) explaining the details of this approach and the code is available at <https://github.com/YtongXie/CoTr>

Technical Differences: While all 3 approaches explore the potential application of using Transformer based networks for medical image segmentation. There are key differences between them. Unlike **TransUNet** which is a 2D segmentation

58 Medical Imaging Technology



Overview of CoTr architecture. It consists of CNN and DeTrans encoders as well as a decoder. Multi-scale features are extracted from the CNN encoder, projected to embeddings and processed in DeTrans encoder to capture long range dependencies. The decoder processes features from the DeTrans encoder to compute the final segmentation output.

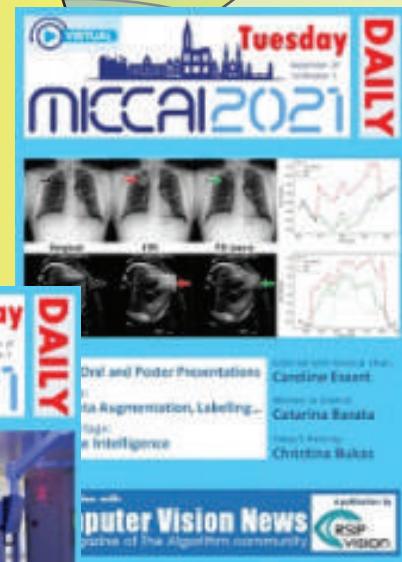
model, CoTr and UNETR utilize volumetric inputs and hence can benefit from the spatial context of data. UNETR and TransUNet both use the transformer layers of the ViT model whereas CoTr leverages a deformable transformer layer that narrows down the self-attention to a small set of key positions in an input sequence. In addition, each of these models utilize the transformer layers differently in their architecture. TransUNet uses the transformer layers in the bottleneck of a UNet, while CoTr utilizes them in between the CNN encoder and decoder by connecting them in different scales via skip connections. On the other hand, UNETR uses the transformer layers as the encoder of a U-shaped architecture and generates input sequences by directly utilizing the tokenized patches. The transformer layers of UNETR are connected to a CNN decoder via skip connections in multiple scales.

Conclusion

Convolutional neural networks (CNNs) have been the de facto standard for 3D medical image segmentation so far. However, Transformers have the potential to bring a fundamental paradigm shift with their strong innate self-attention mechanisms and hold the potential to serve as strong encoders for medical image segmentation tasks. The pre-trained embedding can then be adapted for various down-stream tasks (example, segmentation, classification & detection). In the years to come, we will see new breakthroughs powered by Transformers for medical imaging - the future is exciting, so we should brace ourselves.

MICCAI 2021

**Did you miss out
on MICCAI 2021?**



**No worries!
We got you covered!**

**Keep in contact
with the
community!**

60 Startup Village



MICCAI Startup Village Award 2021

Deepcell is a life science company offering technology at the intersection of microfluidics, genomics, and AI machine learning.

They have been awarded the winner prize at the MICCAI 2021 Startup village. Once again, we chose well when we decided to feature them, before the award: co-founder and CTO **Mahyar Salek** told us much more about them.

Cell morphology, which is the study of the shape, structure, form, and size of cells, is the backbone of cell research being performed in laboratories around the world, typically utilizing traditional microscopes and other devices to view the cells of a given sample.

Deepcell, which spun out of **Stanford University** in 2017, is bringing innovation that could **reshape the way we think about cells**, with a novel imaging and sorting platform powered by microfluidics and supported by machine learning. It has developed a way to automatically analyze and understand cells based on how they look and created hardware that not just analyses these cells but also acts on them. It is developing a **highly advanced camera with an artificial brain** that can identify, annotate, sort, enrich, and preserve single cells.

The way this works is that a sample from a cancer patient, for example, is fed into the device, then images are taken which are fed into a machine learning brain, and these **images are classified in real time to identify which cells are cancerous and which are not**. After that, it isolates those cancer cells and applies deep sophisticated visual analysis. Based on that analysis, one could access viable cells to perform a range of interesting biological experiments.

"This is something that's never been possible before Deepcell technology," Mahyar tells us. *"Imagine you are trying to figure out which drug is effective for your patient. Today, people try out different drugs, which may or may not work, and have to deal with all the related side effects and consequences that entails. I don't think that should have a place in 21st century medicine. Imagine instead, you have a sample from your patient, and you try different drugs on the actual cells. You could isolate different reactions and morphologies to make a connection between how the cells look and how effective the drug is on a patient before they take it."*

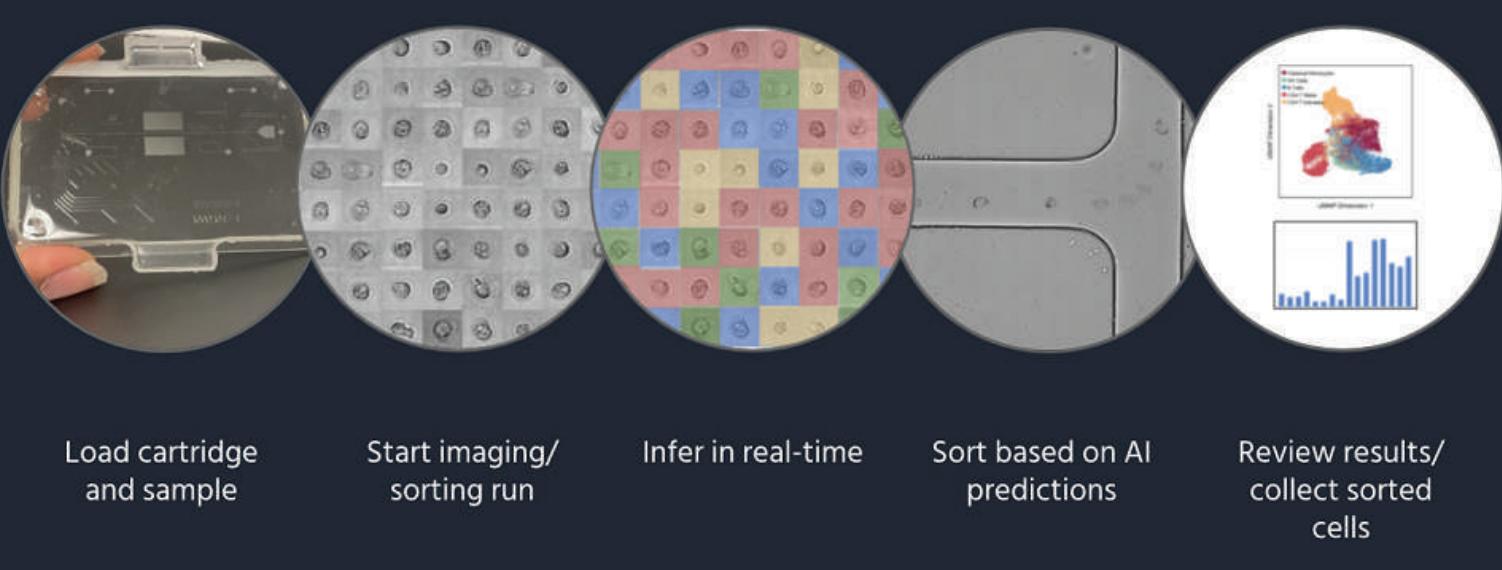
In a way, this is bridging the gap between the centuries-old practice of looking at cells under the microscope, and the modern wave of biology that wants to really **understand the molecular makeup of our cells**. Biology and life sciences are rapidly becoming data science and engineering fields. They have always been about trial and error – you observe something, you try something, you learn something – but data science is allowing a lot more control over how we develop new medicine and new insights.

"Molecular understanding of cells has seen a revolution in the last few years and advancements in machine learning, hardware, and data have all paved the way for technologies like deep learning to flourish," Mahyar points out. *"We have replaced human intelligence with what some would argue was much more like a super intelligence."*

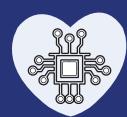
Deepcell has the **largest dataset of high-resolution single cell images anywhere in the world**. With such powerful data, it wants to reach an audience that truly understands the value of applying machine learning techniques to medical imaging.

If all this is piquing your interest, you will be pleased to hear that Deepcell is hiring! They have many open positions in AI, machine learning, and data science, so make sure that you pay them a visit.

"I want to start a conversation with the brightest minds out there," Mahyar says. *"No single company or genius in the world can solve the kinds of problems that we want to solve here at Deepcell. We need a whole community behind us and that's why we're reaching for the best people out there. We want people with the highest skillsets as well as the highest ethics to join this effort and to help us build the next revolution in life sciences and data science. If you are interested in applying big data skillsets to innovate in this area, then we have the most exciting dataset to offer you!"*



62 Congrats, Doctor!



MEDICAL
IMAGING
NEWS

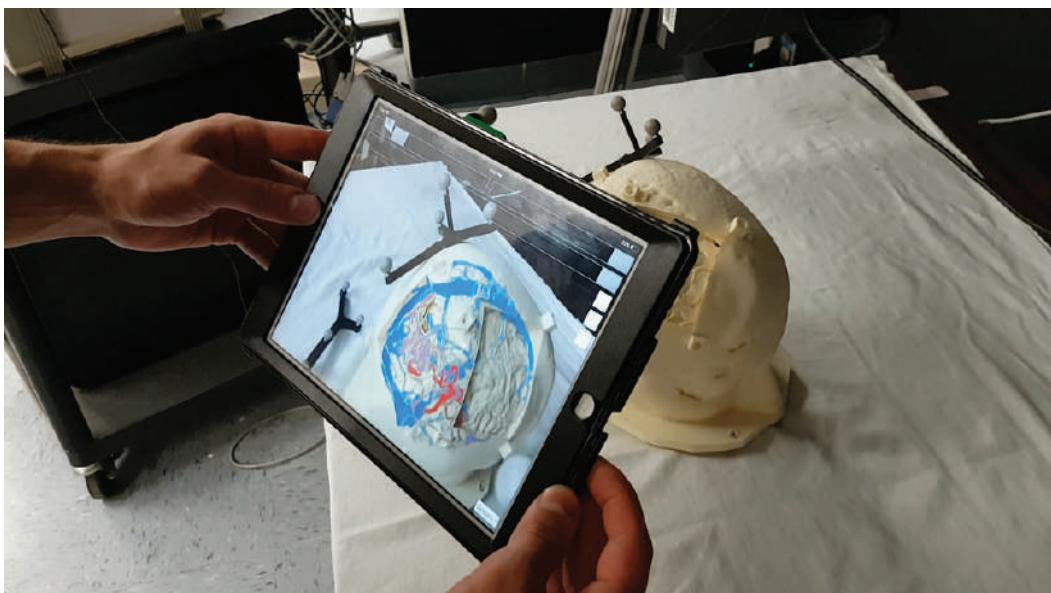


Etienne Léger recently completed his PhD at Concordia University under the supervision of Marta Kersten-Oertel. His research interest lies in Human Computer Interaction, evaluating how new methods can potentially improve neurosurgical workflows. His research focused on developing and assessing neurosurgical guidance tools making use of novel paradigms, methods and hardware to make it more intuitive and interactive. He believes that through new hardware integration, neurosurgical guidance can be made more accessible, which can lead to improved patient outcomes.

It is estimated that 13.8 million patients per year require neurosurgical interventions worldwide, be it for a cerebrovascular disease, stroke, tumour resection, or epilepsy treatment, among others. These procedures involve navigating through and around complex anatomy in an organ where damage to eloquent healthy tissue must be minimized. Neurosurgery thus has very specific constraints compared to most other domains of surgical care. These constraints have made neurosurgery particularly suitable for integrating new technologies. Any new method that has the potential to improve surgical outcomes is worth pursuing, as it has the potential to not only save and prolong lives of patients, but also increase the quality of life post-treatment.

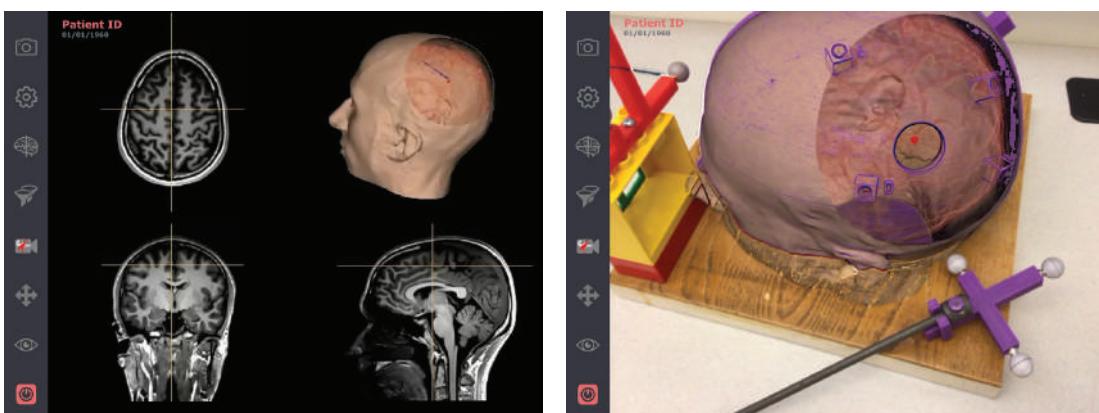
In his work, Étienne developed novel neurosurgical image-guidance methods, making use of currently available, low-cost off-the-shelf components. In particular, a mobile device (e.g. smartphone or tablet) is integrated into a neuronavigation framework to explore new augmented reality visualization paradigms and novel intuitive interaction methods. The developed system, called MARIN for Mobile Augmented Reality Interactive Neuronavigator, aims at improving image-guidance using augmented reality to improve intuitiveness and ease of use. Further, gestures on the mobile device are used to increase interactivity with the neuronavigation system. These touchscreen interactions enable partially mitigating the problem of accuracy loss or brain shift that occurs during surgery. It also gives the control over the visualization back to the surgeon, enabling them to switch between different visualization methods, be it traditional cut-planes guidance, virtual 3D models extracted from the preoperative scan or a virtual reality view where segmented structures are overlaid into the surgical field in real-time. The AR view is also customizable: structures (e.g. vessels, cortex surface, etc.) can be individually added or removed from the view and the augmentation can be limited to only

parts of the operating field through a movable AR window for instance. All of these customizations can be done directly on the device using touchscreen interactions and enable real-time control over the guidance for the surgeon.



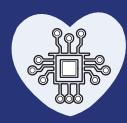
Our developed mobile AR system displaying CTA acquired vessels (virtual anatomical data) over a phantom head (the real world as captured by the iPad).

The results of his work show the feasibility of using mobile devices to improve neurosurgical processes. Augmented reality enables surgeons to focus on the surgical field while getting intuitive guidance information. Mobile devices also allow for easy interaction with the neuronavigation system thus enabling surgeons to directly interact with systems in the operating room to improve accuracy and streamline procedures. To encourage further research and accelerate the pace of innovation, Étienne released the developed application under an open source license, making it accessible to others to reuse and keep improving upon.



On the left: screenshot of the system when used in standard IGNS mode.

On the right: screenshot of the system when used in augmented reality mode.



Computer Vision for Automated Medical Diagnosis

Yuyin Zhou is a postdoctoral researcher at Stanford University, working with Matthew Lungren and Lei Xing on medical image analysis and other related machine learning problems. She is also co-organizing the very first Computer Vision for Automated Medical Diagnosis workshop at this year's ICCV. With the conference only a few weeks away, Yuyin is here to tell us what to expect.

Machine learning has been a helpful tool for doctors in dealing with different medical imaging problems for some time now. It can also support disease diagnosis and treatment planning. Over the past few years, there has been a great deal of progress made in this area because of huge advancements in computer vision and artificial intelligence techniques. Problems such as **medical image registration**, **structure detection**, and **tissue and organ segmentation** have achieved state-of-the-art performance,



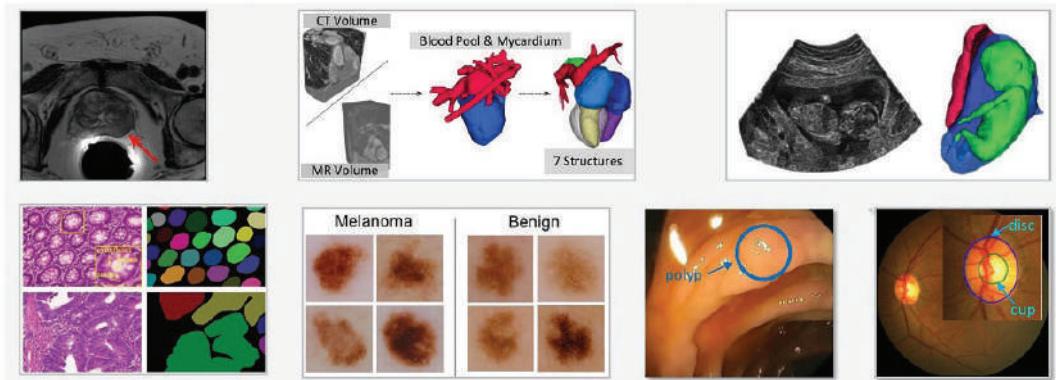
Yuyin Zhou

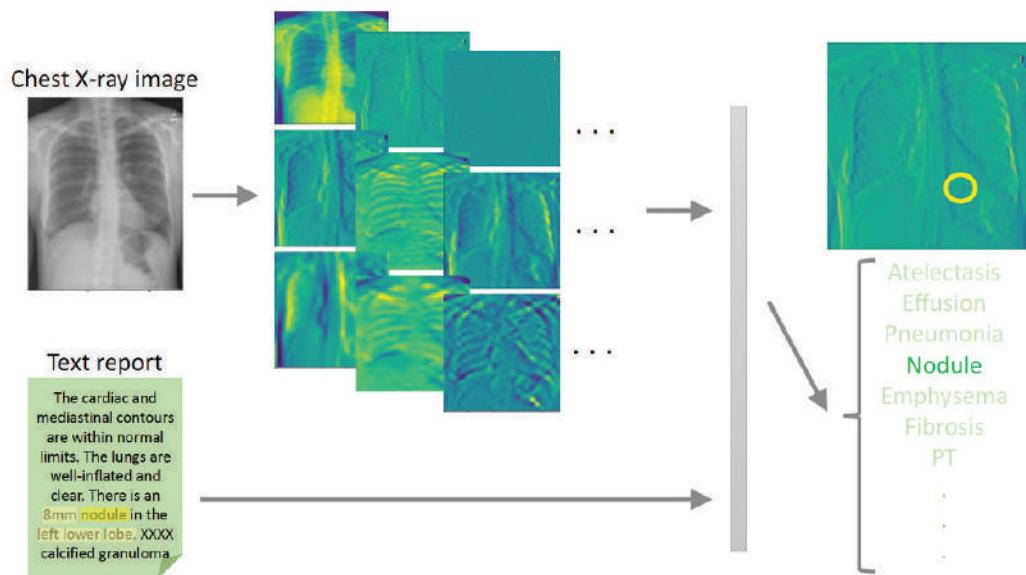
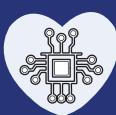
while **many new medical devices have been developed** in conjunction with industry.

In spite of this, the **safe and reliable adoption of such technologies** in hospitals in the real world remains limited, and many problems, such as cancer diagnosis, are still not solved.

"This is the key reason why we have created this workshop," Yuyin reveals.

Computer vision for different medical diagnosis problems





"We want to bring researchers in the computer vision, machine learning, healthcare, and clinical fields together to discuss the current progress and the related challenges which are yet to be addressed. What are the next steps? What possible solutions should we be looking for? Are there any important research directions in the field that we haven't explored yet?"

Their first meeting at ICCV later this month already has a stellar line-up of speakers on board, including **Russell Taylor** from **Johns Hopkins University**, who will be discussing Autonomy and Semi-Autonomous Behavior in Surgical Robot Systems.

"I think this topic is one of the most important and has not been addressed enough in the computer vision community," Yuyin tells us.

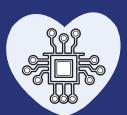
"We've also got Demetri Terzopoulos,

*a distinguished professor at UCLA who is heavily involved in the medical industry. He is going to talk about how to bridge the gap between medical research, computer vision, and industry. We are also going to feature my advisor, **Matthew Lungren**, an Associate Professor of Radiology at **Stanford**, who will speak from a doctor's perspective."*

Another benefit of this module is that it does not require long run-time, allowing fast and easy implementation in the clinic.

Other confirmed speakers include **Yizhou Yu** – How should machines analyze medical images to aid diagnosis? – and [Lena Maier-Hein](#) – Statistics meets machine learning in biomedical image analysis. There will also be 10 engaging oral talks from **UCLA**, **University of Oxford**, **Google**, and more.

66 ICCV Workshop Preview



MEDICAL
IMAGING
NEWS

This is not the first medical computer vision workshop – **CVPR** has had a similar event for a number of years now. The community recognizes and understands the importance of the topic, so the foundations have been laid for this new meeting at **ICCV** to be a success this year and for many more to come.

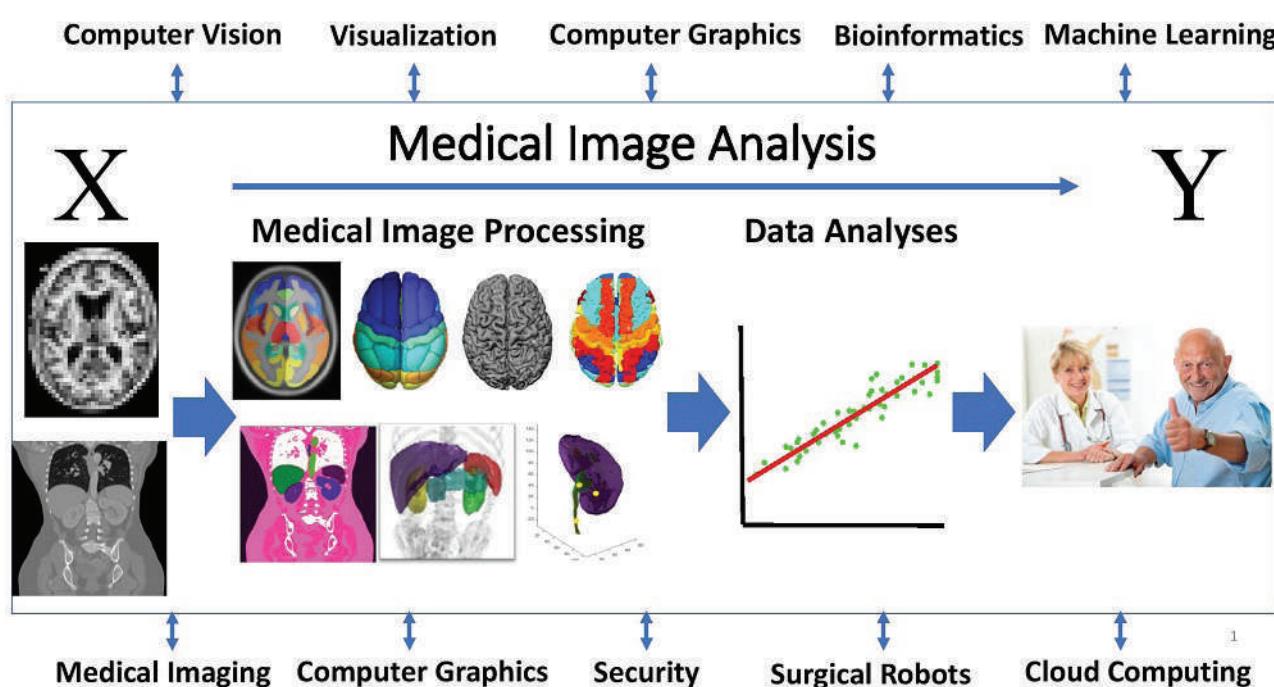
However, the team do not intend to create a carbon copy of another event. They plan to discuss topics which haven't been covered before. But with so much on the menu at **ICCV** this year, including another medical workshop, how should attendees choose between this one and everything else on offer?

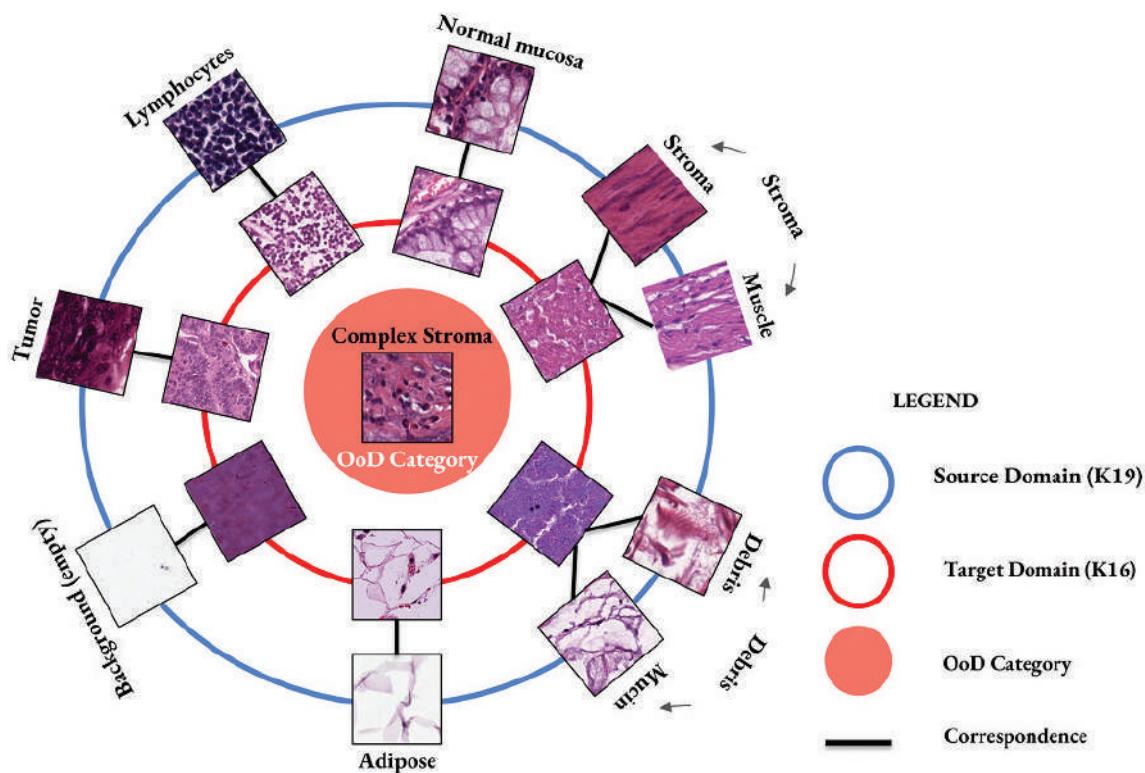
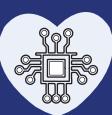
*"What really sets our workshop apart from others is that we are focusing on the **general challenges in the medical computer vision arena**," Yuyin points out.*

"We're going to give a more holistic and complete view of this field and we want to discuss things from broader perspectives—not just medical computer vision like CT and MRI, but also NLP, medical robots, surgical planning, and how to better adapt existing computer vision and machine learning expertise into all of these different medical problems."

The workshop is very close to Yuyin's own body of work. During her PhD career at Johns Hopkins University, the team have been working on the **Felix Project**, which aims to detect pancreatic cancer earlier.

*"We started from pancreas segmentation and went deeper into **pancreatic tumor segmentation and detection problems**," Yuyin explains.*





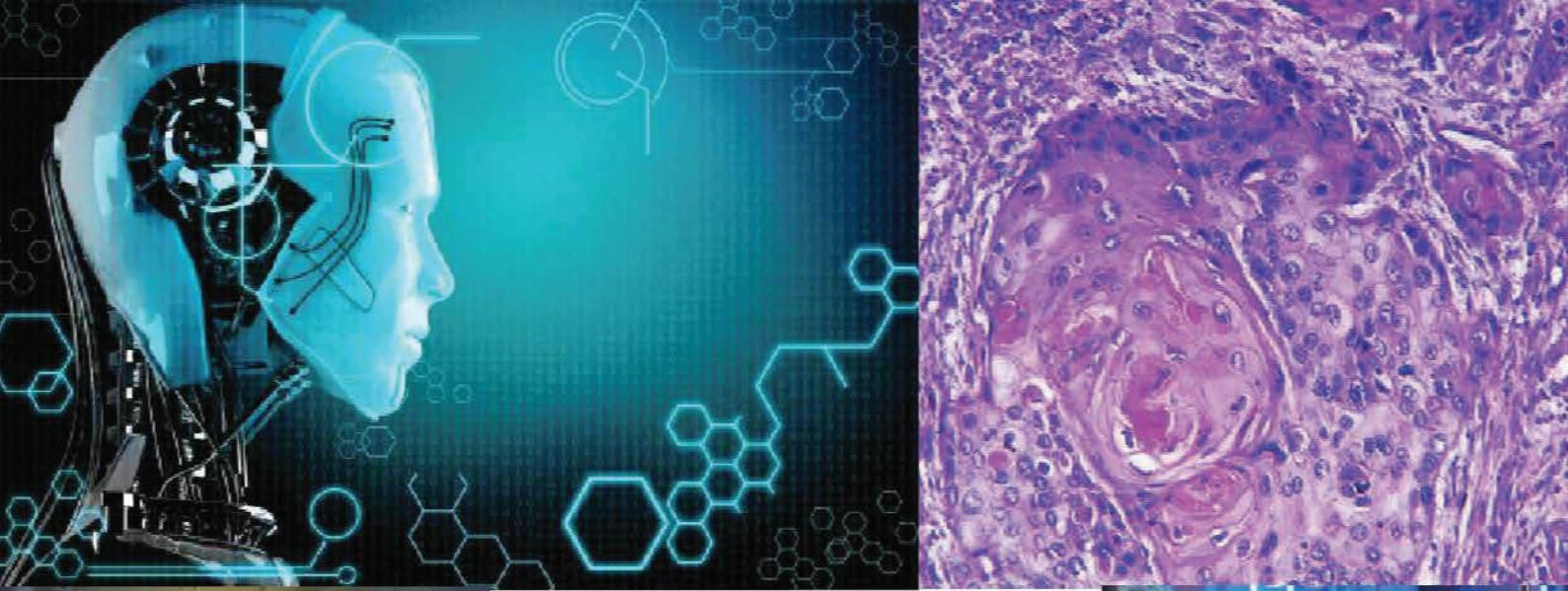
*"For my PhD research, I performed accurate segmentation and detection from different medical images, especially for smaller and more challenging subjects like tumors. I also began to study the issues of **generalization**, **transfer**, and **robustness** because there can be a huge gap when you transfer your technology from one hospital to another hospital, or one population group to another. This is why these existing technologies cannot be fully trusted to be applied to real-world clinical applications yet. A lot of my research focus has shifted to studying how to make these existing techniques more robust and have better transfer characteristics and better generalization to different problems."*

The team are currently finalizing the details for the ICCV workshop. Yuyin's

diverse band of co-organizers include Lequan Yu from the University of Hong Kong; Maithra Raghu from Google Brain; Qi Dou from the Chinese University of Hong Kong; Yuankai Huo, Assistant Professor at Vanderbilt University; Holger Roth, a Senior Applied Research Scientist at NVIDIA; and many other professors and senior scholars from different affiliations.

"We hope that everybody can join our workshop," Yuyin adds, finally.

"We have such an amazing line-up of speakers and people covering different topics. Please take part in the discussions and ask lots of questions. You will definitely learn something new. I guarantee it!"



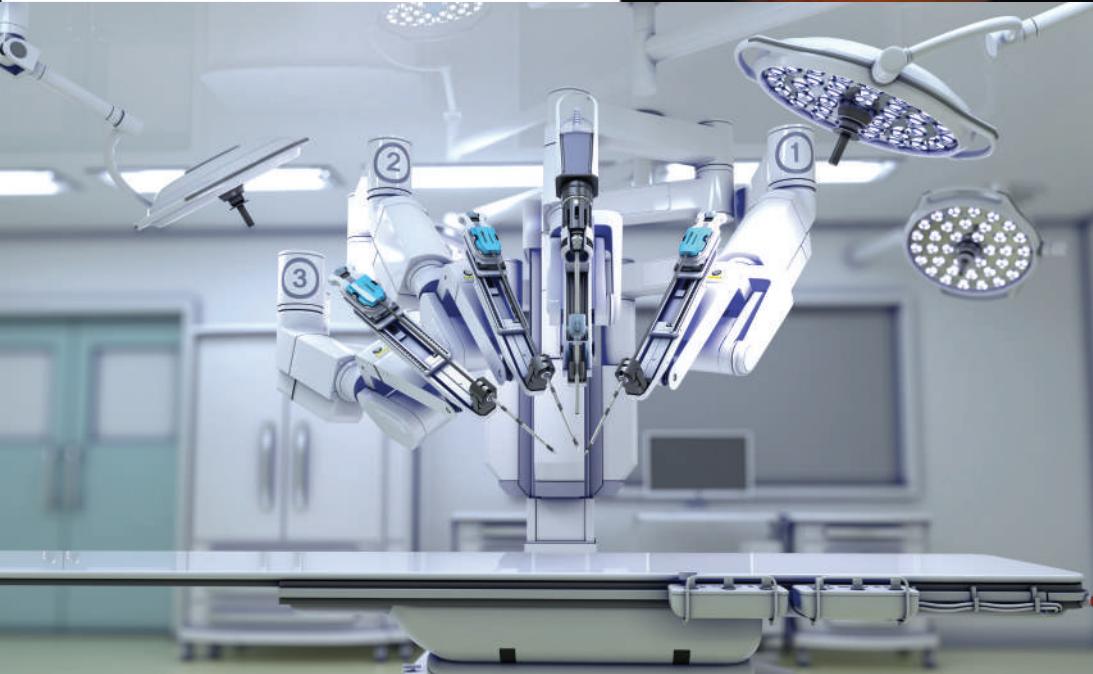
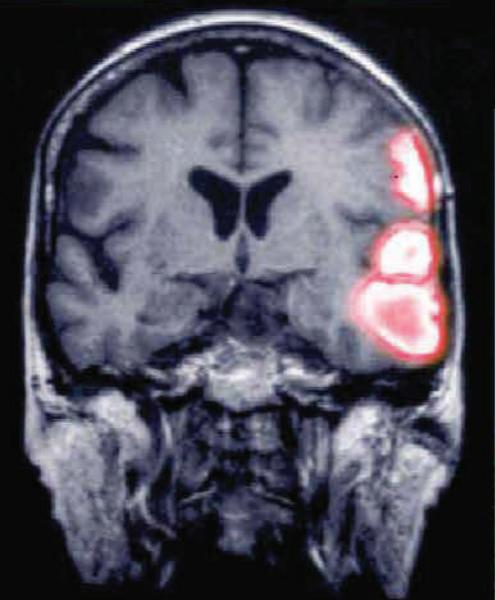
IMPROVE YOUR VISION WITH **Computer Vision** News

SUBSCRIBE

CLICK HERE, IT'S FREE

To the magazine of the
algorithm community

and get also the **New**
supplement Medical
Imaging News!



A PUBLICATION BY

