Computer Vision News

The magazine of the algorithm community

A publication by RSIP VISION

June 2018

Spotlight News

Upcoming Events

Review of Research Papers:
Polygon-RNN++
Deep Learning CNN for Microscopy

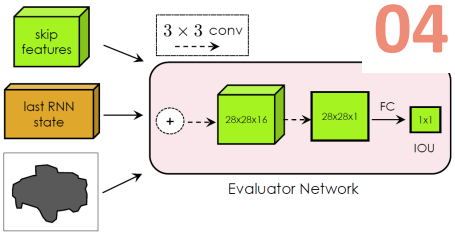
Woman in Science:
Marta Kersten-Oertel

Guest:
Jeff Clune - UBER

Application:
Altia Systems

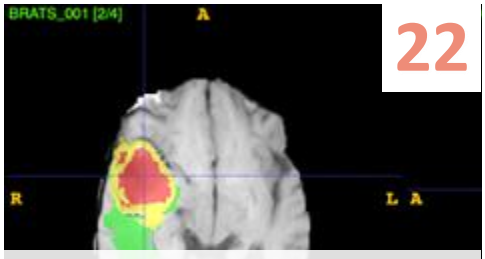
Challenge:
Medical Decathlon

Contents



04

Research
Polygon-RNN++

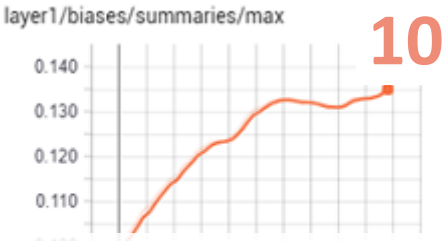


22

Challenge - Medical Decathlon
with Michela Antonelli

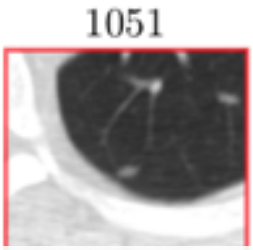
30

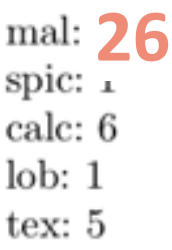
Woman in Science
Marta Kersten-Oertel



10

Focus On: Deep Learning
in TensorFlow and Keras





26

Medical Imaging Research:
RSIP Vision and Tufts

34

Application
PanaCast by Altia Systems

18

Guest: Jeff Clune - UBER

29

Spotlight News

38

100 days to MICCAI
with Alejandro Frangi

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editor:
Assaf Spanier

Publisher:
RSIP Vision

Contact us

Give us feedback

Free subscription

Read previous magazines

Follow us:

Dear reader,

CVPR is here again, beginning on June 18 in Salt Lake City, Utah. Besides the exceptionally rich technical program, participants will once again receive every day the official conference magazine **CVPR Daily**, a joint project of **CVPR with RSIP Vision**. The novelty of this year is that also those who won't have the chance to attend the conference can receive the CVPR Daily in their mailbox. All they have to do is to **fill this short form**: feel at CVPR, without being at CVPR…

Another news item which I wish to highlight is the **Medical Decathlon**, a very ambitious challenge aiming at finding the algorithm that can segment every organ and pathology using different types of images. You will not be surprised to learn that **NVIDIA**, **RSIP Vision** and **DeepMind** are sponsoring this challenge. Read about it at page 22.

Those of our readers who are interested in **Computer Assisted Radiology and Surgery** will enjoy reading the **CARS Preview**, published by RSIP Vision in view of **CARS 2018** in Berlin.

Last but not least, you can meet our consultants also at the **RE•WORK's Deep Learning in Robotics Summit** in S. Francisco, sponsored by RSIP Vision. Read what **Jeff Clune of Uber** says about it at pages 20-21.

Enjoy the reading and, as always, take us along for your next Deep Learning project!

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

Research: Polygon-RNN++

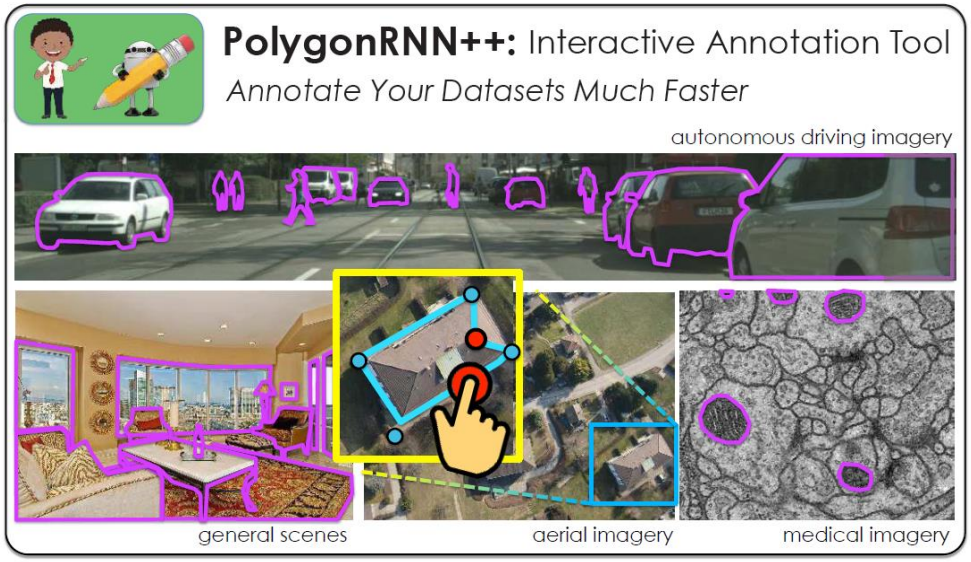
by Assaf Spanier

"A model for object instance segmentation that can be used to interactively annotate segmentation datasets."

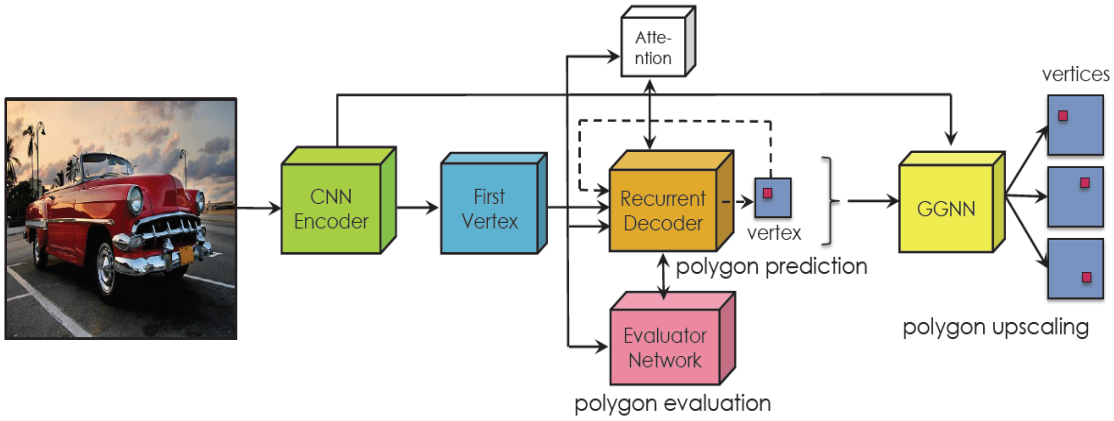
Every month, Computer Vision News reviews a research paper from our field. This month we have chosen to review a new improved version of Polygon-RNN (presented at CVPR 2017): **Efficient Annotation of Segmentation Datasets with Polygon-RNN++**. We are indebted to Sanja Fidler and her team (**David Acuna**, **Huan Ling** and **Amlan Kar**), for allowing us to use images from the paper to illustrate this review. Their article is here and their code is here. This work has been selected for CVPR 2018, where it will presented during the poster session of Tuesday at 10:10am.

Introduction:

Manual labeling of images with object masks is tedious and almost prohibitively time-consuming. The authors propose their **Polygon-RNN++**, a new improved version of Polygon-RNN (presented at CVPR-2017), to help produce polygonal annotations of objects. Polygon-RNN++ produces a polygon by **serially iterating vertex by vertex**. This serial vertex prediction method of the model allows easily incorporating a human in the loop. The user can propose a vertex correction, which is then fed back into the model, which will use it to re-predict more correctly all other vertices. The model includes several important improvements: 1) new CNN encoder architecture, 2) Reinforcement Learning paradigm, and 3) use of a Graph Neural Network for high resolution results. The authors evaluated Polygon-RNN++ on the **Cityscapes** dataset and showed that it outperforms the Polygon-RNN, both for automatic evaluation (10% improvement) and interactive evaluation (requiring 50% fewer clicks by annotators). Moreover, Polygon-RNN++ shows powerful generalization capabilities and improvements over existing pixel-wise methods.



Methods:



First, we'll give an overview of the network's structure and then go into further detail about each component:

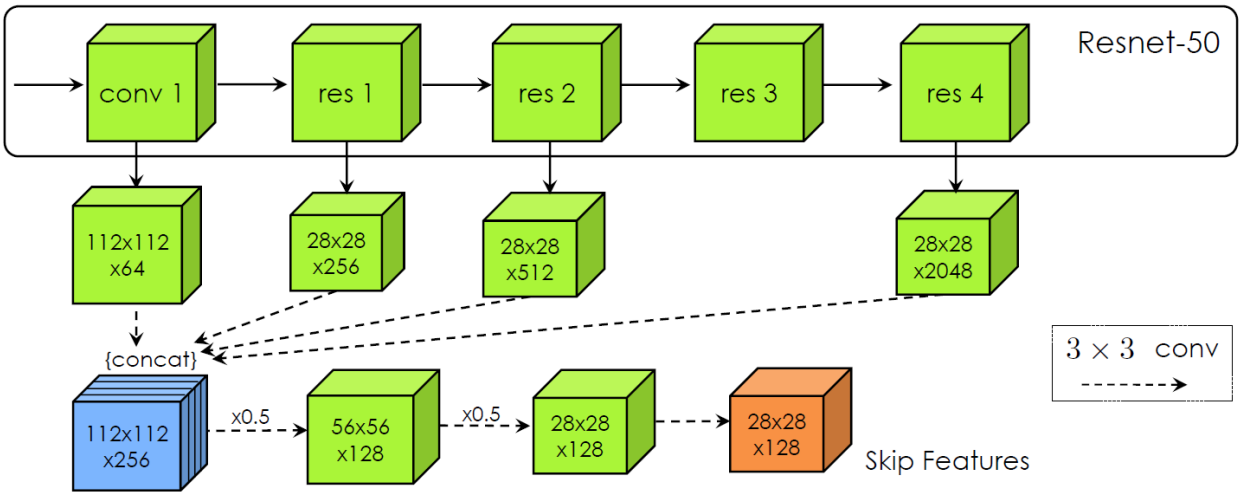
1. A CNN encoder [in green in the figure below] is used to extract features from the image cropped around the user-marked bounding box.
2. Next, another network [in blue] gets the features produced as input to predict the first vertex, from which construction of the polygon for the object will start.
3. The image features and first vertex are the input for an RNN Decoder [orange]. Each iteration of the RNN produces the next vertex of the polygon [marked by a red square position within a gray box]; the iterations are represented by a broken line.
4. The RNN decoder network includes a Visual Attention mechanism [white], which uses weights to focus the RNN decoder on a certain area in which to search for the next vertex .
5. An Evaluator Network [pink] gets a set of candidate polygons proposed by the RNN decoder as input and selects the best polygon from among them.
6. Finally, a network called GGNN (gated graph neural network) [in yellow], works at a higher resolution to refine the polygon produced, by adding vertices and adjusting the overall polygon.

Now, let's look more closely at each component of the network:

1. First, let's describe the structure of the CNN Encoder [green]:

The Encoder network is based on a ResNet-50 architecture, with the following modifications: (1) Reducing the stride of the network and introducing dilation factors. (2) The original average pooling and FC layers were removed. (3) A skip-layer architecture was added to certain convolutional layers in the network, and all skip layer outputs are concatenated (the skip layers capture both low-level features such as edges and corners and high-level semantic features). (4) Finally, a combination of conv layers and max-pooling operations was used to obtain the final feature map. In the figure below the 112x112 blue tensor is fed directly to

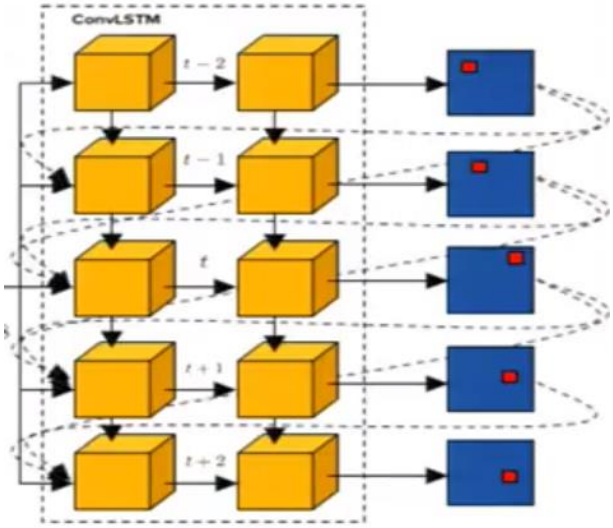
the GGNN, while the 28x28 orange tensor is the input for the First Vertex network.



2. First Vertex:

A separate network is used to predict the first vertex. It adds two DxD layers to the 28x28 orange tensor (in the above figure). The first layer predicts edges, while the second predicts the vertices of the polygon. The first vertex is sampled from this vertices-predicting final layer.

3. RNN Decoder



Two-layer ConvLTSM was used with 64 kernels in the first layer and 16 in the second, each with a kernel size 3x3. the output at each time step t is a matrix of size DxD of zeros and ones, where 1 indicates a vertex and 0 otherwise. D is the resolution used by the system for rough polygon prediction (the authors used D=28). When the polygon is closed an end-of-seq token is signaled.

Network loss: Polygon prediction is formulated as a reinforcement learning problem. The policy (denoted by p) for selecting the next vertex vt is computed by maximizing the reward r as the IoU between the mask enclosed by the generated polygon and the ground-truth mask m. To maximize the expected reward, our loss function becomes $L(\phi) = -E_{v^s \sim p_\theta}[r(v^s, m)]$ where

$v^s = (v_1^s, \ldots, v_T^s)$ and v_T^s is the vertex sampled from the model at time t, and $r = IoU(mask(v^s), m$

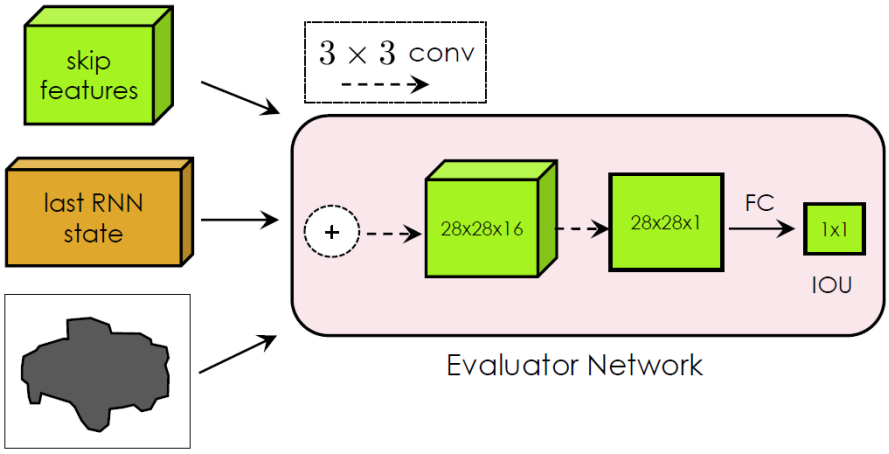
4. Visual Attention Mechanism

Previous RNN hidden state image feature map is used by the RNN to focus only on the relevant information in the next time step. For this, at time step t, the following weighted feature map is computed:

$$\alpha_t = softmax(f_{att}(s, f_1(h_{1,t-1}), f_2(h_{2,t-1})))$$
$$F_t = x \circ \alpha_t$$

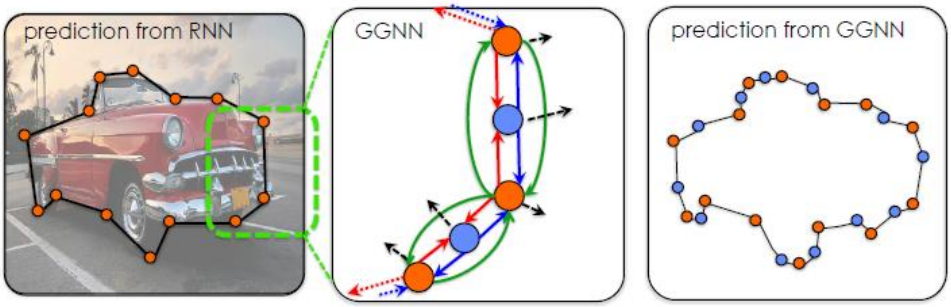
where $\circ$ is the Hadamard product, x is the skip feature tensor, $h_{1,t}$ and $h_{2,t}$ are the hidden state tensors from the two-layers ConvLSTM.

5. Evaluator Network



An evaluator network chooses among a list of K=5 candidate polygons. This network takes as input the skip features (112x112 blue tensor in CNN Encoder figure), the last state tensor of the ConvLSTM, and the predicted polygon, and tries to estimate its quality. In training, we minimize the mean squared error $(\phi) = [p(\phi, v^s) - IoU(m_{v^s}, m)]^2$, p is the network predicted IoU, m_{v^s} is the mask for the sampled vertices and m is the ground-truth mask.

6. Gated Graph Neural Network -- GGNN



The GGNN network element is used in order to generate refined polygons at a much higher resolution. GGNN has been proven efficient for semantic segmentation, when used at pixel-level. The authors take the polygon predicted by the RNN Decoder (orange vertices in the left figure) and, at a higher resolution, add midpoints (in blue) between every pair of consecutive orange vertices. The GGNN uses three types of edges (red, blue, green) to arrive at improved predictions of the relative location for each of the nodes (vertices) -- the black dashed arrows (zoomed-in middle figure). Right figure is the high resolution polygon output by the GGNN.

Dataset and Evaluation:

The authors use the **Cityscapes dataset**, which to this date is one of the most comprehensive benchmarks for instance segmentation. It contains 2975 training, 500 validation and 1525 test images with 8 semantic classes. The ground-truth polygons are pre-processed according to depth-ordering, to obtain polygons for only the visible regions of each instance.

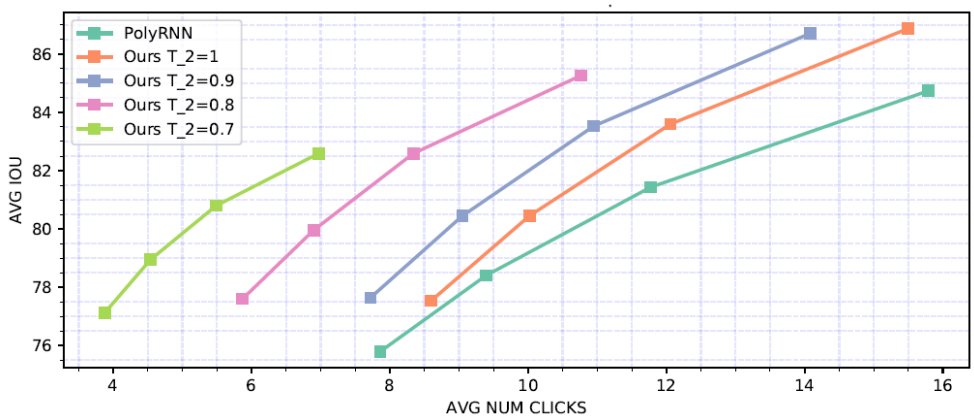
Two evaluation metrics were used: 1) Automatic Mode -- Intersection over Union (IoU) metric for evaluating the quality of the generated polygons and 2) Interactive Mode -- the average number of annotator clicks required to correct the model's predictions.

Results:

Model	Bicycle	Bus	Person	Train	Truck	Motorcycle	Car	Rider	**Mean**
Square Box	35.41	53.44	26.36	39.34	54.75	39.47	46.04	26.09	40.11
Dilation10	46.80	48.35	49.37	44.18	35.71	26.97	61.49	38.21	43.89
DeepMask	47.19	69.82	47.93	62.20	63.15	47.47	61.64	52.20	56.45
SharpMask	52.08	73.02	53.63	64.06	65.49	51.92	65.17	56.32	60.21
Polygon-RNN	52.13	69.53	63.94	53.74	68.03	52.07	71.17	60.58	61.40
Residual Polygon-RNN	54.86	69.56	67.05	50.20	66.80	55.37	70.05	63.40	62.16
+ Attention	56.47	73.57	68.15	53.31	74.08	57.34	75.13	65.42	65.43
+ RL	57.38	75.99	68.45	59.65	76.31	58.26	75.68	65.65	67.17
+ Evaluator Network	62.34	79.63	70.80	62.82	77.92	61.69	78.01	68.46	70.21
+ GGNN	**63.06**	**81.38**	**72.41**	**64.28**	**78.90**	**62.01**	**79.08**	**69.95**	**71.38**

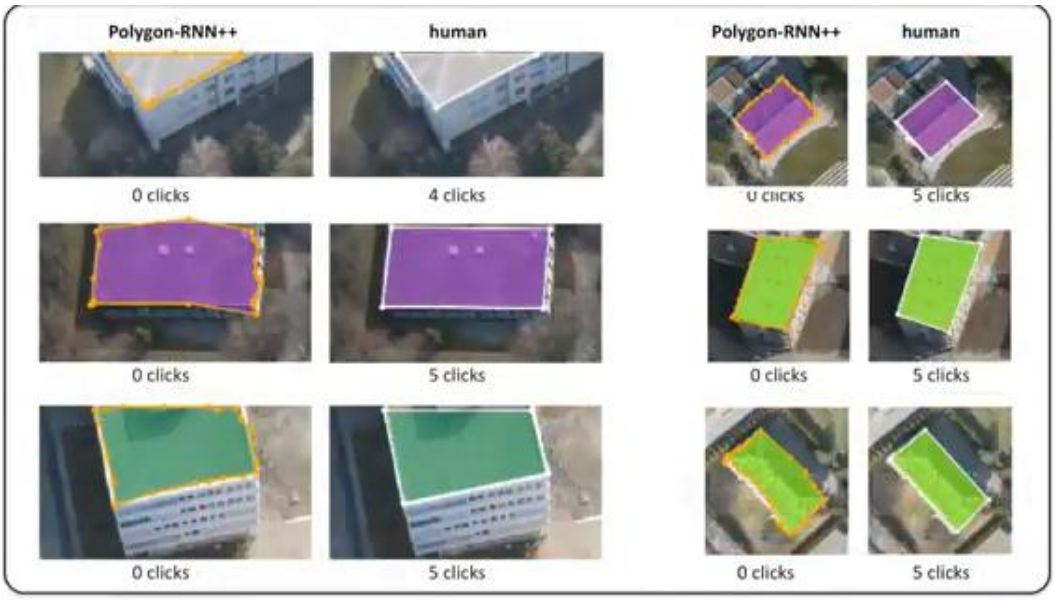
Automatic Mode Evaluation. Comparison of the model against: (1) SquareBox considers the provided bounding box as a prediction; (2) Dilation10, (3) DeepMask, (4) SharpMask, as well as (5) Polygon-RNN considered as state-of-the-art baselines; (6) Ablation study. The full model outperforms all other methods by almost 10% IoU; it also achieves best performance for each class.

The goal of **Interactive Mode Evaluation** is minimizing annotation time, while obtaining high quality annotations.



T_2 is the IoU between the predicted polygon and the GT mask, with polygons achieving agreement above T_2 considered to need no interference from the annotator.

In the results above are the average number of clicks per instance required to annotate all classes on the Cityscapes *val* set with different values of T_2. At T_2 = 0.8 the new model is still more accurate than Polygon-RNN at T_2 = 1.0. At T_2 = 0.7, it achieves over 80% IoU with an average of 5 clicks per object, over 50% reduction.



It generalizes to out-of-domain imagery: *aerial scenes*

Conclusions:

This paper proposed Polygon-RNN++, a model for object instance segmentation that can be used to interactively annotate segmentation datasets. The model builds on Polygon-RNN presented in CVPR-2017, while introducing important improvements to outperform the previous approach, in both automatic and interactive modes.

For the interested reader: in the original article the authors show (a) the model's robustness to noisy annotators and (b) a capability to generalize to novel domains.

Debug and analysis mechanisms for deep learning in TensorFlow and Keras

by Assaf Spanier

"This reduces the overwhelming number of neurons to a small set of groups, distilling the process and power of the neural network's deep learning."

The training of **deep learning neural networks** is a non-trivial process, requiring the update of dozens of parameters. In addition, the behavior of these networks is unpredictable. When writing "normal" code, like a program for sorting an array of numbers, we know the expected behavior of all the algorithms it is made up of and its expected output. In the case of deep learning, however, the behavior of the algorithm is determined by the data input into the network, not by deterministic code.

Tool

Beyond the non-deterministic nature of **deep learning models**, we must take into account that these models include tens of millions of parameters, each of which can affect network performance. Networks also have several crucial hyperparameters, which highly impact behavior and performance, such as number of iterations, number of layers, number of hidden units etc., making things quite challenging. When you train a **CNN network** and it achieves an error rate of 5%, it is nearly impossible to tell whether this performance is optimal without in-depth analysis. Yet, software tools and packages for thorough debugging and analysis of deep learning networks are in their infancy. So what can you do to debug your deep learning set-up? Let's get to know some of the better known existing tools and get some practical tips on how to use them.

This article is divided into three parts. **In the first**, we'll give you some practical tips for deep learning network training. **In the second**, we'll demonstrate how to implement some of these tips in practice in TensorFlow and Keras. **In the third and final part**, after you've trained your network, we'll show interpretive approaches to understanding and analyzing what's been learned, and what is happening 'under the hood' of the deep learning network.

First thing first, a number of practical tips when training neural networks and a recommended sequence to follow:

1. Review the overall structure of the network, and make sure the output dimensions of each layer are appropriate to the input dimensions of the next layer.
2. Make sure the loss value is reasonable. There are a number of tests you can run for this with the network in its preliminary state, and all weights initialized at very small values:
    a. The loss value produced should be the maximal loss, (as determined by the loss function you have chosen to work with and the number of the categories).
    b. Verify that the network is functioning properly by checking that when you add a regularization term to the above loss function the loss value goes up a little.
3. You should run an overfit test -- that is, run the network on a simple, clear-cut dataset. There are two types of datasets that might be relevant here: 1) A set of synthesized data that are very easy to tell apart -- such as images that are all either clearly red or clearly blue. 2) The other extreme -- a set of unclassifiable data, where you expect no correlation between input and output -- such as synthesized images where each pixel has a randomly determined value. If the network fails these very simple classification tests, there is a problem with its structure or it suffers from very flawed weights initialization.
4. Continued overfit testing -- train the network on a very small subset of your real-world data samples. If the network fails to learn on this small set, then, again, there is likely a problem with the code and network structure.
5. If the network managed to learn successfully on a small dataset, you can start adding more and more data, while at the same time monitoring the activations of hidden units with histograms to see that they are “healthy”.
6. Tuning the learning rate hyper-parameter -- if the network isn’t learning (that is, it’s loss isn’t decreasing), you can lower the learning rate -- you won’t reach convergence at these rates, but the important thing is to verify that accuracy is going up and loss is going down as training progresses. In the same context, you can lower batch-size to 1 (a single image) -- just to get a very quick feedback and make sure the general trajectory is that the network is learning.
7. Setting the regularization hyper-parameter: in general, you should start with a small regularization term. Then, as you tune the learning rate, as described in (6) above, your loss term should be going down. If it doesn’t, you can slowly increase your regularization term and re-run the training process.
8. Hyper-parameters in general should be tuned as follows: first, test a range of values (in logarithmic scale) by running them for a limited number of

cycles to get a "general feel" for the right ballpark for your hyper-parameters, making sure your loss term is going down. Then, later in the process, you fine-tune your hyper-parameters over larger numbers of cycles.

9. Make sure there are no `nan` or `inf` values throughout the network weights -- their presence usually indicates that something is wrong with the weights update, either the learning rate is too high or low, or there is a problem with the update computation.
10. Analyze the rate of weights update -- the rate of update between batches should be more or less constant and should range around 0.001.

Now, let's demonstrate the use of some of these tips. To do this, there are a number of debugging tools we shall get to know today: **TensorBoard** and **tfdbg**. Some of the capabilities that those tools allows you to follow are:

1. Loss value and hidden unit activations histograms monitoring capability.
2. Monitoring the rate of weights update.
3. Verifying there are no `nan` or `inf` values throughout the network, and that the fit output dimensions of each layer are appropriate to the input dimensions of the next layer.

Tool

TensorBoard provides a suite of visualization tools to make it easier to understand and debug deep learning methods written in TensorFlow. For TensorBoard visualization you can use the following functions: `tf.summary.scalar` to trace a single statistic, and `tf.summary.histogram` for aggregate statistics. For instance, you can add the following code snippet to network layers constructed in TensorFlow to export its statistics to TensorBoard to trace distribution, variance, and max and min values of each parameter which we want to monitor.

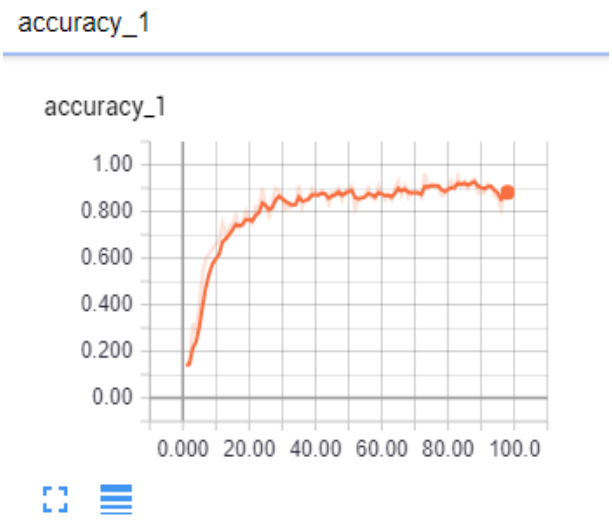
```
def variable_summaries(var):
 """Attach a lot of summaries to a Tensor (for TensorBoard
visualization)."""
 with tf.name_scope('summaries'):
   mean = tf.reduce_mean(var)
   tf.summary.scalar('mean', mean)
   with tf.name_scope('stddev'):
     stddev = tf.sqrt(tf.reduce_mean(tf.square(var - mean)))
   tf.summary.scalar('stddev', stddev)
   tf.summary.scalar('max', tf.reduce_max(var))
   tf.summary.scalar('min', tf.reduce_min(var))
   tf.summary.histogram('histogram', var)
```

For demonstration purposes, you can include the above functions in constructing the simplest FC layer. In the function, we see `variable_summaries` used for the `biases` and `weights`, while for the activations values themselves we're interested only in `histogram`.

```
def nn_layer(input_tensor, input_dim, output_dim, layer_name,
act=tf.nn.relu):
      with tf.name_scope(layer_name):
       # This Variable will hold the state of the weights for the layer
       with tf.name_scope('weights'):
            weights = weight_variable([input_dim, output_dim])
            variable_summaries(weights)
       with tf.name_scope('biases'):
            biases = bias_variable([output_dim])
            variable_summaries(biases)
       with tf.name_scope('Wx_plus_b'):
            preactivate = tf.matmul(input_tensor, weights) + biases
            tf.summary.histogram('pre_activations', preactivate)
       activations = act(preactivate, name='activation')
       tf.summary.histogram('activations', activations)
       return activations
```

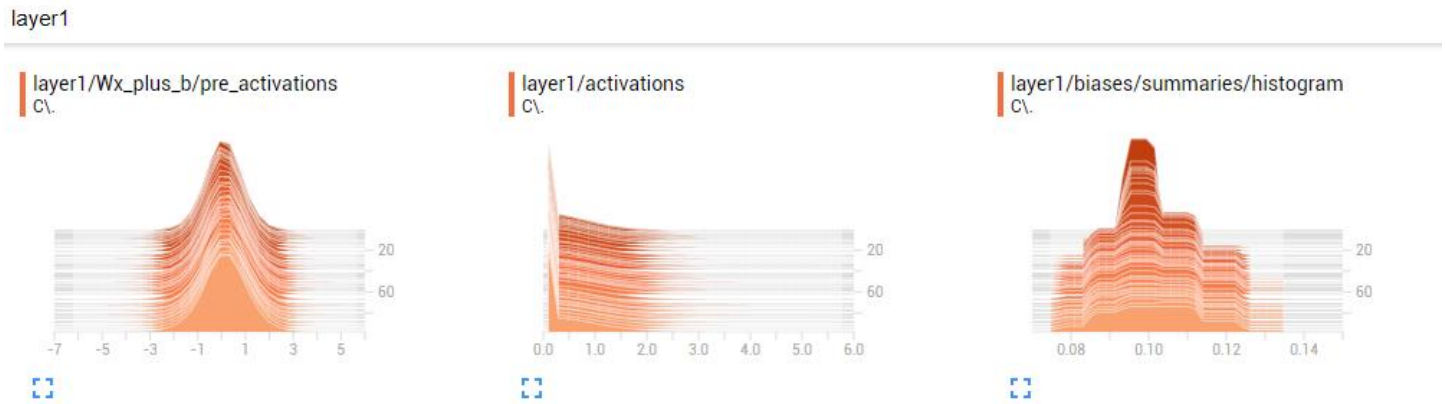
When you run TensorBoard in your browser, it looks like the images below (to learn how to run TensorBoard follow the link):

For example, accuracy:

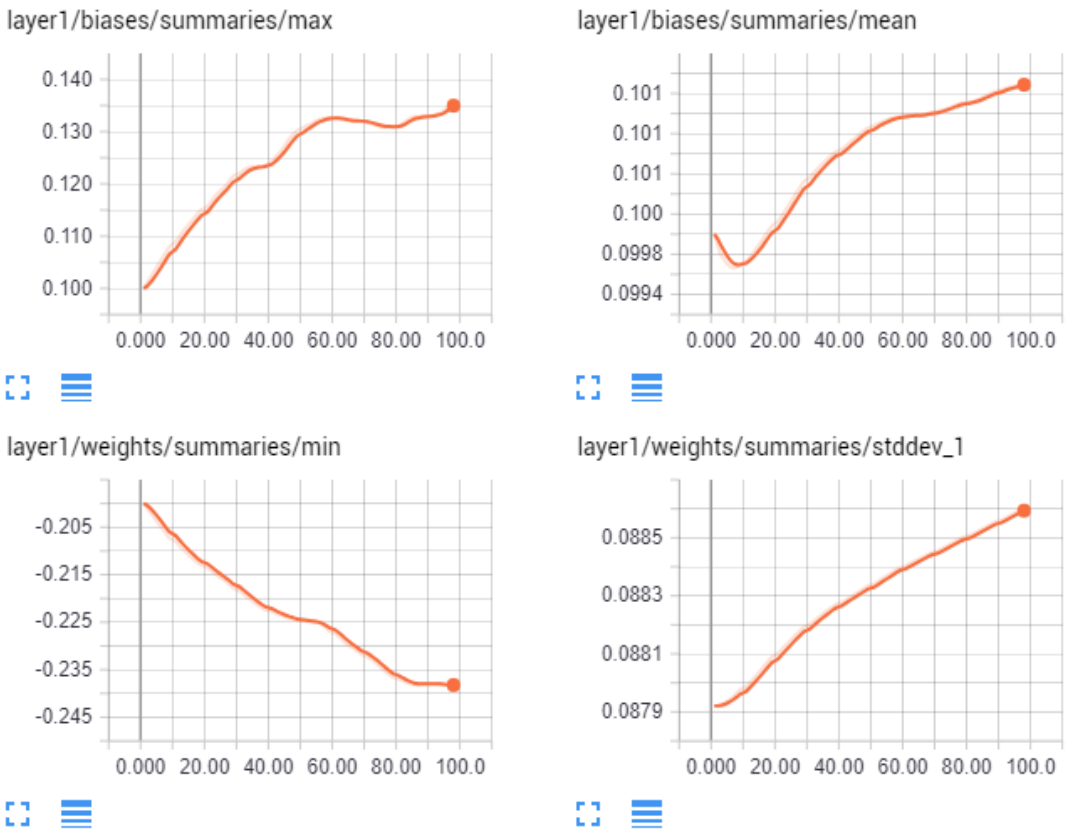


The histograms for preactivations and activations:

Tool



The stat sets for the max, min, mean and standard deviation values:



All of the above can also be done using Keras. As always with Keras, everything is simpler and more elegant but with less control over details. The following code snippet demonstrates using TensorBoard with Keras. The code in fact includes three lines: The first defines the TensorBoard object and the statistics to be collected (a histogram of the parameters, the structure of the graph itself, and the capability to look at the images on which the network is training). The second line compiles the model. And the third, runs the training while collecting the statistics determined, by calling `callbacks` in the `run` function.

Important side note: unfortunately, currently you cannot use ImageDataGenerator together with `TensorBoard` in Keras.

```
tbCallBack = Keras.callbacks.TensorBoard(log_dir='./Graph',
                                         histogram_freq=True,
                                         write_graph=True,
                                         write_images=True)

# TrueCompile the model
model.compile(loss='categorical_crossentropy',
              optimizer=optimizers.RMSprop(lr=1e-4),
              metrics=['acc'])

model.fit(x_train, y_train,
          batch_size=batch_size,
          epochs=epochs,
          validation_data=(x_test, y_test),
          shuffle=True, callbacks=[tbCallBack])
```

Now let's move to **tfdbg** which is a **specialized debugger for TensorFlow**. It lets you monitor the structure, weights and parameters of TensorFlow graphs.

To work with tfdbg, **you must include** the following lines in your code:

Tool

```
from TensorFlow.python import debug as tf_debugsess =
tf_debug.LocalCLIDebugWrapperSession(sess)
```

In the following example, we'll see how to use tfdbg to monitor whether there are any `nan` or `inf` values anywhere in the network.

The code below demonstrates how to register a simple filter named `tfdbg.has_inf_or_nan` to determine if there are any nan or `inf` values in any intermediate tensor of the graph:

```
sess.add_tensor_filter('has_inf_or_nan', tf_debug.has_inf_or_nan)
```

After running your network (with the above code), you will be prompted with the tfdbg, then can use "`run -f has_inf_or_nan`" which executes the training procedure until any tensor contains `nan` or `inf` which triggers that filter and stops the training execution process.

Finally, after the network has successfully been trained and achieved a low loss, we would like now to try, analyze and understand what the network has actually

learned: what has the learning process taught each individual neuron to represent? We will demonstrate a number of interpretive approaches which you can use as building blocks towards this objective.

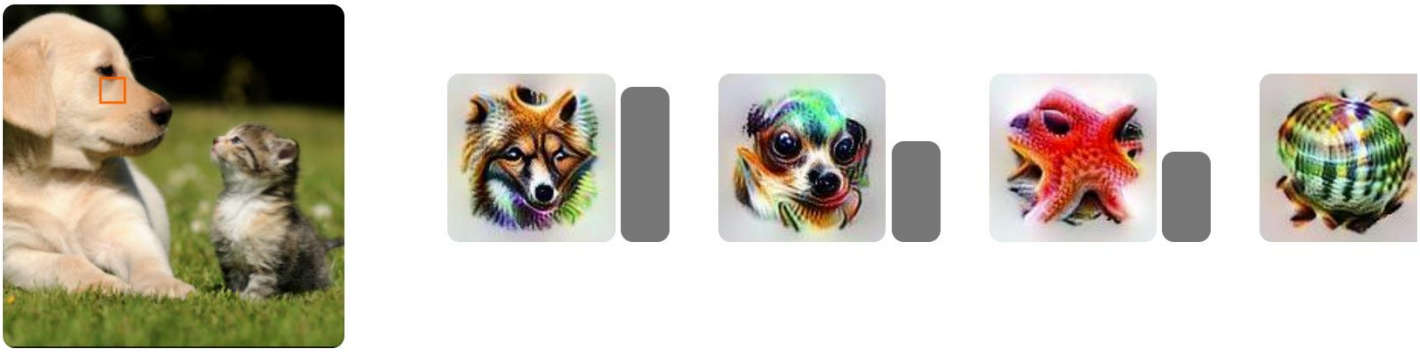
A **semantic dictionary**, similarly to a regular bilingual dictionary, is a list of pairs (neuron, iconic visualization), which provides a visualization of what each individual neuron of each layer detects. This enables us to better grasp the underlying mathematical objects of deep learning networks' hidden layers -- the first step to interpretability. Neuron activations now map to iconic visualizations, instead of abstract indices -- many visualizations will bring to mind natural language human concepts, such as "*floppy ear*", "*dog snout*" or "*fur*".

We can use the semantic dictionary as a building block to construct an activation vector, a weighted composite of semantic dictionary icons -- visualizing what specific combinations of neurons firing represent.

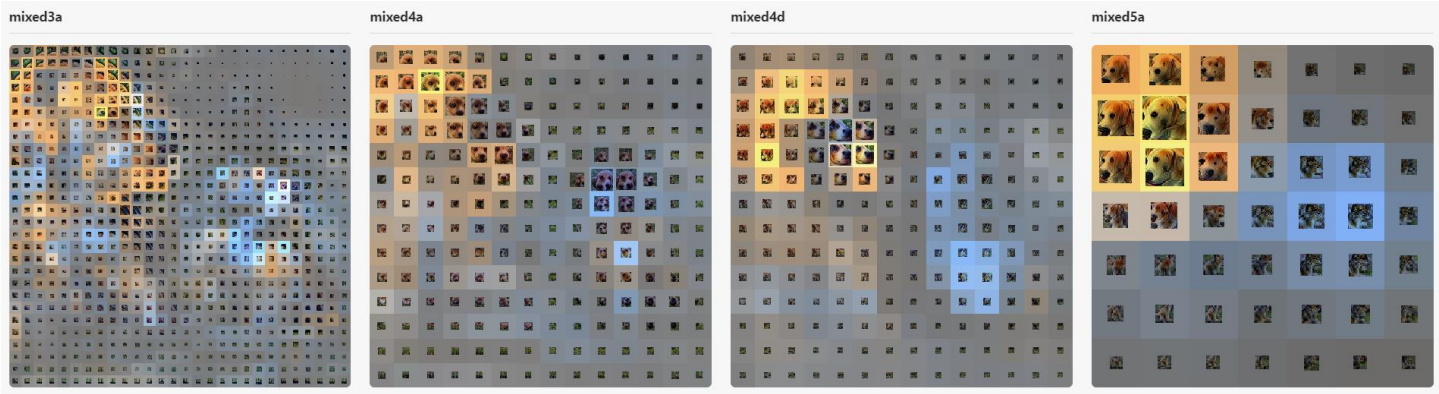
We can group neurons to construct composite activation vectors along different conceptual axes:

Spatial attribution of neurons: this is usually done using saliency maps. To produce a saliency map, we take a label and go backward through the network to the image, using a simple heatmap to show what pixels of the input image most contributed to the classification. Based on the saliency map and the semantic dictionary, we replace image areas with their activation vectors. Because these are visualizations, unlike a simple heatmap, they can represent the relevance of an area to multiple classifications (i.e., dog, snout, white, etc.).

Tool

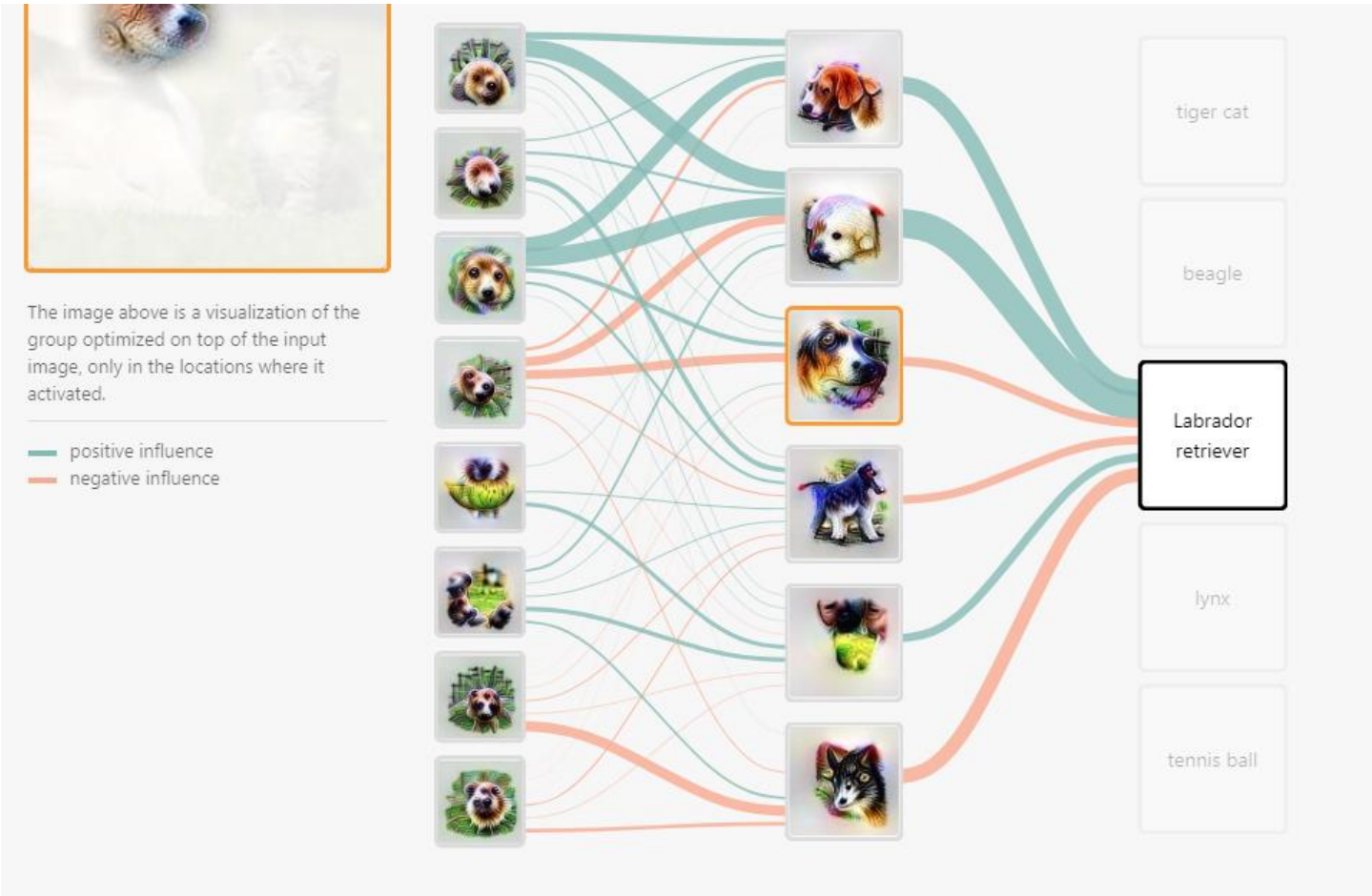


The figure below demonstrates interpretive visualization using the above image of labrador retriever and tiger cat. As the image goes through successive reductions in resolution as it is processed along layers of the neural network (from left to right) -- we get decreasingly dense grids. Each grid cell's color is determined by the overlay of two saliency maps -- orange for 'labrador retriever', and blue for 'tiger cat'. Then, the activation vector visualization (based on semantic dictionary icons) for each grid cell is overlaid on top of the saliency maps, with size representing magnitude.



Matrix factorization: is used to aggregate activations along channels and spatial areas. This reduces the overwhelming number of neurons to a small set of groups, distilling the process and power of the neural network's deep learning.

The figure below demonstrates factorization at two consecutive network layers. The first layer factorized into 8 neuron-groups, the second into 6. For the label 'labrador retriever', it shows the influence of each neuron group at each layer towards the image's classification as that label.



Tool

Figures in this last part are from this link, where you can also find further reading on interpretability.

Guest

Jeff Clune is a Senior Research Manager at Uber AI Labs and the Loy and Edith Harris Associate Professor in Computer Science at the University of Wyoming. We talked with him ahead of RE•WORK's Deep Learning for Robotics Summit in San Francisco, where he will be a speaker. To register for the Summit, see our Upcoming Events page. Read also our interview with Raquel Urtasun.

Jeff, can you tell us about your work?

I work in machine learning and deep learning. The goal is to try to invent the next generation of machine learning and deep learning algorithms to improve what computers are capable of doing with the ultimate goal of trying to build artificial general intelligence, but also with the near term goal of trying to help all of the various constituents in the world who want to take advantage of machine learning to try to do wonderful things within their organizations.

Do you have your eye on autonomous driving in particular?

No. I would say I'm generally interested in artificial intelligence, especially in terms of trying to create artificially intelligent robots and software in any form. I am interested in reinforcement learning, deep learning, and neuroevolution. I don't really focus on one particular domain such as autonomous driving.

Has what you have found in your research met your expectations?

Yeah, if you had asked me in my late teens what I would want to do more than anything, when I was in college for example, I would have told you that I wanted to be a professor. If you had asked me a few years later as I was leaving college and then while I was in Silicon Valley during the dot-com boom, what I would have wanted to do more than anything, it would have been to work at an impressive, disruptive startup, especially in their R&D labs, inventing the technology that they would use to change the world for the better.

So this is off your bucket list now?

That's right. Now I have both. I'm a professor, and I'm working at the largest startup in the world, disrupting and improving the world through new artificial intelligence technologies. To some extent, I have satisfied both goals simultaneously, which I would not have originally thought was possible.

What do you know about robots that the public doesn't know?

That's a great question. I think the short answer is that most people are probably going to assume that robots work a lot better than they do because they see YouTube video demos of only the things that work extremely well. Robots have already profoundly changed our economy, but only in very narrow cases such as factory floors, where

we can remove all of the complexities of the world, fix the robots in place, and ask them to do the same thing over and over again. What robots cannot do are the simplest things that animals are capable of: walking through a park, picking up an acorn, throwing a can in a trash bin, picking up objects, etc. There are things that a two year old or even a one year old can do that our robots can't. I think probably if the public thought about why there aren't robots out there in the world helping us, they might realize that robots can't do these things. Hollywood and the best YouTube clips paint a different picture. Probably their expectations are out in front of our current technological power. That said, that's exactly what we're working on in the field of deep learning and deep reinforcement learning for robotics. We're trying to create robots that have those capabilities. Progress is rapid, and soon the capabilities of robots may catch up a little bit with the portrayal on YouTube and in Hollywood movies.

Is there one specific capability that you, personally, would like to realize?

I can say that there is one capability that I have been a part of pushing the state-of-the-art forward on that is an important capability… that is robots that can adapt to changes—either in their own bodies such as with damage, injury, or their batteries wearing down—or to changes in the environment, like if there is a natural disaster and a building has collapsed: First of all, that is an unknown, new environment that has never been seen before. Second of all, the robot could become damaged in such an environment. What you don't want is what happens now which is the second that a robot is off of its pre-programmed game plan, it's hopeless, lost, and ineffective. It just gets stuck. Instead, we want our robots to adapt, to either soldier on with their mission, such as find survivors, or at least, take themselves back to a repair station so we can repair them and send them back out into the world.

If they are able to repair themselves, we might be able to one day send them to the galaxies…

That's true. The classic example is that the Mars rover got stuck in the sand. Despite years of the best engineers trying to get it unstuck, they could not adapt to that situation. We would rather send robots that can adapt on their own and also don't require us to remotely program them... so that they can automatically make autonomous decisions including adapting to unforeseen situations to carry out missions.

What will happen if one day robots will be able to produce new robots?

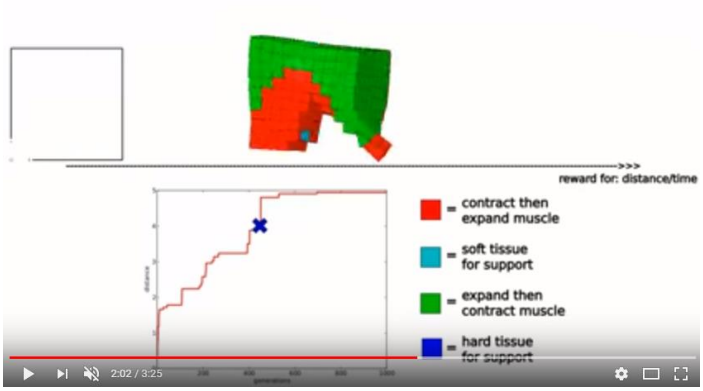
I think that's the beginning of an extremely powerful feedback cycle, where optimization can take over. . I think it will happen in software far before it happens in hardware, like robots. You might have AI that makes better AI. To some extent, that happens a little bit now, but not in any way like what I think your question is asking about.

Ultimately, we do think that robots and AI will build better robots and AI. Then you can think about rapid accelerations of their capabilities on the cognitive side and on the hardware side for robotics in AI.

In a few weeks you are going to speak for the first time at RE•WORK's Deep Learning for Robotics Summit, in San Francisco. What are we going to learn from you?

There are two main new things that I am going to highlight. The main point of my talk is how we can get robots to adapt to changes in their environment or to damage. That specifically will cover a paper that we have published with my excellent collaborators - Antoine Cully, Jean-Baptiste Mouret, and Danesh Tarapore - that was on the cover of Nature. That paper discusses how you can get a robot that has become damaged to conduct a very few number of experiments and, within one to two minutes, figure out how to continue on with its mission… literally one to two minutes of watch time, and the robot will figure out how to adjust to the situation and continue to go about what we've asked it to do. The second thing I want to discuss is what most people in deep learning for robotics and deep reinforcement learning have been focused on, which I call traditional reinforcement algorithms, such as DQN (deep Q-network), an algorithm from Deep Mind that was on the cover of Nature. We recently showed in a series of papers that came out at Uber that evolution is actually a competitive alternative to these traditional reinforcement learning algorithms. That is following on the heels of work at OpenAI, which showed that a different evolutionary algorithm is a competitive alternative. So between the work at OpenAI and our work at Uber, there are basically two new champions in the tournament that can all take on any given dragon (a dragon might be a hard engineering problem). What we're finding, which I think is very interesting, is that each of these different algorithms is very good at different types of problems. One message that I will emphasize to the practitioners or the people from industry who are at the RE•WORK Summit is that if you have a particular problem, and you know that you have four algorithms, and one of them is likely to work and the others aren't, then you probably shouldn't try one of them, stick with it, and really try to get it to work. You should probably try all four of them really quickly and see. One of them might work far better than the other ones. Ahead of time, it's not clear

which algorithm that will be. There are new arrows in our quiver or tools in our toolbox that the industrial community should be aware of that they can use to try to tackle the hard problems that they try to solve on a daily basis.



Why is Uber interested in this kind of things?

The short answer is that Uber has a tremendous number of fascinating, challenging engineering problems that can benefit from machine learning. Uber benefits greatly by inventing new state of the art algorithms to attack these problems. The more tools we have, the more capable we are to solve the problems we have. Machine learning will make many, many facets of any modern company better, including Uber.

Also, why are you going to RE•WORK's robotics summit in San Francisco to talk about these things?

The summit is a wonderful opportunity to share the cutting-edge of research in machine learning and deep learning with the people out there in the industry looking to employ these technologies. One of the hardest things, if you're in industry, is knowing what is possible and what is not. It's hard to distill out of the hype of popular press articles, the things that sound really good, but aren't really ready versus the things that you can take off the shelf now and mostly expect to help you this quarter on a problem that you are focused on. I think this is a wonderful opportunity to share with the applied machine learning community some new algorithms that we have invented and the research community has invented that they might be able to take advantage of this year.

Would you like to share with our readers why you recommend they come to the RE•WORK robotics event to listen to you and other speakers?

What has impressed me about the RE•WORK summit is how they marshaled some of the brightest minds in academia at the forefront of deep learning to share their expertise and innovations with the industrial community. The lineup of speakers is superb. Therefore, I would recommend that people come hear from the rockstars of academia about what technology they might not yet be aware of that could help their businesses. For a very long time there have been a lot of hype and hope surrounding artificial intelligence, machine learning, even deep learning, and especially robotics. All of those things have historically been very challenging to get working, but recently have started to actually work, not just in the academic papers, but in industrial applications and in companies. They're being deployed, and they're making and saving a lot of money. My guess is that most of the people who are attending RE•WORK are doing so because they are already convinced of that, but to the extent that there are skeptics, I would say that if you are not currently finishing, if not formulating, your artificial intelligence strategy, you may be left behind. This technology is working. It's getting better, and you'll want to be actively employing it throughout your company, or somebody else is going to take your place.

With Michela Antonelli

Michela Antonelli is a research associate at University College London (UCL) in the Translational Imaging Group. She is also one of the organizers of this year's Medical Segmentation Decathlon challenge, sponsored by NVIDIA, DeepMind and RSIP Vision.

Michela, can you tell us about the challenge at medicaldecathlon.com?

The aim of this challenge is to find the algorithm that can segment everything. With everything, I mean different types of organs/pathologies using different types of images. There are a lot of niche segmentation algorithms around, but what is missing is an algorithm that can be generalized and applied to different types of data set. This is the real challenge that we are proposing.

You were personally charged with organizing the data, weren't you?

Yes, I collected all the data and it has been a very tough work! Clinicians sent me data in different formats with different names and I had to make it uniform and try to correct any mistakes. The whole data set has to be used by the same algorithm, so there has to be a common way of treating it to make running the algorithm as easy as possible.

We managed to release the data set with the most open license possible, so people can use the data, change the data, and do whatever they want with it. It was a challenge to find hospitals that could give us data with this kind of license that could be released to the public in such an open way.

Is there one specific hospital that was particularly cooperative you would like to name?

Everyone has been really cooperative – I don't want to forget someone! Amber Simpson at the Memorial Sloan Kettering Cancer Center. Bjoern Menze, Geert Litjens, and Patrick Christ, were all very helpful.

Do you have any tips for the participants to the challenge?

They should try to be as general as possible, because this challenge is divided into phases. In the first phase, they have seven tasks where they can train their algorithms and produce results. In the second phase, there are three mystery tasks that they don't know anything about and they have

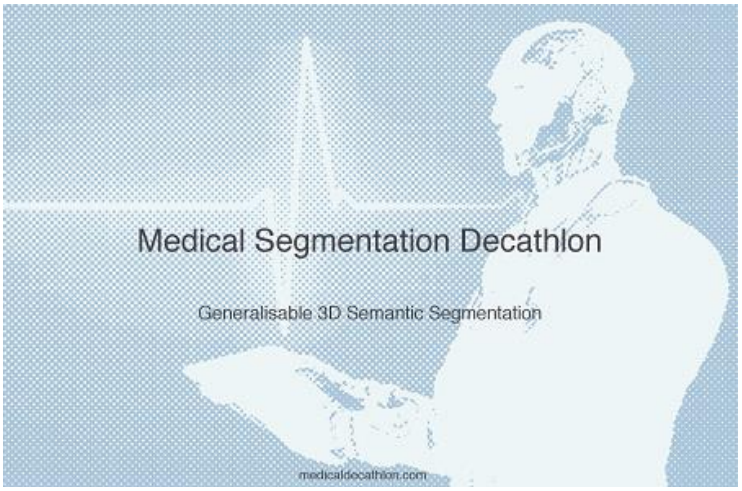
very limited time to train their algorithms. They can't change the parameters, so I suggest not focusing on obtaining the most amazing results in the first phase. Think about the fact that the algorithms will be used for the other mystery tasks, so don't be too specific about those seven data sets.

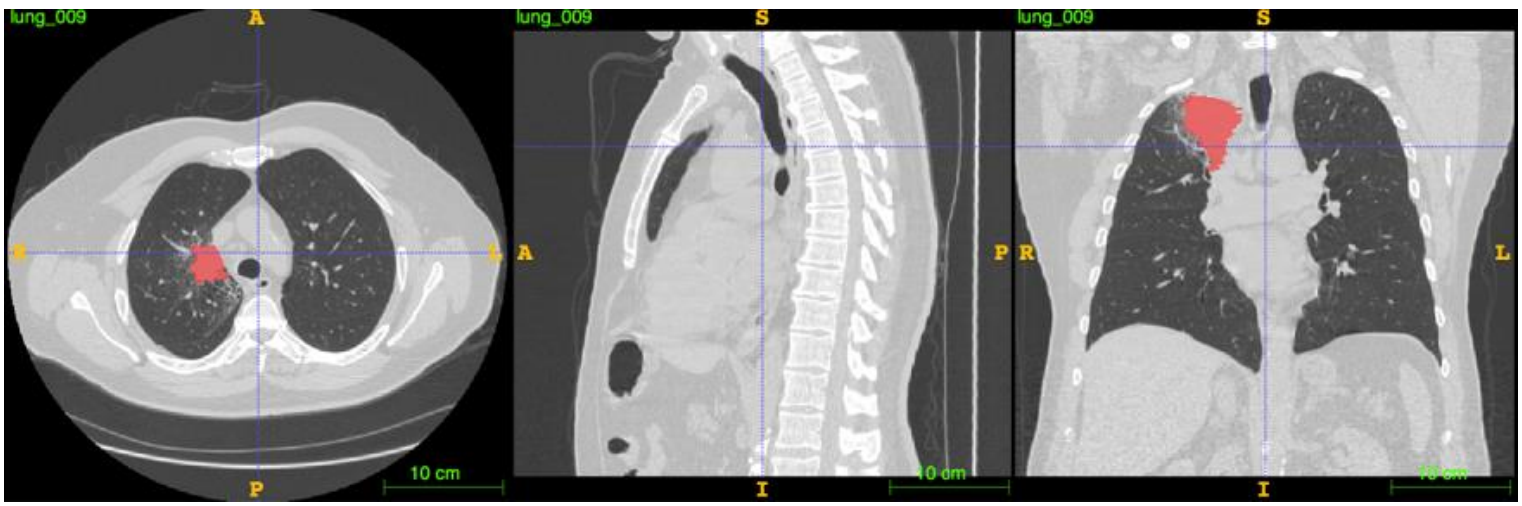
All the instructions are on the website and it is very important to understand that if you don't participate in the first stage, you can't participate in the second. It is mandatory that anyone who wants to participate has to register and submit their results before the deadline of the first stage. After this date, no one can register anymore, because the mystery tasks will be released at this point.

What was particularly challenging in organizing this challenge?

Besides the collection of data, another very challenging aspect is how to combine all the numbers that arise with these segmentations, because we are talking about thousands of data sets and there are several indexes that can be used for evaluating each segmentation. There is a team that is working on how to combine the data and how to combine the results, because finally we need one single number - we need the winner and we need to say that this algorithm is the best one. We have to find the statistical significance for this combination and not only to decide which index to apply for evaluating the segmentation, but also, how to combine all these indexes together to end up with a number. I think this is as much of a challenge as collecting all the data.

"It is very important to understand that if you don't participate in the first stage, you can't participate in the second!"

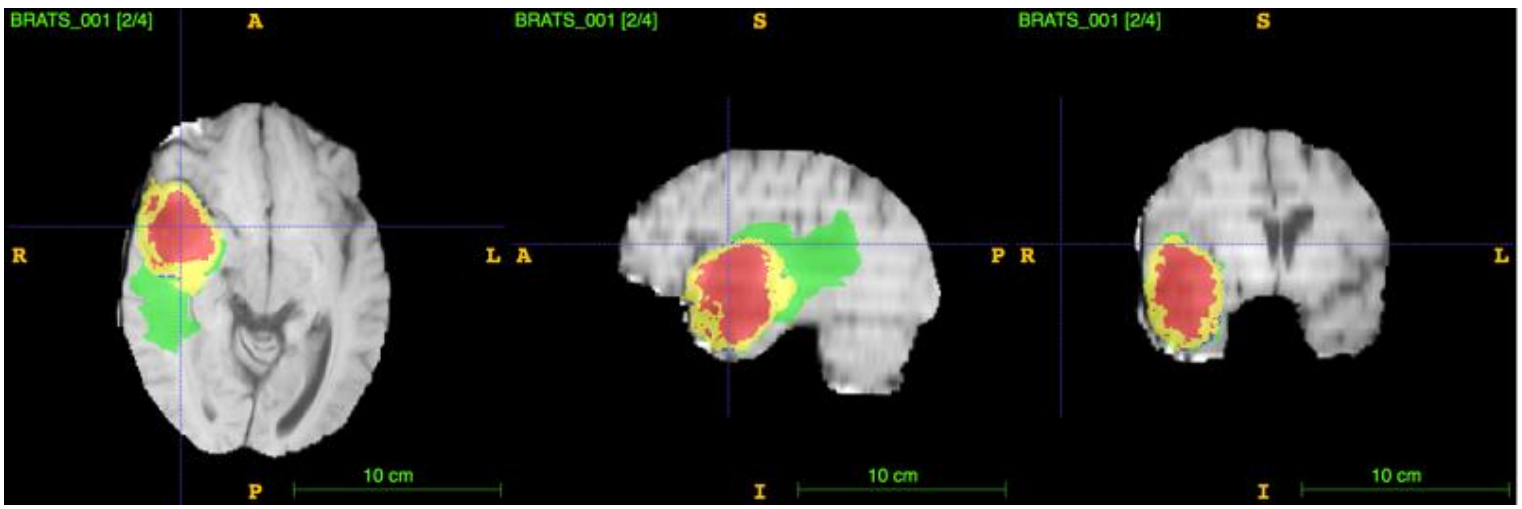




"...to build a computer-aided system which can be used to better classify the lesion inside the prostate and avoid unnecessary prostate biopsy!"

Finally, can you tell us about your regular work besides the challenge?

Medical segmentation is one of the main topics of my research. I use MRI images of the prostate to segment the prostate into the two main areas – the transitional zone and peripheral zone. I'm more a Machine Learning person, that is my background, but now I'm trying to apply this very theoretical background acquired during my previous experience at the University of Pisa in the field of medical imaging. What I'm trying to do is to build a computer-aided system which can be used to better classify the lesion inside the prostate and avoid unnecessary prostate biopsy. The idea is to build a system that can first identify where the prostate is, then identify if there are some pathologies inside this area. I do this by applying different techniques, from deep learning to more traditional classification systems. It is a tough area because the images are very noisy and there is a lot of variation between patients. It is not easy even for the human eye to spot differences between different types of lesions inside the prostate.



Feel at CVPR
Without being at CVPR!

June 19-20-21

Get CVPR Daily news every day in your Mailbox!

Click here and subscribe for free to CVPR Daily!

Deep Learning Convolutional Neural Network for the Classification and Segmentation of In Vivo Confocal Microscopy Images

Project

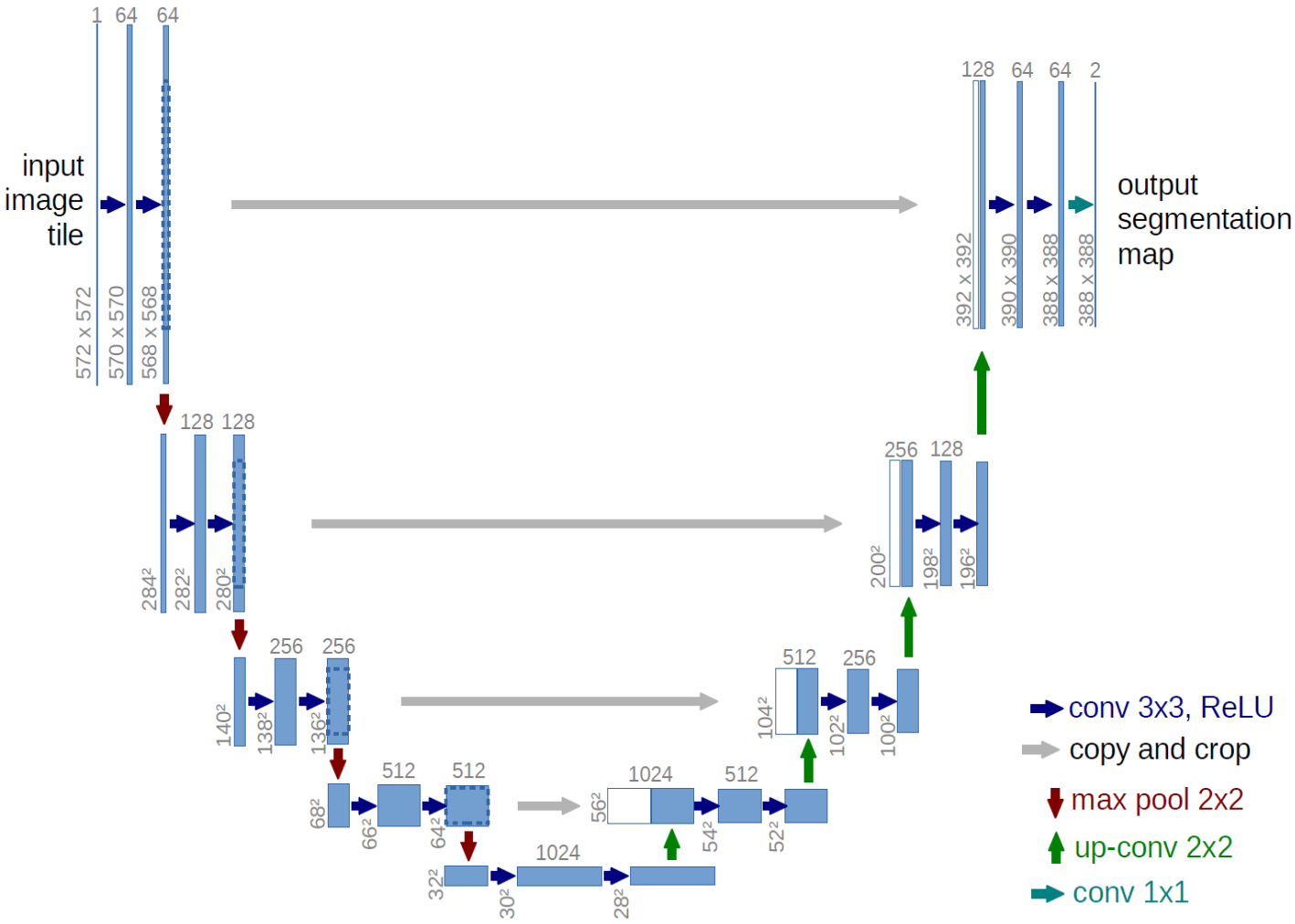
As usual, this section covers computer vision work done by RSIP Vision. This month we shall focus on a very recent R&D project, completed by RSIP Vision in cooperation with Tufts. Findings were presented at the annual meeting of ARVO, the Association for Research in Vision and Ophthalmology.

Introduction:

In Vivo Confocal Microscopy (IVCM) is a non-invasive imaging technique enables to acquire coronal images throughout the cornea at a cellular level. Through these images it is possible to assess the Dendritic Cells (DCs) density, which is strongly correlated with Dry Eye Disease (DED), one of the most frequent ophthalmic disorders. Therefore, this tool provides a non-invasive biomarker for corneal inflammation. Our study proves that automated Convolutional Neural Networks (CNN) detect and quantify DCs, providing an accurate estimation of ocular surface inflammation and significant benefits for diagnosis and treatment.

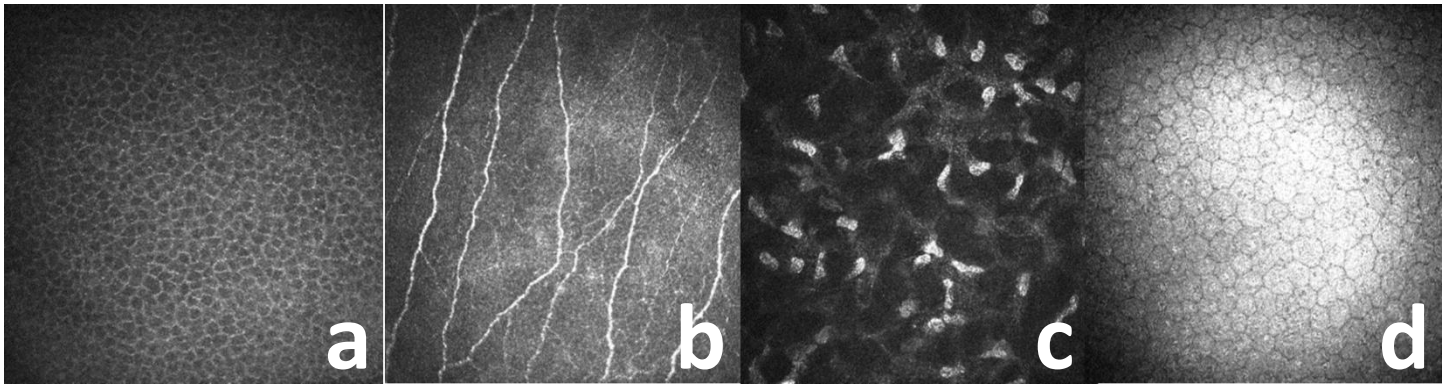
Method:

A 'U-Net' based fully convolutional neural network (CNN) architecture was developed to analyze the 2140 IVCM images from our database: images demonstrative of healthy epithelium, subbasal nerve plexus, stroma, and endothelium. Training and validation datasets were selected randomly from the database.



258 images were selected randomly in the same database to demonstrate DCs from DED patients. Segmentation CNN was trained 5 times on random sets of 233 images and tested each time on the remaining 25 images.

The image below shows (left to right) IVCM images demonstrating healthy a) epithelium, b) subbasal nerve plexus, c) stroma and d) endothelium.



Density was first calculated by expert ophthalmologists and imaging specialist via Image J, a public domain NIH imaging software. Intermethod agreement was calculated between Image J and the CNN.

DC Segmentation:

Within IVCM images, dendritic cells are identified and distinguished from the background as bright individual dendritiform structures with cell bodies. Training was improved by manual verification and correction of all DCs found by the CNN. See image below, where results from Image J can be compared to results from CNN:

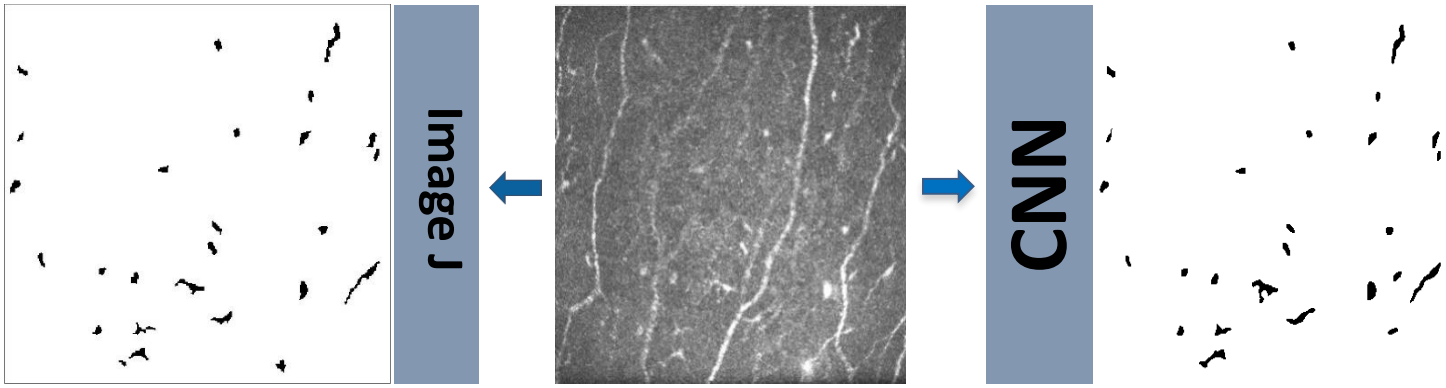


Image J		CNN
120.5±161.6	DENSITY (cells/mm^2)	118.2±155.8
1692.6±2641.0	AREA (µm^2)	1736.7±2458.7
1040.0±2458.7	PERIMETER (µm)	862.1±1135.0

Interclass correlation was calculated between the semi-automated analyses (Image J) vs. deep learning (CNN). Shows intermethod agreement values very close to 1.

Our results demonstrate that Deep Learning allows a standard, objective, accurate and fast analysis of IVCM images, thus providing an effective tool for evaluating pathologies.

Conclusion:

Laser IVCM enables non-invasive visualization of corneal layers. Images of corneal DCs provide precious information about the severity of inflammation. Our results demonstrate that our Deep Learning algorithm allows a standard, objective, accurate and fast analysis of IVCM images, thus providing an effective tool for evaluating pathologies, while at the same time it saves the tedious manual work. As a result, it contributes to improve diagnostic and treatment.

My name is Bot: James Bot!

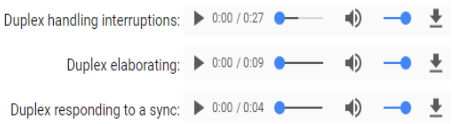
That might be the name of the next secret agent 007, at the service of POTUS. **Dawn Meyerriecks**, the **CIA**'s deputy director for technology development, declared that the Agency was adapting to a new landscape where its primary adversary is a machine, not a human agent. It seems that the CIA has spent more than 30 years preparing for this transition from humans to computers... Read More!

The GANfather: The man who's given imagination to machines:

If you don't like spy movies, we have something for you too. Here's a very nice portrait of **Ian Goodfellow** and his work on GANs. Hence the nickname GANfather. Even if you don't like mafia movies, it's a great reading: Read!

Google Duplex: Real-World Tasks Over the Phone:

You have certainly already seen this and heard this. But if you missed it, you have to see how this bot by **Google** books a table for you. Hilarious and impressive! Read Now!

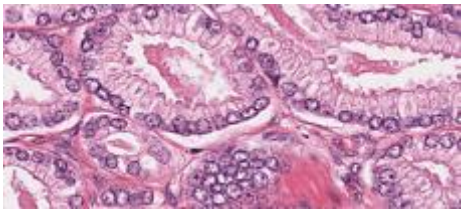


Big list of open source projects from Facebook AI:

It's not me revealing that, it's **Yann LeCun** on his Twitter account. Some of them were released on the same day of the announcement. Enjoy!

Open Images V4 and ECCV 2018 Open Images Challenge:

Another item from **Google**, this one signed by Vittorio Ferrari. Version 4 has 15.4M bounding-boxes for 600 categories on 1.9M images, making it the largest existing dataset with object location annotations. Read Now...



Free to download datasets by Figure Eight (was CrowdFlower):

These datasets were curated on the Figure Eight platform. They are free to download for the entire data science community: medical, handwriting, language... Enjoy!

9 Pitfalls to avoid in building a machine learning program:

Write your code with a solid foundation thanks to these tips. Some are obvious, some are precious, all are good. Read!

A python toolbox for gaining geometric insights into HD data:

where HD does not mean high definition but high dimensional. This library is called **HyperTools** and you will use it to visualize and manipulate HD data in Python. Try!

Other news you shouldn't miss: Notes by Andrew NG! Google's new tools and their price

Marta Kersten-Oertel

Marta Kersten-Oertel is an Assistant Professor at Concordia University where she is the head of the Applied Perception Lab.

Marta, are you originally from Canada?

I am not. I was born in Poland. We moved when I was five years old to Canada with my family.

So you are practically Canadian?

So, yes… I am very much Canadian… a naturalized Canadian.

Can you tell us about your work?

I work in image guided surgery. I’m very much interested in more of the human aspect of things. I’m a prof in human-computer interaction here at Concordia. I’m interested in the perception of the surgeon and how he perceives the images in planning information on the navigation system. I look at things like how we can use augmented reality to help guide the surgeon. Current research also involves things such as having audio feedback instead of visual so as not to overload the whole system… a lot of medical image visualization and perception research in the context of image guided surgery.

Are many surgeons involved in this work together with you?

I did my PhD at the Neuro, and that’s where all of my contacts are from. One important aspect of our work is actually getting feedback from surgeons and bringing the systems into the operating room so that we can see if what we are doing is actually useful. There are a number of surgeons that work with us on these projects.

So it’s a sort of interdisciplinary cooperation?

Yeah, we try!

When you are doing these kinds of cooperations, who is the leader? The scientist or the surgeon?

That is a very difficult question! I really think it is a cooperation though. We bring our ideas to them. One of the things that I did during my PhD was to actually go into the operating room and see how things were working. From an outsider’s perspective, you see things that surgeons are so used to. They don’t realize they are accounting for things like they’re using 2D images when they could be using 3D. Then we bring our ideas to them. They either like them or don’t like them. They’re the boss in how to use the technology. It’s an ever-evolving process where we try to take their comments, fix things, and then bring things back into the OR.

How did you feel the first time you

entered an operating room?

I loved it!

Really? The blood and everything?

It didn't bother me at all. I found it fascinating. For me, it continues to be fascinating. There's someone working inside the brain! The majority of the time - well not always the majority - the patient will wake up without any deficits. They're working in the brain! For me, it's more of a fascinating thing. I love going into the OR and watching how the surgeons work and what they are doing.

Why didn't you become a surgeon?

[*laughs*] That's a good question! The thought occurred to me when I was finishing my PhD. My PhD took quite a long time. I was like, "*Oh, maybe I want to go to med school.*" My husband said, "*Maybe it's just a crisis, and that's not necessary.*" [*laughs*] Of course, he said, "*If that's really what you want to do…*". I came from a family of academics. Becoming a medical doctor never crossed my mind until I was in surgery. If I was to do things again, maybe it's something I would have considered. Instead, now, I try to work as closely with them as I can to help.

Never say never! [*both laugh*] What about the discipline in the operating room, is that fascinating too? The fact that everyone has a defined role?

Yes, whenever we go into the OR, I feel that the surgeons are doing us a favor. I'll be honest. I feel like they're there to help support our research. Sometimes it is important to them, but it really depends on the surgeon. Other times, they're really helping us in trying to see how our technology works. Everyone has their role. Some of us going into the lab always feel very comfortable saying, "*Can you use this now? Give us feedback!*" I was more timid, like "*Would you mind using our technology?*" [*laughs*] I felt like I had a particular role, but I think that's very personal as well.

Is there anything that you have developed which is already in use?

I started my PhD a long time ago. Well, it feels like a long time ago. We started working in augmented reality for neurosurgery. It has become used now in commercial products: Brainlab is using it, for example, in their neurosurgical microscopes. Other commercial products are also using it as well. I've always wanted to provide evidence to say: "*Hey, is this useful or not useful?*" I never worked commercially on it, but I hope that, as we provide more evidence, the commercial companies that are working on these things will realize that "*Hey, this is actually a useful tool, and it should be incorporated into our technology!*"

In the operating room: augmented reality view (yellow and blue anatomical data are overlaid on the dura of the patient).

Can you describe the process by which you and the other scientists make conclusions in your research before the industry adopts it and commercializes it? How does this process work?

I think I'd be overstating things to say that we have had a huge influence on what industry is doing. They have their own research teams and do a lot of things. As we publish, we show the surgeons the tools that are useful. One of my things is really to show where augmented reality is useful and where is it not useful. Not every technology will be useful in every situation. We really looked at which specific surgical tasks would really benefit from augmented reality. As we publish these results, hopefully other people will pick up on them.

How would you fix the communication between industry and academia?

I have a hard time commenting on that because I haven't reached out to say, "*We have this idea that we published. It might be interesting for you.*" I'm not sure how well that works. In terms of non-medical research, I have a shared grant with a company: I'm a co-PI with Emad Shihab and Tristan Glatard. That seems to work really well. One of the hurdles we have is when commercial systems are closed, it's very difficult to bring stuff into the operating room. This is why we have IBIS, an open source platform neuro navigation system. There, we can incorporate any research tools. That's definitely a hurdle. It's understandable that the systems are closed. We overcome this by creating our own tools and bringing our research system in parallel with the commercial ones.

Which company sponsors the grant?

It's a Canadian company called Motsai that does system on module boards for wearable sensors and other types of sensors.

Did you dream of becoming a scientist as a child in Poland or did that happen later?

To be honest, I never dreamed of it. I never knew what I wanted to study. This is just a path that I ended up going down. I'm very happy with it. There was never this goal. It's just something that happened.

So what is the goal now?

[*laughs*] The goal now is, in terms of research, to continue to work on developing solutions to clinical problems. Again, one of my big things is not to develop the solutions, but actually have them being used. The goal would be to have these solutions become used in everyday practice.

If you could have one solution adopted in the OR, what would it be?

I find that, in the OR, the workflow is not always optimized. This has to do with the setup of the technology, the image-guided surgery systems, and the way that the surgeon has to interact with the system because they're sterile. They end up calling a technician who then interacts with the system. What I'd really like is these flowing surgical procedures where there are no interruptions because the technology needs to be setup to be used. They have superman X-ray vision to see the tumors. The augmented reality is there. There are many sensors around the operating room so you don't have to move the camera to capture the surgical tools just because something else is in the way… really improving the workflow of the operating room.

Can I ask you about CARS? Since you are not presenting this time, did you have a chance to go through the program and see any session that would like to follow?

There are actually a few interesting sessions that I'm looking forward to at CARS: the tutorial on Rapid prototyping of image guided therapy applications on open source software platform; the session on Technical Standardization in the OR; the workshop on the Digital Operating Room. And of course: the ISCAS session on virtual and augmented reality! I actually only decided last week that I'll be going to CARS.

How comes?

I was waiting for some funding decisions. I'm also pregnant, so I wanted to make sure that I could fly at that time. The decision was made last Friday, finally!

In Pisa for CARS with second son

Application: Altia Systems

Aurangzeb Khan

PanaCast by Altia Systems is the world's first Panoramic-4K Plug-and-Play USB video camera for video-conferencing. We asked Co-Founder and CEO **Aurangzeb Khan** and CTO **Ram Natarajan** to share with us how their technology works and what makes it so cutting-edge.

While traditional video conferencing solutions can be quite expensive and at the same time limit the visual perspective, PanaCast enables anyone using desktop or mobile devices to communicate in an **interactive, real-time Panoramic-HD 180 degree field-of-view video** that replicates the human panoramic perspective. The easy-to-use, portable camera is reasonably priced and works especially well in collaboration spaces and classrooms, delivering a natural video experience.

Together, they build a completely new kind of **e-vision system**, which consists of an arrangement of three cameras, synchronized and optimized in real-time to create a very wide, immersive field-of-view. It generates a visual experience similar to how eyes see the world.

Traditional cameras have been limited to an 80° field-of-view with low distortion. People have tried to overcome this for almost 70 years. The usual approach is to use an ultra-wide angle lens or fish eye lens, but this leads to several problems. First, it creates a very large amount of distortion. When people try to correct the distortion, it's not linear, and the quality degrades. Secondly, this method still doesn't represent how our eyes experience the world.

"*The problem with traditional lenses in cameras* - Natarajan explains - *is that they are always manufactured with 50 to 70 degrees field-of-view. They cannot produce a larger field-of-view without distorting the image itself. Altia System solves this problem*

by taking three images and combining them. This provides a 180° view without distorting the image."

A standard conferencing camera provides a view around 90 degrees, and the software running in the device automatically adjusts the view just wide enough to include everyone in the video. The camera corrects itself rather than recording footage of the actual room itself. This limits the level of interaction during video conferencing, especially considering that 70% of communication is nonverbal.

Whether during a business meeting or inside the classroom, people instinctively like to get context. They want to include everyone in the conversation and see how they react. PanaCast has a core vision of transforming video by giving it a very natural view. The device not only gives a 180° view, but also performs intelligent sensing. The camera analyzes the video to detect people, for example, it detects people's faces and adjusts the field of view automatically to just the right size to include everyone. It can also count and provide information on the number of people present in the meeting, and later, the device will be able to detect other things as well. Then they can combine the detection with the video to take autonomous action or automate/optimize workflows.

To this, Khan adds: "*We think that really great technology works so well that people forget it's there. The technology adapts to you, and you don't adapt to the technology. We want to keep expanding and improving the experience without people having to make manual effort or break the conversation. We think that there are many opportunities in this area.*"

Altia Systems makes both the hardware and the software for their solutions. Their device has three, full independent cameras which are three mega-pixels each. Each has its own

Application: Altia Systems

imager and its own ISP (Image Signal Processing). Their IP is a streaming architecture running in an **FPGA**, which they call the **PanaCast Vision Processor**. In total, there are nine processors in the system, and one of the chips is the core IP. It also has two microphones and an **audio DSP (Digital Signal Processor)**.

The three adjacent cameras overlap by about 20 percent. Then they run in real time a software algorithm, a dynamic scene stitching algorithm, to do pixel accurate stitching. Computationally, this is a very hard problem. Many try to do this in the cloud or as a post-production step afterwards, but the team at Altia Systems figured out how to do it in real time at the camera level inside the device. It's extremely fast in order to collaborate, and **the incremental latency for the stitching is 5 milliseconds**. Which is exceptionally fast.

When asked about the computer vision techniques that they use to create this impressive technology, Natarajan explains: "*We are using something like multi-view geometry to figure out how to stitch these three videos simultaneously.*"

"Altia Systems figured out how to do it in real time at the camera level inside the device"

Today, new AI technologies do not require viewing the face from the front to recognize a person. It can still detect the person from any angle, even if the person turns around. If someone else enters the room, the camera understands and expands the view to include that person. As a result, Altia Systems enhances the experience by expanding the distance in which these algorithms work and improving the accuracy in the identification of people.

Their software called **Whiteboard** detects certain objects and utilizes that information on the user's behalf. For example, it enables simple whiteboard sharing during video collaboration.

"We are using something like multi-view geometry to figure out how to stitch these three videos simultaneously"

Much has happened since Altia Systems started shipping their solution in Q4 of 2015. The **PanaCast 2** received numerous awards including the **CES 2016 Innovation Awards Honoree recognition** followed by the **rAVe Best UC Solution** at Integrate Australia 2017. In 2017, they created a real-time 3D system which synchronizes two devices at about the same separation as the human eyes. This produces the same effect as someone seeing in three dimensions.

Today, over 1,400 companies and 200 universities use their technology in nearly 40 countries for conferencing, educational experiences, and content creation.

Khan adds: *"If you look at how the world works now, what we are doing is very normal, right? In the old days, it used to be a phone call. Today, it's much better because we see each other. You can get a sense of human connection. A lot of companies are building a new way for collaborating called huddle rooms, which are small spaces. They put the screen on the wall and a table under the screen. In this case, the PanaCast device is perfect because it gives you full 100% outreach and 100% AI coverage, which means not only can you see everybody, but also you can take action and generate big data. It's a very exciting time for us because we're seeing a lot of growth into this market."*

"We want to make this whole technology ubiquitous. That's one of the key things we are working on."

Alejandro Frangi

Alejandro Frangi is Professor of Biomedical Image Computing at the Electronic and Electrical Engineering Department of the University of Sheffield. He is also General Chair at MICCAI 2018 which will be held on September 16-20 in Granada, Spain.

Alex, what's new in this edition of MICCAI?

I'd like to update on two main fronts: one is about new features we have planned for this edition and the other is a bit of an update on the submissions and statistics, because there is also some news on that front.

With respect to the first part, we have a number of features: one is the mentorship program that we have set up. We are still looking for people who might be willing to volunteer to be a mentor but at the moment we have nearly 40 people who have signed up to be mentors so that means that we'll probably be able to support between 50 and 80 mentees, depending on how many people each of them takes. This is all thanks to the work of Marius Linguraru, who has done a fantastic job. This is really new and a lot of people have been asking for it.

We have also made an important improvement on the challenges. This year for the first time we have an online system to submit applications for challenges that walks you through a number of questions that you need to answer, that relate to best practices for a good design of a challenge. So it helps you to sign up, but at the same time it helps you to make sure that you have thought everything you have to think about to design a challenge. So this has gone through and I think it's in the order of something like 13 challenges that we will be running. Kudos to Lena Maier-Hein!

We also have an impressive increase in the number of workshops and tutorials: this year we have about 13 tutorials, while last year we only had four, and this has been a number that has been low for a few years now, so we have tripled that number. Among those there are some tutorials actually organized by industry, including a deep learning tutorial by NVIDIA that is very exciting and people are welcome to sign up for that.

The other thing is we have about 46 workshops. Last year we had 26, so that's also an impressive increase. Fortunately this year we could accommodate most of the requested workshops – we actually had to rent an overflow space in a hotel around the

"We got considerably more companies from the area of artificial intelligence, machine learning, deep learning…"

corner from the conference venue!

Those are the most relevant aspects. The review process has been largely the same the previous years and it's the second year that we have moved to CMT and TPMS a paper matching system for assigning papers. We would like to hear any feedback from reviewers and authors about their experience. And speaking of feedback, one other feature that's new this year and I think has been working very well – and I want to thank the community actually for engaging with it – is that we have created an online suggestion box. So if you go to the webpage of MICCAI you can actually submit comments about pretty much every area of the conference organization, post critiques or suggestions. We got quite of number of them and we have dealt with most of them. So that's a way people can provide additional comments.

I understand that you will give visibility to industry and especially to start-ups. Can you expand on that?

One of the things we noticed in a very clear way, that was very marked in the Quebec edition of MICCAI, is we got considerably more companies from the area of artificial intelligence, machine learning, deep learning beyond the traditional imaging vendors. So that led to a bit of a revision of the sponsoring perspective so we accommodated for the different needs that these different companies have. Given the increase in small companies that are approaching us to sponsor the event - the ones that applied to us were mostly spin-offs but they are already in their third or fourth or fifth year, so they are not that small at this point fortunately - we also realized that there may be a larger number of smaller companies that are particularly keen to recruit good people and at the same time they want to start airing the developments that they are doing.

So for them we created an extra level that has a smaller cost – similar to a bronze sponsor but it gives them a bit of wider exposure, which means that

they also have the possibility to have a booth, so they can showcase their technology and hopefully attract people to them. In any case Terry Peters is the person to contact if you are a prospective sponsor. The other thing to say is that we try to work with each of the companies to tailor. We have certain categories but at the same time we work with them to tailor the products, for some of the main companies like NVIDIA we have been working to develop the ideas of several challenges by them or their tutorials. So please do talk to Terry and we'll try to accommodate in the interests of the industry.

"We got 33% more submissions this year than any year before!"

What is the key factor for success for the next edition of MICCAI? What will make it special?

There is more than one thing, but the one thing that I can tell you now is that we got 33% more submissions this year than any year before. Just to give an idea, last year we got roughly 1000 expressions of interest and 800 final submissions. That was already high compared to previous years. This year we got about 1400 intentions to submit, so that's about 40% more; and about 1060 final submissions, so that is about 30% more than the previous year. So there is a huge interest in the community and I think that gives us a lot of hope as well. At the beginning we were not clear whether this was an increase in the number of papers but not necessarily maintaining the quality – that could have been the case – but I can tell you now that that is not the case. During May we notified the outcome of the papers. Something that has happened this year is that we got a similar amount of direct accepts as the previous year, a similar percentage, which is about 14%. These are papers that did very good in the reviews and will get straightforward access. We got lower than before direct rejects. In the past we used to have 30-40% direct rejects. This year we got about 26% direct rejects, which means that there is a huge amount of borderline papers, nearing about 60% of the papers. We need to take approximately one third of them. So the society is looking at the moment whether this year, given the increase in papers, we want to maintain the acceptance rate. That means that instead of 260 papers we will easily have in the area of 320-350 papers, which also opens the question of what do we do with the orals and what do we do with all the overflow.

A few weeks ago we had approval from the board to actually have for the first time parallel oral sessions, but we will do it experimentally this year on only one of the three days – which will be the Tuesday, the middle day. If the community likes that idea then possibly in the future we can look into doing that throughout the whole conference, which is going to be crucial if we want to grow the conference to the size of a CVPR or NIPS, because otherwise we can't scale it up. So this year we are fortunate enough that we have two amphitheaters in the conference venue and that those are available for us. So the idea will be that we will be able to split the orals of the Wednesday and that will be a day where we're going to put primarily the clinical ready papers. So we are going to possibly divide them by clinical application.

The balance between MIC and CAI was considered last year as being a growing problem. Do you think that things are getting better?

I think that the real issue is not MIC and CAI, it is the fact that CAI at the moment is largely about computer vision aspects in CAI. So a lot of the CAI papers are not really medical robotics, which is what you would also consider core CAI, but is essentially image-guided interventions. So the big challenge for our community is to get back the people from medical robotics. This year one of the key things that is also quite exciting is that the main robotics conference IROS is taking place in Madrid about two weeks after ours. This year another thing that is very important is that we have three very very good keynote speakers, two of which are coming from the computer-aided interventions area. We have Paolo Dario from the BioRobotics Institute of the Scuola Superiore Sant'Anna Pisa; and Kirsten Grauman from the University of Texas – she's more from machine learning domain; and Bradley Nelson who is from ETH Zürich and he also will be talking about robotics and intelligence systems. Basically, we are trying to attract this community back to our conference. The numbers of submissions are actually pretty much aligned with the previous years from what we've seen. We don't know what will happen in terms of acceptance but our experience from previous years is that the real problem is not that we are more selective for one or the other track, it's more that we have different numbers of submissions in the first place so it is clear that we need to attract more submissions. But there's a limit of what we can do there!

Bringing the keynote speakers in from the field is certainly one of the things.

Yes, that's one of the things we can do and have done.

At MICCAI 2017 in Quebec with Julia Schnabel, who is Program Chair at MICCAI 2018

An image from the Multimodal Human-Computer Interaction Workshop, held one day before FG2018 in Northwestern Polytechnical University. First left is speaker Albert Ali Salah (Bogazici Univ.); at the center of the front row is organizer Dongmei Jiang (Northwestern Polytechnical Univ.); photo courtesy of Xiaohan Xia.

Hot off the press:
CARS 2018 Preview
is here!

CARS 2018 PREVIEW

Computer Assisted Radiology and Surgery June 20-23 Berlin

Exclusive Interviews with:
Leonard Berliner
Anna Roethe
Woman in Science: Marta Kersten-Oertel
With presentations by Gabor Fichtinger, Mario Cypko and Matija Ciganovic

In cooperation with
Computer Vision News
The Magazine of The Algorithm community

A publication by
RSIP VISION

Click to read!

AI EXPO WORLD SERIES

DELIVERING AI FOR A SMARTER FUTURE

AI EXPO EUROPE 2018
27-28 JUNE 2018
RAI, AMSTERDAM

AI EXPO NORTH AMERICA 2018
28-29 NOVEMBER 2018
SANTA CLARA CA,
SILICON VALLEY

AI EXPO GLOBAL 2018
25-26 APRIL 2019
OLYMPIA GRAND,
LONDON

GET INVOLVED

+44 (0) 117 980 9020 | enquiries@ai-expo.net

www.ai-expo.net

Co-hosted events

BLOCKCHAIN EXPO WORLD SERIES

IoT TECH EXPO WORLD SERIES

FREE SUBSCRIPTION

Dear reader,

Do you enjoy reading Computer Vision News? Would you like to receive it **for free in your mailbox** every month?

Subscription Form
(click here, it's free)

You will fill the Subscription Form in **less than 1 minute**. Join many others computer vision professionals and receive all issues of Computer Vision News as soon as we publish them. You can also read Computer Vision News in PDF version and find in our archive new and old issues as well.

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

CVPR - Computer Vision and Pattern Recognition
Salt Lake City, UT MEET US! June 18-22 Website and Registration

AI Expo Europe - Conference & Exhibition [see previous page]
Amsterdam, Netherlands June 27-28 Website and Registration

ICIAR - Intern. Confer. on Image Analysis and Recognition
Porto, Portugal June 27-29 Website and Registration

RE•WORK AI in Industrial Automation + DL for Robotics
San Francisco, CA MEET US! June 28-29 Website and Registration

International Summer School on Imaging for Medical Apps
Sibiu, Romania July 2-6 Website and Registration

MIDL 2018 - Medical Imaging with Deep Learning
Amsterdam, Netherlands July 4-6 Website and Registration

ICVSS 2018 - International Computer Vision Summer School
Scicli RG, Italy July 8-14 Website and Registration

MIUA 2018 - Medical Image Understanding and Analysis
Southampton, UK July 9-11 Website and Registration

TMLSS - Transylvanian Machine Learning Summer School
Cluj-Napoca, Romania July 16-22 Website and Registration

Global Summit-Expo on Multimedia & Artificial Intelligence
Roma, Italy July 19-21 Website and Registration

2nd International Summer School on Deep Learning 2018
Genova, Italy July 23-27 Website and Registration

INTECH - International Conf. on Innovative Computing Technology
Luton, UK Aug 15-17 Website and Registration

3DV - International Conference on 3D Vision
Verona, Italy Sep 5-8 Website and Registration

Did we forget an event?
Tell us: editor@ComputerVision.News

FEEDBACK

Dear reader,

How do you like Computer Vision News? Did you enjoy reading it? Give us feedback here:

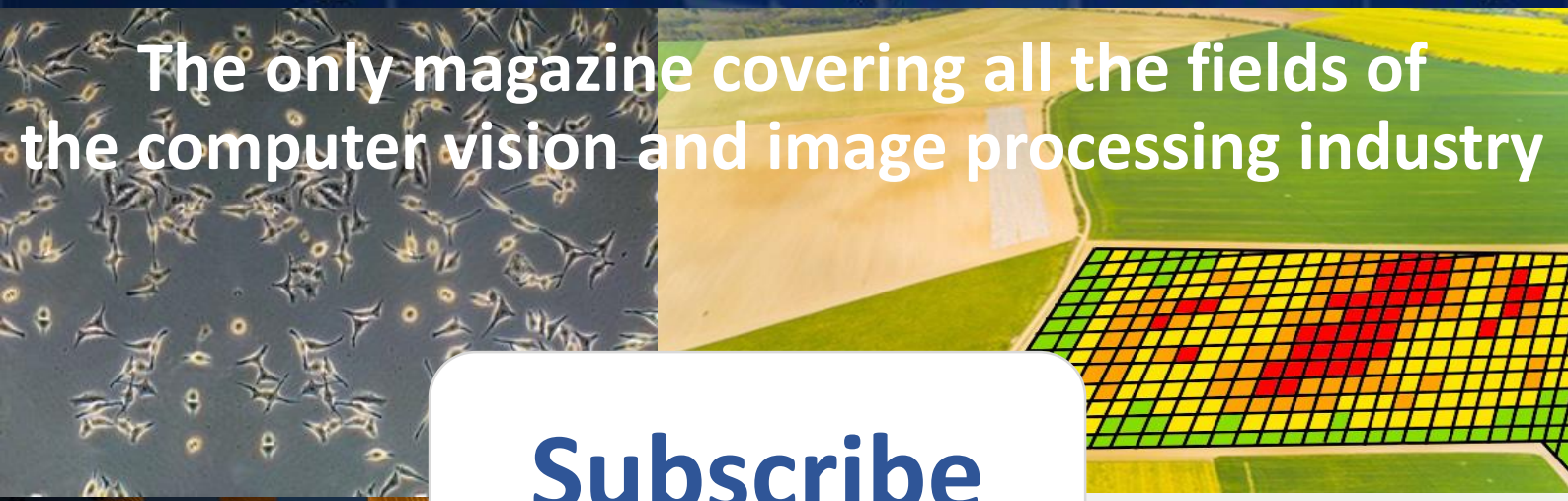
Give us feedback, please (click here)

It will take you only 2 minutes to fill and it will help us give the computer vision community the great magazine it deserves!

Improve your vision with

Computer Vision News

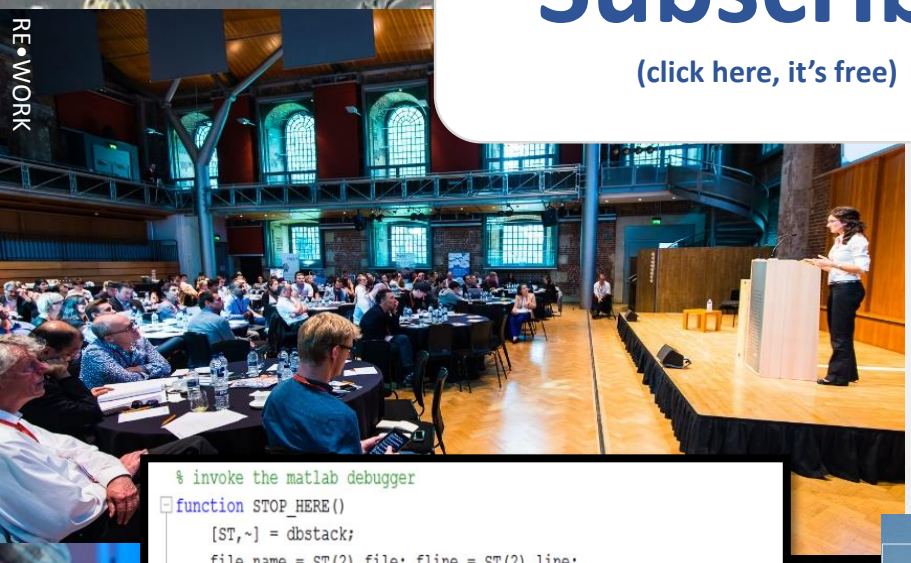
The Magazine Of The Algorithm Community

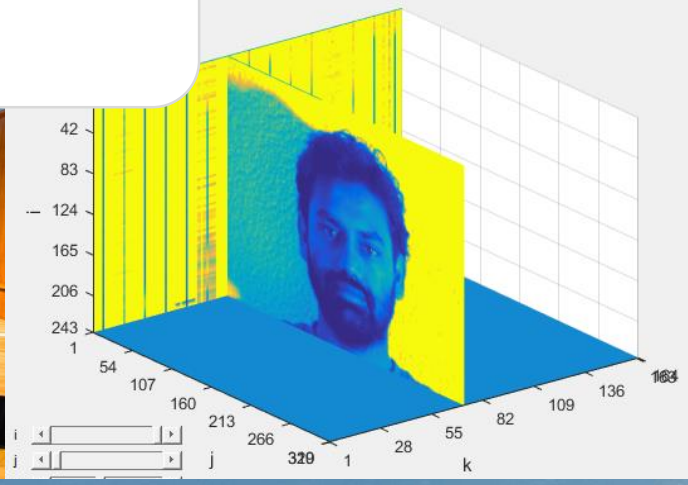


The only magazine covering all the fields of the computer vision and image processing industry

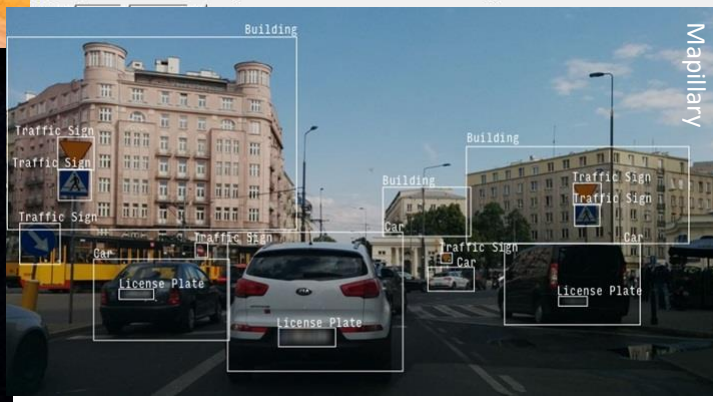
Subscribe

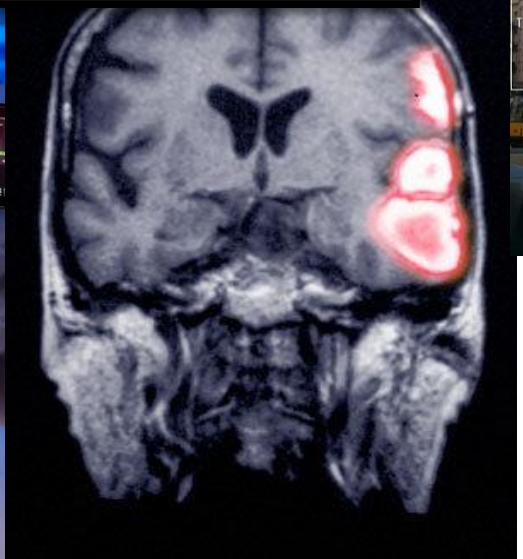
(click here, it's free)






```
% invoke the matlab debugger
function STOP_HERE()
    [ST,~] = dbstack;
    file_name = ST(2).file; fline = ST(2).line;
    stop_str = ['dbstop in ' file_name ' at ' num2str(fline+1)];
    eval(stop_str)
```





A publication by

RSIP VISION