Computer Vision News

The magazine of the algorithm community

A publication by RSIP VISION

INCLUDES BEST OF CVPR 2018 41 PAGES !!!

July 2018

Upcoming Events

Infographics:
Robotics Landscape in the United States

Project Management:
DEEP LEARNING From Plan to Practice

Presentations by:
Pia Bideau
Jingya Wang
Holger Caesar
Emanuel Laude
Ksenia Konyushkova
Yongqin Xian
Juan Caicedo

Exclusive Interview:
John Smith, IBM Fellow
IBM Watson Research Center

We Tried For You:
CVXPY and Colaboratory

Guest:
Jan Kautz - NVIDIA

Women in Computer Vision:
Adriana Kovashka, Nasrin Mostafazadeh and WiCV Workshop

by Assaf Spanier

CVXPY is a dedicated Python toolbox aimed at solving convex optimization. Its user-friendly coding style...

In this week's section we will demonstrate the **CVPXY library**. As a bonus, we will show how you can run this on **Colaboratory**, a new Google environment for running python online that requires no setup to use.

CVXPY is a dedicated Python toolbox aimed at solving convex optimization. Its user-friendly coding style inspired by CVX (Grant and Boyd, 2014) allows the user to express a wide range of convex optimization problems, in mathematically intuitive syntax. CVXPY is open source with GPL license available at http://www.cvxpy.org/

Tool

One programming approach for convex optimization is to use a Domain Specific Language (DSL). In this approach you get to express your problem in a mathematically intuitive way, and then it is automatically converted into the format needed by the solver (e.g., CVX (Grant and Boyd, 2014), YALMIP (Lofberg, 2004), QCML (Chu et al., 2013), PICOS (Sagnol, 2015), and Convex.jl (Udell et al., 2014)).

CVXPY offers a new approach to DSL: on the one hand it's a dedicated convex optimization DSL, and at the same time it's an ordinary Python library, which allows you to take advantage of all the features of Python.

In this article we will demonstrate the use of the CVXPY library to solve **4 classic problems**:

1. **Least-squares problem with box constraints**
2. **Lasso with Python built in multi-threading**
3. **Shortest path**
4. **Image in-painting**

Least-squares problem with box constraints:

We shall start with a simple example, solving the least-squares problem with box constraints. In this problem, the equation b=A*X must be solved with X constrained between two limit values. This problem has applications in many fields such as nonnegative matrix factorization, image reconstruction, and more.

We will find X by minimizing ||A*X – b||^2

Now let's look at the code:

Lines 1-2 import the needed libraries.

Lines 3-7 define the data matrices.

Now we define the problem to be solved by the solver, comprised of:

Line 8 defines the objective function: sum of squares of A*X-b;

Line 9 defines the box constraints.

Line 10 instantiates a problem object with the defined function and constraints.

Line 11 runs the solver to solve the problem instance.

Line 12 prints the result.

```
1. import cvxpy as cp
2. import numpy as np
3. m = 30, n = 20
4. np.random.seed(1)
5. A = np.random.randn(m, n)
6. b = np.random.randn(m)
7. x = cp.Variable(n)
8. objective = cp.Minimize(cp.sum_squares(A*X - b))
9. constraints = [0 <= x, x <= 1]
10. prob = cp.Problem(objective, constraints)
11. result = prob.solve()
12. print(x.value)
```


Lasso with Python built in multi-threading:

The Lasso statistical solver was originally formulated to solve the least-squares problem. It uses the same sum of squares error, but adds a regularization term called the L1 penalty, commonly weighted by some factor. In order to implement

this weighted factor in CVXPY you need to use the parameter function, as you can see in the code below.

```
x = Variable(n)
lambda = Parameter(sign="positive")

 error = sum_squares(A*x - b)
L1 = norm(x, 1)
prob = Problem(Minimize(error + lambda*L1 ))
```

The value of lambda isn't known *a priori* as it is different for every problem. Therefore, a common approach is to evaluate several values for the lambda parameter. To do this efficiently, we will take advantage of the fact that CVXPY is a Python library to use Python's multithreading capabilities to run parallel instances each with a different lambda value. Obviously, this capability is something that non-embedded DSLs could not provide.

The get_x function below is a function that each thread will run -- it gets one lambda value and invokes a `solver` for finding the solution (x).


```
def get_x(lambda_value):
        lambda.value = lambda_value
        result = prob.solve()
        return x.value
```

```
gamma_vals = numpy.logspace(-4, 6)
pool = multiprocessing.Pool(processes = N)
x_values = pool.map(get_x, gamma_vals)
```

Shortest path:

Now we will solve the problem of finding the shortest path in a weighted graph. In this example, we will see how to use python classes to define the graph. The vertices and edges of the graph will each form a class.

The `Vertex` class has two functions: `constructor` and `prob`. The `constructor` defines the variables to be optimized, and the `prob` function defines the specific constraint for the vertices, which is minimizing the net flow, i.e., minimizing the sum of weights passed through between source and sink.

```
class Vertex(object):
def __init__(self, cost):
            self.source = Variable()
            self.cost = cost(self.source)
            self.edge_flows = []
def prob(self):
            net_flow = sum(self.edge_flows) + self.source
        return Problem(Minimize(self.cost), [net_flow == 0])
```

The `Edge` class has three functions: `constructor`, `connect` and `prob`. The `constructor` initiates the cost (weight) of the edge, the `connect` method forms the graph by defining the vertices of the given edge, the `prob` defines the constraint to be optimized using the CVXPY `Problem` function.

```
class Edge(object):
            def __init__(self, cost):
                        self.flow = Variable()
                        self.cost = cost(self.flow)
            def connect(self, in_vertex, out_vertex):
                        in_vertex.edge_flows.append(-self.flow)
                        out_vertex.edge_flows.append(self.flow)
            def prob(self):
                        return Problem(Minimize(self.cost))
```

Finally, after the entire graph has been constructed, the following code defines the global optimization constraint for finding the shortest path, i.e. minimizing the sum of weights passed through between source and sink.

```
prob = sum([object.prob() for object in vertices + edges])
prob.solve()
```

Image in-painting:

The goal of image in-painting is to reconstruct a corrupted image. We will represent the reconstructed image as *U* with size [*m* x *n*]. The corrupted pixels of *U* are defined by the index matrix (`corr_index`).

The reconstruction *U* is found by minimizing the total variation of U, subject to matching uncorrupted pixel values. We will use the *L2* total variation, defined as:

$$\mathbf{tv}(U) = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \left\| \begin{bmatrix} U_{i+1,j} - U_{ij} \\ U_{i,j+1} - U_{ij} \end{bmatrix} \right\|_2.$$

Tool

The code below demonstrates the image reconstruction process:

Lines 1-2 import the needed libraries.

Line 3 loads the corrupted image.

Line 4 loads the a `0-1` matrix called `corr_index`, which has a value of 1 for uncorrupted pixels and 0 otherwise.

Line 6 defines a CVXPY variable class the size of the image -- the value field of this class holds the in-painted image.

will hold the in-painted image.

Line 7 defines a total variance minimizer.

Line 8 defines the constraint that all pixels with a `corr-index` value of `1` must remain unchanged.

Line 9-10 define the `problem` instance and run the `solver` to solve it.

Once the solver converges, the matrix `U.value` contains the in-painted image.

```
1.  import matplotlib.pyplot as plt
2.  import numpy as np
3.  corr_img = np.array(Image.open("data/lena512_corrupted.png"))
4.  corr_index = Image.open("data/lena512_corr_index.png")
5.  rows, cols = Uorig.shape
6.  U = Variable(shape=(rows, cols))
7.  obj = Minimize(tv(U))
8.  constraints = [multiply(corr_index , U) == multiply(corr_index ,
    corr_img )]
9.  prob = Problem(obj, constraints)
10. prob.solve(solver=SCS)
```

As in the least-squares problem we define the objective function we want to minimize, which is a sum of squares of A*x-b

Last but not least, colab!

All you need to do is go to colab.research.google.com/, open a new notebook, write and run your code.

However, two things will be missing:

1. The CVXPY library, which is not a built-in Python package.

   You need to install it, which can be done by the following command:

   !pip install cvxpy
   (Yes, you can run this pip command even though this is a web service).

2. Obviously, images and other files to work on will be missing.

The following hack can be used to overcome this issue...

The following hack can be used to overcome this issue, the code snippet below can be pasted into a colab cell, when you run it a pop-up window will enable you to upload your needed files.

```
from google.colab import files
def getLocalFiles():
  _files = files.upload()
  if len(_files) >0:
    for k,v in _files.items():
      open(k,'wb').write(v)
getLocalFiles()
```


Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editor:
Assaf Spanier

Publisher:
RSIP Vision

Contact us

Give us feedback

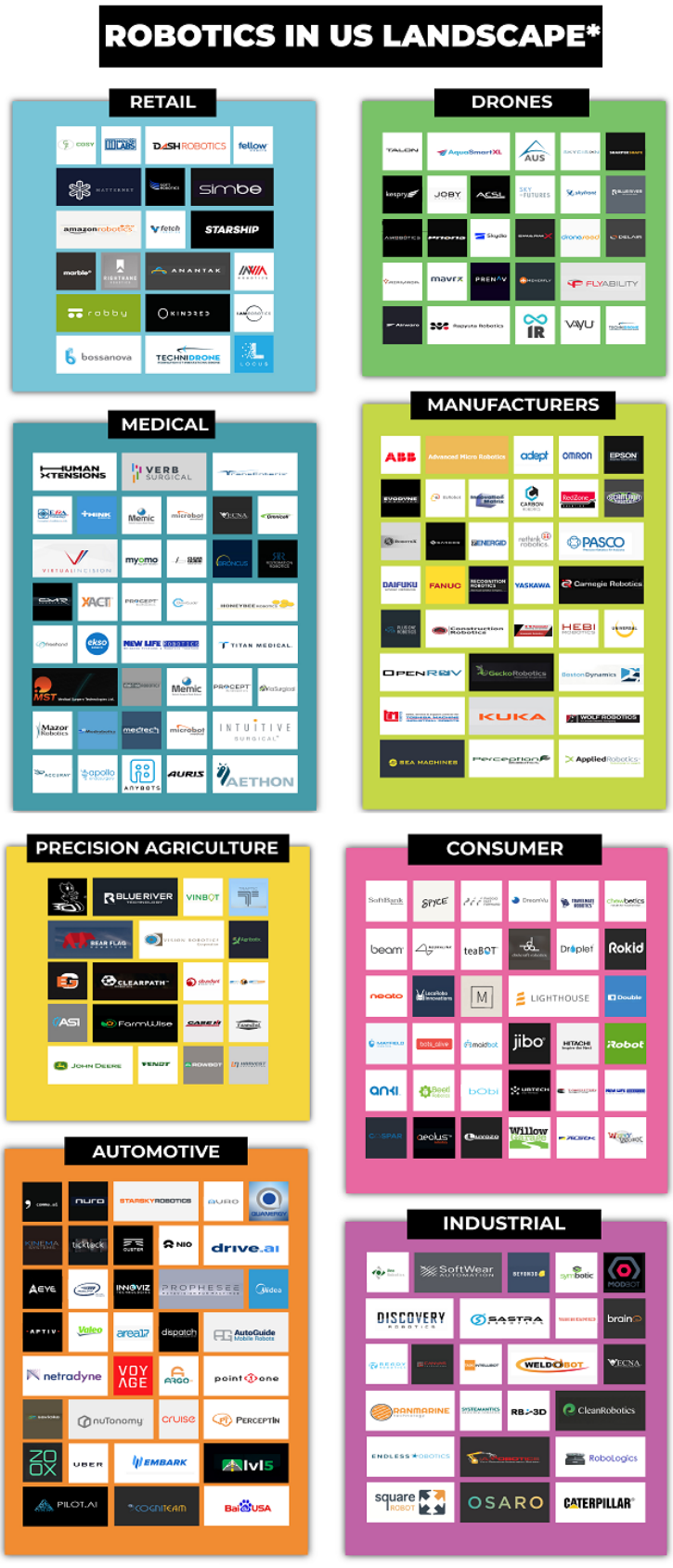
Free subscription

Read previous magazines

Follow us on:

Robotics Landscape in the United States

We collaborated with RE•WORK to highlight leading companies in the robotics space who are working to solve challenges across sectors in the United States. Take a look at the breakdown of companies employing robotics in their industries**. For a full size view with all the names, please click on the image**:



Deep Learning: from plan to practice

RSIP Vision's CEO **Ron Soferman** has launched a series of lectures to provide a robust yet simple overview of how to ensure that computer vision projects respect goals, budget and deadlines. This month we learn about **Deep Learning: from plan to practice**. It's another tip by **RSIP Vision** for **Project Management in Computer Vision**.

In many situations, the project manager knows very clearly that Deep Learning is the optimal solution to a segmentation problem. Still, it is a long way until this vision is put into practice.

We want a ground truth, feed it into the training system and hope that results will be good.

Let's take medical segmentation as an example: we often want to segment organs, bones or pathologies. The general framework is very clear: we want a ground truth, feed it into the training system and hope that results will be good. However, getting a valid ground truth for medical segmentation is far from being simple.

First, who is going to do the annotation? An expert radiologist or laymen who are given basic training? Due to the complexity and diversity of the human body, even expert radiologists may find that judgement for borderline cases is not clear-cut, resulting in a wide intra- and inter-observer variability.

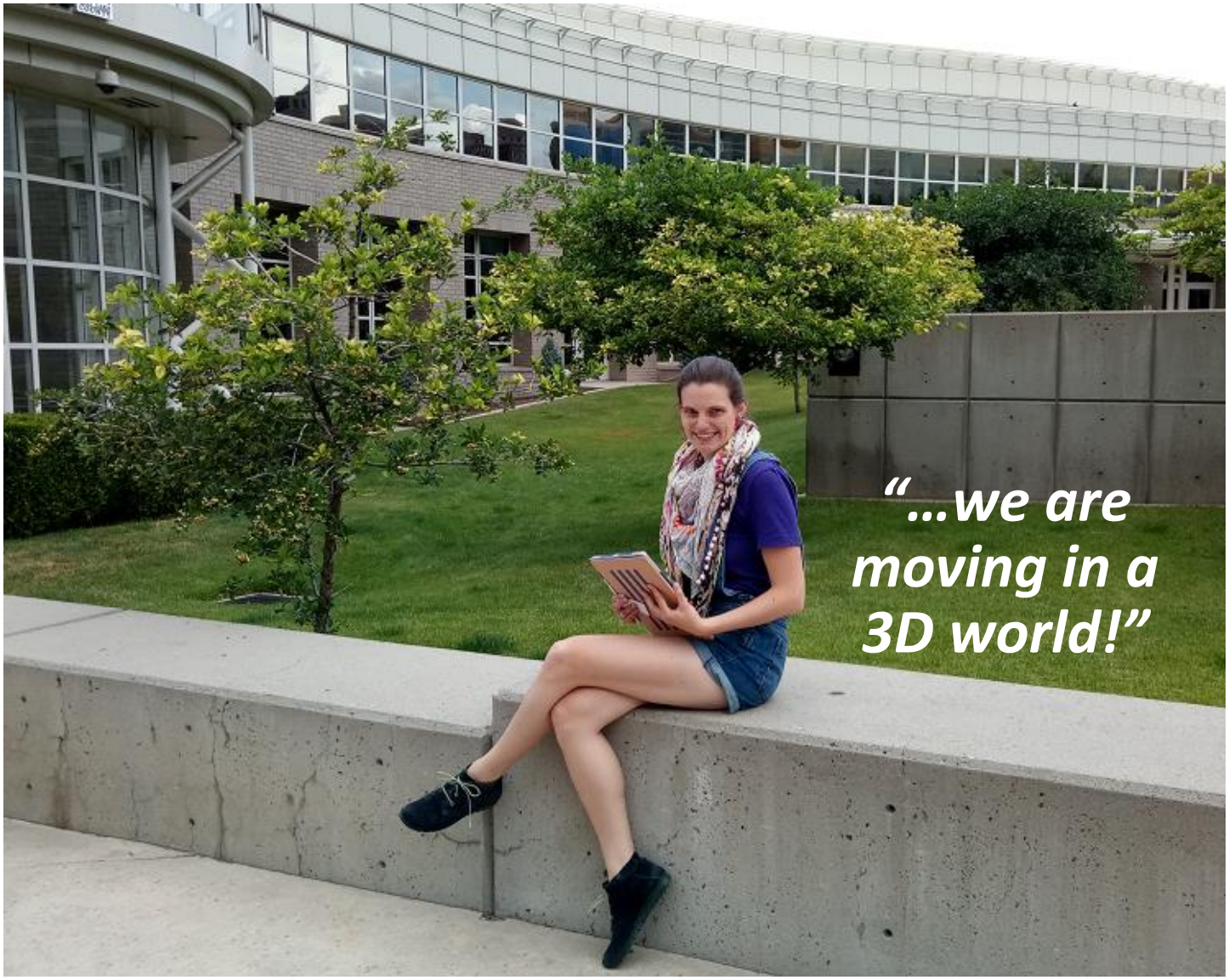
Another important aspect is quantity: it would be better to have thousands of images, but in practice this is not always possible, especially when precise annotations are required. We also need to understand what we have in those images: when a dataset includes many unrelated items or complex cases with multiple pathologies, its quality may suffer as a result.

The quality itself depends on the resolution of the images coming from the different modalities (MRI, CT, etc.) in all their types. The project manager needs to make sure that the data received is in line with what the project requires. When data is gathered from big databases, there are many DICOM files which need to be sorted out to find what is relevant and what is not.

Extra tools are needed to facilitate the definition of the ground truth.

Ground truth might be difficult to obtain for other reasons as well: some organs are very complex, like the airways leading to the lungs, and many times radiologists lack the time and patience to do this tedious task through a large dataset of images. In that case, extra tools are needed to facilitate the definition of the ground truth.

We conclude saying that the first part of the project needs a thoughtful planning in order to achieve the ground truth.



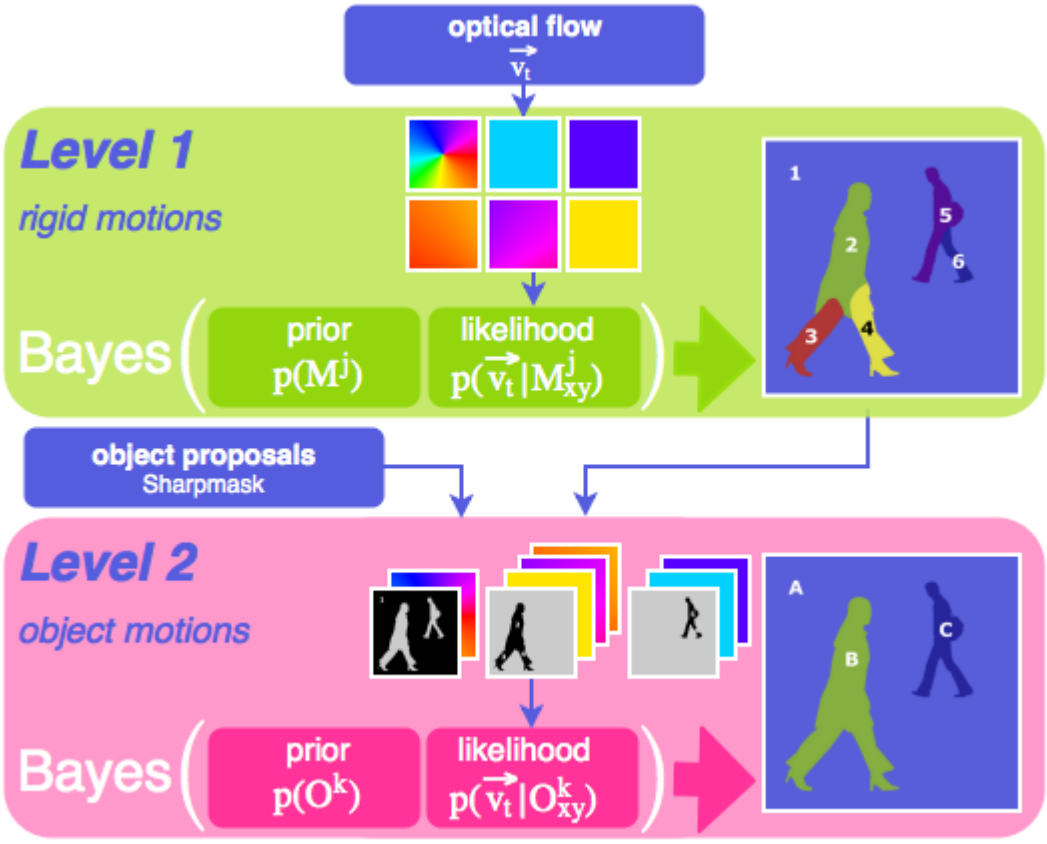
Pia Bideau is a PhD student at the University of Massachusetts at Amherst. She spoke to us ahead of her poster presentation.

Pia's work is about **motion segmentation**, which is the **segmentation of independently moving objects in a video**. Both the camera and objects can be moving – for example, cars, humans, animals – and we want to detect all those objects that are moving differently to the camera.

Pia explains that motion is very important for human vision, for scene understanding, and to get to know the world we are moving in. When a person is walking through a forest or just walking along the street, it is easy to detect objects as soon as they move. A moving car will draw your attention. A squirrel will run up a tree. You will detect them immediately just because they are moving.

The challenge of this work comes because **we are moving in a 3D world**. We have objects that are very close to us and objects that are very far away. Objects that are close might move more on the image plane than objects

"If you solve all those problems together, they can support each other!"



that are very far away, but actually they do not necessarily move. A tree which is close to us is displacing a lot, and a tree which is far away is moving just a little, but none of them actually move. Only the person or the animal is moving. There are a lot of connected topics like object segmentation, depth estimation, and motion segmentation which we have to consider all at the same time.

Pia explains: "*We look at how the camera is moving. The camera can be translating, the camera can be rotating, and an object can be moving. All those three ingredients are part of the motion in the world which creates the optical flow. We first estimate the camera rotation and we subtract this off the optical flow, such that we only have a translational flow field, which has the camera translation and the object motion. If you just look on the direction of the optical flow, not on the magnitude, then you can see which objects are moving into a different direction. We developed a system which is considering the anti-optical flow and the derived flow likelihood, which computes how likely is a motion model given the observed optical flow.*"

In terms of next steps, Pia tells us that she will focus more on solving all those problems at the same time – **depth estimation, motion segmentation, object recognition, as well as estimating optical flow**. She predicts that if you solve all those problems together, they can support each other.

Pia Bideau presented her work on a poster at CVPR on June 19.

"The reason we transfer the knowledge in the attribute space is because normally for the human re-identification, the ID labels from different domains, different data sets, are independent – they don't have overlaps – but in our study, we want to transfer the knowledge."

Jingya Wang is a final year PhD student at Queen Mary University of London. She spoke to us ahead of her poster presentation.

Jingya tells us that her work is about **jointly learning the attribute and identity of person re-identification under an unsupervised setting**. Matching the person from non-overlapping camera views in different locations. Most existing person re-identification works under a supervised setting that needs a large amount of human annotation. Her work focuses on the transferable unsupervised setting, transferring knowledge from the existing dataset to the unseen new dataset.

The work focuses on the jointly-modelled global identity and the local attribute information, because there is a heterogeneous problem for multi-task learning, so she proposes a progressive knowledge fusion mechanism by encoder-decoder networks for intermediate space that progresses transfer for the domain adaptation.

Jingya explains that previous re-identification work has focused on the unsupervised setting that needed a larger amount of pairwise data. In this work, they want to jointly learn the attribute and identity space to get the better feature extraction for the person re-identification.

"This is an open-set recognition problem."

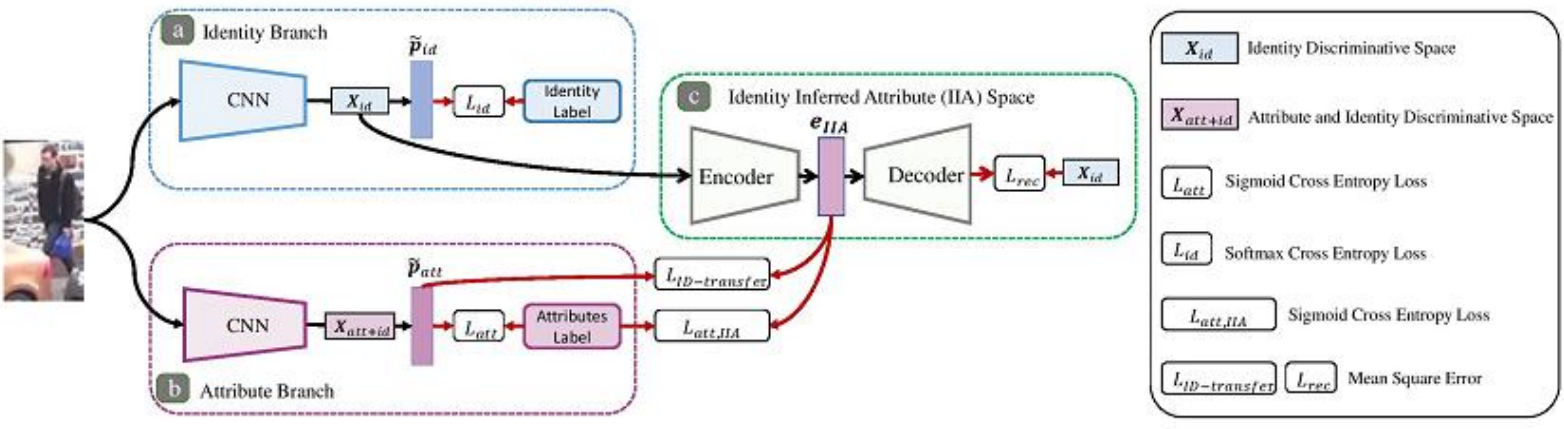
In terms of challenges, she says that because this work uses surveillance video, it's not like most computer vision works that use well annotated or well-structured annotation. **This is unsupervised**. Also, because it is surveillance video, it has poor image quality, the background is normally blurred, and sometimes there are different view conditions because of different camera locations. She adds that **person re-identification is more challenging than facial recognition**.

Jingya explains more about the computer vison methods used: "*The first one that we introduced was progressive knowledge fusion mechanism by encoder-decoder intermediate space. Second, we proposed a normal domain adaptation method, but it should be the consistency scheme, and because of this we can transfer the knowledge in the attribute space. The reason we transfer the knowledge in the attribute space is because normally for the human re-identification, the ID labels from different domains, different data sets, they are independent – they don't have overlaps – but in our study, we want to transfer the knowledge. This is an open-set recognition problem. In our work, we want to introduce attribute space, because it's more uniform space for transferring the knowledge, because they share the most common description.*"

The next step is to deploy a better domain adaptation model.

The next step is to deploy **a better domain adaptation model**. Being this an open-set domain adaptation problem, it is very challenging in the re-identification setting, and Jingya would like to explore more on that aspect.

Jingya Wang presented her work on a poster at CVPR on June 19.



Nasrin Mostafazadeh

Nasrin Mostafazadeh is a senior AI research scientist at Elemental Cognition where she works on the next generation of AI systems that not only comprehend stories, but also explain themselves.

I am a senior research scientist at Elemental Cognition, which is this fundamental AI research startup that is focused on pushing the boundaries of AI forward through to AI systems that can explain themselves, collaborate with humans via dialogue, and machine reading to have a better shared understanding of the world.

You are here at CVPR to give a talk and just for one day. What is your relationship with computer vision?

I would say that I have a complicated relationship with computer vision.

We love complicated relationships!

[*laughs*] I'm joking.

[*laughs*] I'm not! [*both laugh*]

I got exposed to working on vision and language when I was at Microsoft Research for about a year. There they had an ongoing research project on a line of research which was mine which was narrative structure understanding and story understanding, which they wanted that to be multi-modal. Hence, through that, I got involved with vision. Before that, I had never done any vision and language work. That was the start. After that, I built up on top of that line of work. I kept working on different vision and language projects, which involved deeper language understanding and common sense reasoning, which is the field of AI that is really close to my heart. I'm just super passionate about doing something on common sense reasoning. All the vision and language work that I've done has been trying to push that forward. That's the complicated relationship. I'm not part of the vision community, I would say. I'm a part of the vision and language community. I was very happy to come to CVPR after all and become a part of the community. So that's the complicated relationship. [*laughs*]

You mentioned in your talk that ten years ago there were concepts that were difficult to explain to the community, and today it's much easier. Can you tell our readers what that was and why you think that changed?

So actually I made a point that the change happened and the kind of revolutionary effect that deep learning had on the community happened in other fields. I would characterize different AI challenges that we have in two categories. One being pattern recognition. The other being perception and reasoning. Deep natural language kind of falls beyond pattern recognition and into a task that requires reasoning and common sense. Anything like that, which deep natural language understanding is one of them, requires

lots and lots of knowledge and has not been really revolutionized in the past decade, I would say. Whereas a lot of pattern recognition such as image processing and speech recognition have been revolutionized, which is the example that I was making in my talk.

What made people change their approach? The deep learning revolution or something else?

What has happened in the vision community is the abundance of data which has enabled existing algorithms which have been around for decades and decades such as deep learning to work actually. On top of that, we have much more computing power such as a lot of GPU's which many big companies have access to these days. They have given rise to the so-called deep learning revolution, which has definitely made a huge change in pattern recognition problems.

Regardless of your excellent accent, I understand that you are not American-born. Where do you come from?

Oh, wow! My accent is excellent then? [*laughs*]

I think so!

Oh, thank you! I'm Iranian. I came to the US five and a half years ago. I was born and raised in Iran. I went to school in Iran.

When did you discover that you would like to study language?

Ah well, the story goes all the way back to high school. Actually, I started working on natural language processing and natural language understanding in high school. The story that I told about how I got into natural language understanding from robotics was in high school. I started working on RoboCup competitions when I was like 15 or so. Then through that classical natural language example that I just told you, I really was excited to work on a problem that people called AI-complete. I can tell you the story of how I got exposed to natural language.

Please do!

It's just a bit long.

[*laughs*] We like long stories!

Here's the story. We were me and a couple of my amazing friends, we had this robotics team, as I said. Because of that team, we wanted to compete in international RoboCup competitions. It meant a lot of hard work. For about more than a year, maybe a year and a half, we just pushed on that agenda. We had a team, and we wanted to win. Through that, we didn't go to any of our classes in school. We had this amazing school that let us do this. The end of that part is that it was successful. We became actually the second in the world,

BEST OF CVPR 2018 SALT LAKE CITY • JUNE 18-22 Tuesday DAILY

which was a dream at the time. We were kids coming second in the world.

Could you believe it?

We couldn't believe it! As I said, we skipped our classes. The year after, when I came back, we had this physical education teacher. She gave me a hard time. She said that I won't let you pass this course because you haven't shown up for about a year. She handed me this stack of papers for translation. They were English documents that she just wanted to be translated into Farsi. I was the manual labor that she had. She said: "*Unless you translate these documents, I won't pass you!*" Anyways, I take these documents home. Days pass, and I think "*This is too much!*" Then I was like, I've spent the past few years of my life just sitting in front of a computer and programming. What if there was a program that could do this for me? Then I looked up how to automatically translate English to Farsi, something like that. Then I remember the Wikipedia page was maybe the third hit with machine translation. That's how I started knowing that natural language processing was a thing. I didn't know that the field existed. Through that, I got exposed to natural language processing. I got interested in it.

Then it turns out that there is this task that is incredibly challenging for a system, but is super easy for a four- or five-year-old kid. I just chose to work on that moving forward. That's my story of how I got into it.

How did it develop over time?

It's kind of funny. I was one of these people who stuck with what I was doing and didn't let go. There are people who have done a minor in psychology or a major in philosophy and then it turned into computer science. I've been doing this forever. I've been doing this for many years.

Did you ever have any second thoughts?

I have been lucky, I would say, that I knew what I wanted to do. At the same time, I know there's a downside, right? I've never dived deep into history or geography or other topics that I know I don't like, but maybe I should have pursued. [*laughs*]

But you don't like them?

Maybe, but I've passed generic breadth courses.

So what can you tell to all of the PhD students who are having second thoughts about what they are learning?

There are two things in life, right? I would say, you either pursue something that you are great at or you pursue something that you are super passionate about. You're extremely lucky if those two collide.

And if they're useful, that's even better!

That's another thing. You think, you know, at this point you want to maximize your chances of getting a job in computer science. [*both laugh*]

The truth is, there should also be a demand,

right? You're the luckiest if there is also a demand for what you are doing. I would say that it's about balancing. Maybe it's an art of how to find something that you enjoy and, at the same time, you are good at. Sometimes you have to pay the price of doing something that you love so much that you can't let go. I think if you have second thoughts, you should really try to talk to many people in that particular doubtful area that you are at. I found that the most useful thing whenever you are going through something, you are not alone. So many people have experienced that. Just talk to people, and collect your data points. Go about making a meaningful and significant decision by having prior knowledge. It's okay to doubt. It's absolutely fine to not to know what you want to do. Everyone around me has been like that. It's just that I have been super lucky that I knew what I wanted to do from early on.

Did you see other students working with you that had second thoughts?

Everyone… I've had so many close friends that have dropped out of their PhD. I've had friends that have dropped out of college. I've had friends who, right now, they got a PhD. They got a job, but say that they don't like computer science. This happens all the time. I do try to do my best to help people make the right decisions for themselves, but this is very personal, right? Even the idea of should I get a PhD or not in the AI world is very personal. It really depends on what you want to do, if you're super passionate about research.

It sounds like you are extremely passionate about this subject.

Yes!

What is your second biggest passion?

In life? If I wasn't doing AI research, in an alternate universe, or maybe when I somehow get to a position where I could do this, I would have been Anthony Bourdain. I'm so sorry that he's gone now.

Ah, cooking…

I always said that I would have been a chef, traveled, and worked… and explore history and culture through food and understood humanity through food.

Tell me something about Iranian people that we don't know.

In the US, people don't know much. I think through the lens of media, and this is changing a lot with a younger population of Americans, it is a totally different perspective. I think on average, not in major cities of the US, but people have this image of Iran.

I'll tell you what I know: warmth and hospitality.

That's a good characterization. I think you know this too: women in Iran, in terms of university graduates, actually there are more women than men. They have a majority. There are higher percentage of women.

We see it also in our community.

Yes, in vision, I'm sure. There are many Iranian women. That's something that

BEST OF CVPR 2018 SALT LAKE CITY • JUNE 18-22 Tuesday DAILY

people have asked me: were you able to study the same way that men did in Iran?

Well, our magazine has interviewed many impressive women from Iran.

I think you already know too much!

Sorry! [*both laugh*] What is the thing that you regret that you have in Iran that you don't have here?

Family!

They are there?

Of course. Iranians, almost all of them, or a lot of them, get single entry visas. Then they become students in the US. Because there is no US embassy in Iran. It's a huge risk to leave the US after you've come here to obtain a new visa. People end up staying here for long, long times without seeing their families. Also, getting tourist visas for families is very hard, if not impossible. I was lucky. I got a multiple entry for two years. I could go back to Iran twice, but I was a huge exception. Now, I've been here for four and a half years without being able to present my work at different conferences when they were outside of the US or to see my family. There are many other Iranians who have had it much worse than I have had. They have been here for eight years, nine years without seeing their families. It's just cruel, right?

It's super hard. People keep up with it because they have to, but it's really, if you think about it, it's just terrible. It's really hard.

"It's just a blip..."

What about ups and downs?

I'll tell you this story. Once I read about a financial recession, decades ago. Many people really stressed about their job prospects. They thought that everyone would go down, and they would lose their houses. Now, you zoom out the graphs of the economic growth throughout the years, and look at that point in time: it looks like a little blip. A tiny blip, which at the time seemed like a disaster. I would say that in life, you have so many such blips. The only thing you can do is to make sure that when you are living it, be mindful of the fact that when you zoom out, it would have been nothing. Try not to stress. It's hard when you are in the middle of something bad that is happening to you. It's so easy to lose sight of the big picture, things that you have, and how much things really matter. Just saying "*shake it off*", doesn't work. I think sometimes reminding yourself of the fact that however desperate, helpless, and sad you feel, it's just a blip in your life. That can maybe help you to get through it more easily.

How long will it take until an AI can replace me in doing this interview, with laughs and all?

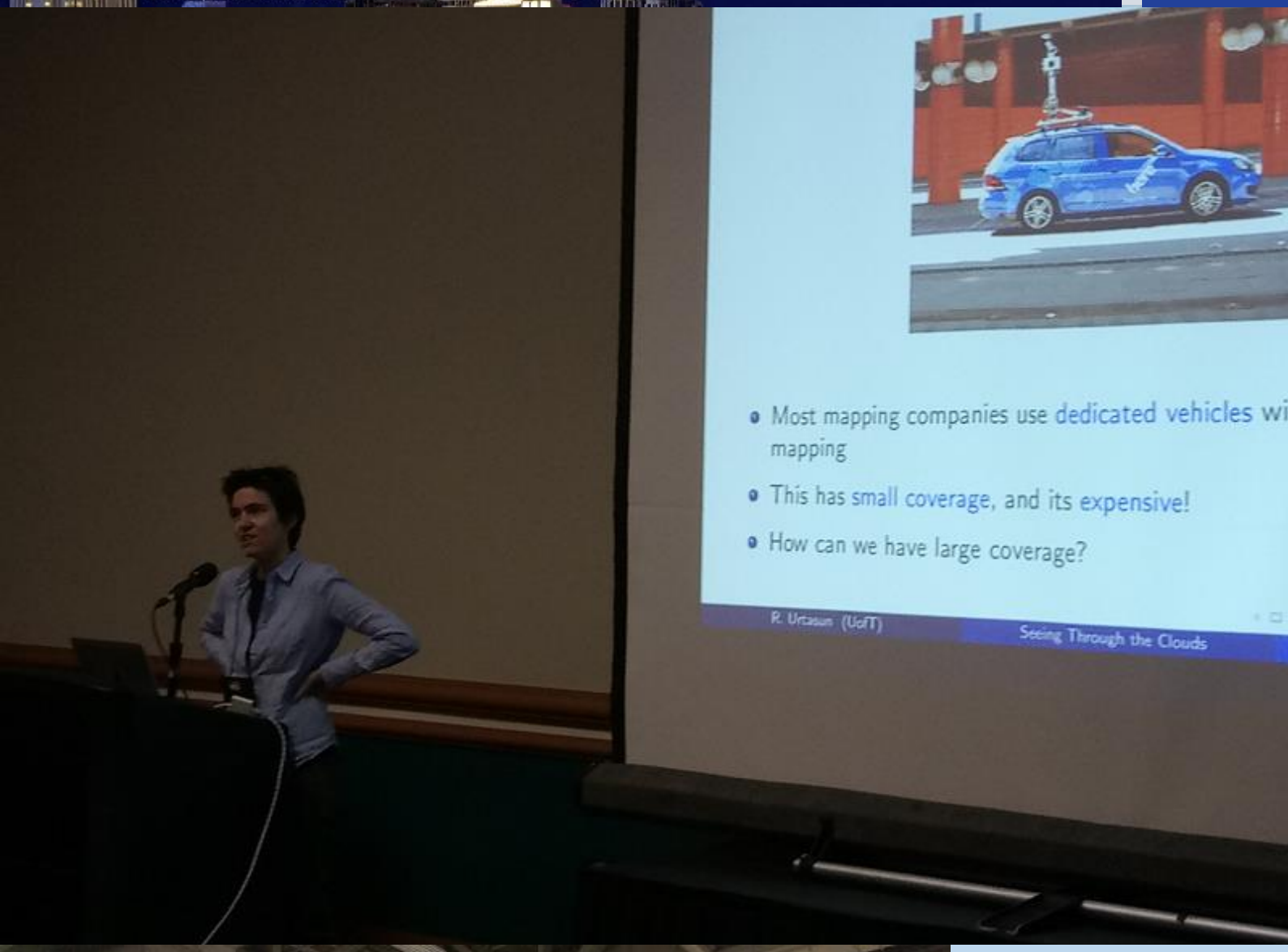
The full interview with all the breaks and pauses, in the same way that you did it? I will say 30 years from now!

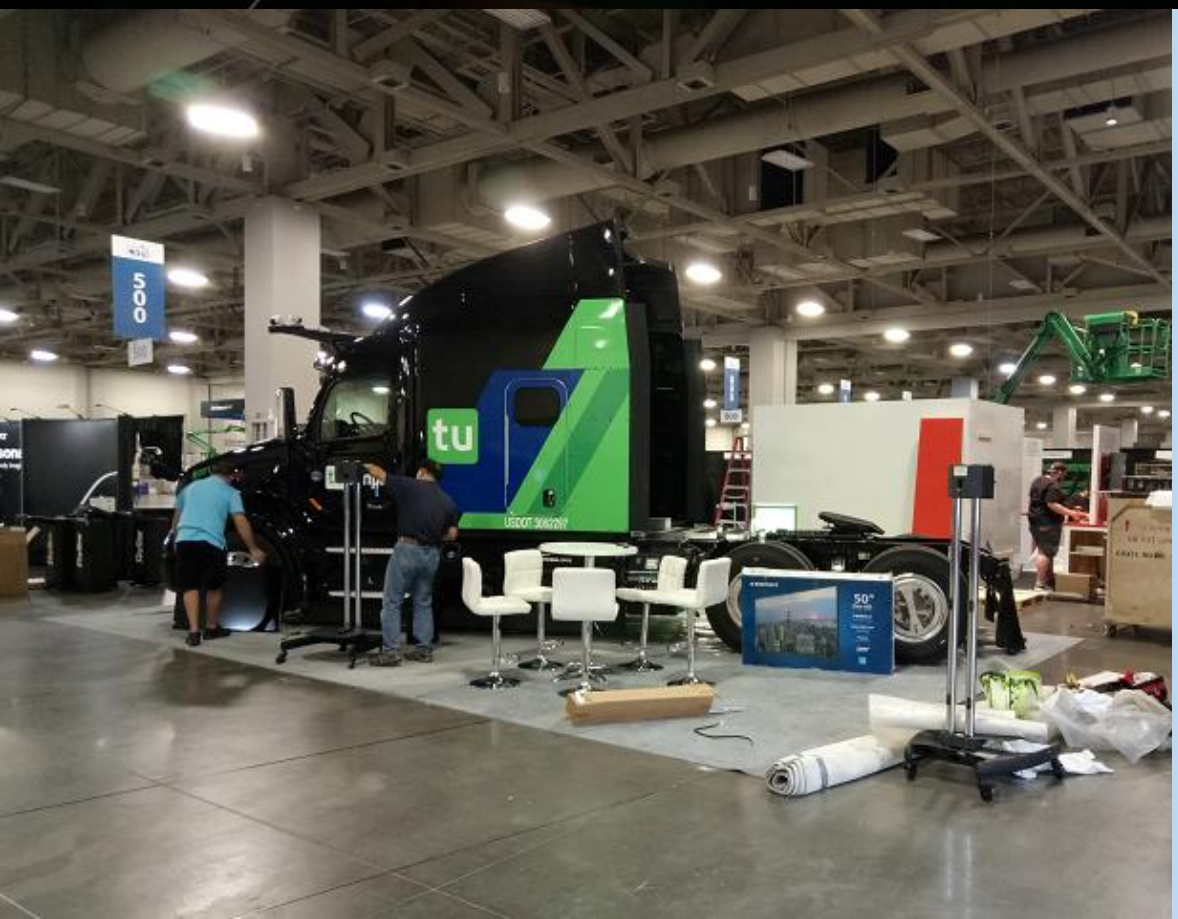
Jitendra Malik during his talk at the VQA workshop.

BEST OF CVPR 2018 SALT LAKE CITY • JUNE 18-22 Tuesday DAILY

Top image:
Mehdi Moradi (IBM) presenting at the Medical Computer Vision and Health Informatics workshop. Standing left, co-organizer Tal Arbel.

Bottom image:
Tal Arbel introducing Alejandro Frangi. Tal was Program Chair at MICCAI 2017 and Alex will be General Chair at MICCAI 2018, later this year.





Top image: Raquel Urtasun (Uber ATG) presenting at the DeepGlobe workshop.

Bottom image: The Expo is being built. This is an autonomous truck developed by TuSimple

BEST OF CVPR 2018 SALT LAKE CITY • JUNE 18-22 Wednesday DAILY

Jan Kautz - NVIDIA

Jan Kautz is Senior Director of Visual Computing and Machine Learning Research, which effectively means leading the research team at NVIDIA called Learning and Perception.

Jan, I understand that you are opening a new lab in Toronto, under the leadership of Sanja Fidler.

Yes, we have been engaged with the community in Toronto for a while now and have an office there. We wanted to expand our deep learning and machine learning research and Sanja is a very stellar individual who we are lucky to work with, so we decided that it would make sense for us to open a research lab in Toronto given all the talent there.

As far as I know, NVIDIA is a hardware company. Why do you need to hire the best software talent that is around?

We shouldn't call NVIDIA a hardware company any more. We make the hardware, but also software building blocks that are used by our customers and partners. We do both hardware and software, which makes us a platform company. Software has become a very important part of NVIDIA's business. That is why we do a lot of research on the software side. We still do research on the hardware side as well, but the software side is becoming more important.

That is why you are taking the best talent from us! [*both laugh*] How many employees you have in total today?

We have roughly 12,000 employees in total worldwide.

Can you tell me about your own work?

I am part of NVIDIA Research. Its goal is to push the state-of-the-art forward, so that the research has an impact on the community, but of course, ultimately also on NVIDIA. We try to develop new technologies through research that will help NVIDIA directly or one of our partners at some point in the future. NVIDIA works with a lot of partners and often our research is helpful to them. My group specifically, the Learning and Perception Research Group, do a lot of work in computer vision and machine learning. On the computer vision side, we address a lot of the computer vision stack. Going from low-level computer vision, such as image processing and computational photography, all the way up to high-level computer vision such as scene understanding and video understanding. We try to touch on all of computer vision and address areas that we believe can have an impact on us.

Why do you have such a strong and active presence at CVPR? You are also a diamond sponsor of the event.

Computer vision and machine learning are two very important parts of NVIDIA nowadays, for obvious reasons. Computer vision due to self-driving cars and robotics - the two dominant application areas of computer vision. Machine learning in general, of course, is very important.

How many papers are you presenting at CVPR this year?

We have been very successful this year, with 14 full papers at CVPR. 11 of them are in collaboration with my group. You don't want to talk about all 14 of them?

Yes. You have 13 seconds for each!

[*both laugh*] There's the work which we call Super SloMo, a method for taking standard footage that's recorded, say at 30 frames per second, and we can slow it down to an arbitrary rate using a neural net. It works very well. We have trained it in a self-supervised manner on lots of videos. Now we can insert or hallucinate intermediate frames to slow down a video; that works really nicely.

What problem in the real world will that solve?

One of the use cases is aesthetics – people like to look at things in slow motion. There are also some professional use cases. For example, if you are a professional ballerina or athlete and you want to see every nuance of your form, you could do that with this method by slowing it down and looking at it in slow motion. Another paper, in collaboration with an intern from UMass Amherst, is called SPLATNet. That is also very interesting and has won the best honourable mention award. To generalise, it is about dealing with sparse, high-dimensional data. In particular, we looked at point clouds and how you process point clouds. Looking at processing not just the point clouds themselves, but also point clouds together with images. Say, you have a LiDAR scan as well as RGB images, you can work or process them jointly. For instance, to do semantic segmentation, which is really quite neat.

Click to view NVIDIA video

Can you tell us something nice about NVIDIA that the public doesn't know?

[*laughs*] NVIDIA is a very open company internally, which is great: we share a lot of information across teams. It is very collaborative, which is probably not visible from the outside, but it is a very fun place to work as a result.

That sounds awesome. What can you share with us about what's coming up at NVIDIA in the future?

You can expect to see research expand. We are eager to push more on the research front and be visible at all the major conferences. It is important to us to be part of the community and really give back to the community as well.

I am very happy to be at CVPR and it is great to see so many people attending and there being such a strong interest in computer vision. The research that people are doing and showing here is great to see.

"Things are objects with a clear shape and size. Stuff are all the other amorphous background regions that don't have a clear shape and size."

Holger just finished his PhD at the University of Edinburgh under the supervision of Vittorio Ferrari and is now a research scientist at nuTonomy.

Holger spoke to us after his poster presentation about his work on stuff and things, which he says is taking the very popular **COCO dataset** and augmenting it with pixel-level stuff annotations.

Firstly, what are **stuff** and **things**? Things are objects with a clear shape and size. Stuff are all the other amorphous background regions that don't have a clear shape and size, have no instances, and are often defined by texture.

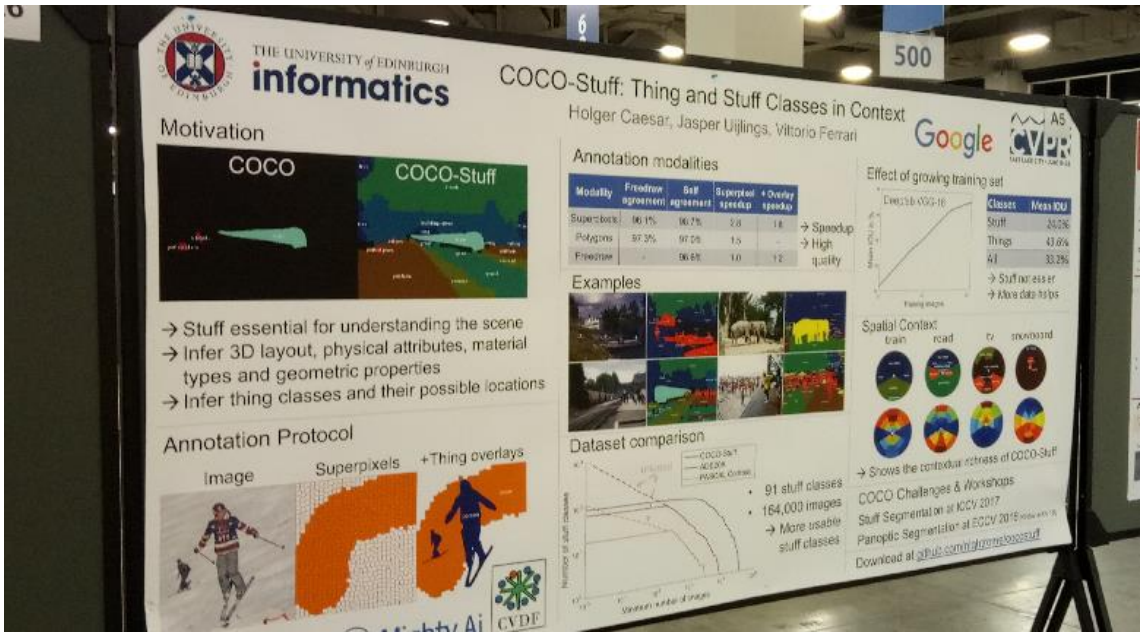
Holger explains: "*We need stuff to know more about the scene. It's not just about cars, you have to know about the road as well. You want to know about the kind of scene, like a highway maybe. You want to know about the 3D set-up of the scene and you can infer that from stuff. It's also about material types and all those aspects. Even if you don't care about stuff, you can use stuff to find things in the scene. Maybe with our annotations, you can improve your results on COCO itself.*"

At ICCV 2017, Holger tells us that he organised and presented the **COCO-Stuff Segmentation Challenge**. It was a new segmentation challenge only on the stuff labels. The next step, he says, coming soon at ECCV 2018 in September, is the Panoptic Segmentation Challenge, which will cover both stuff and things.

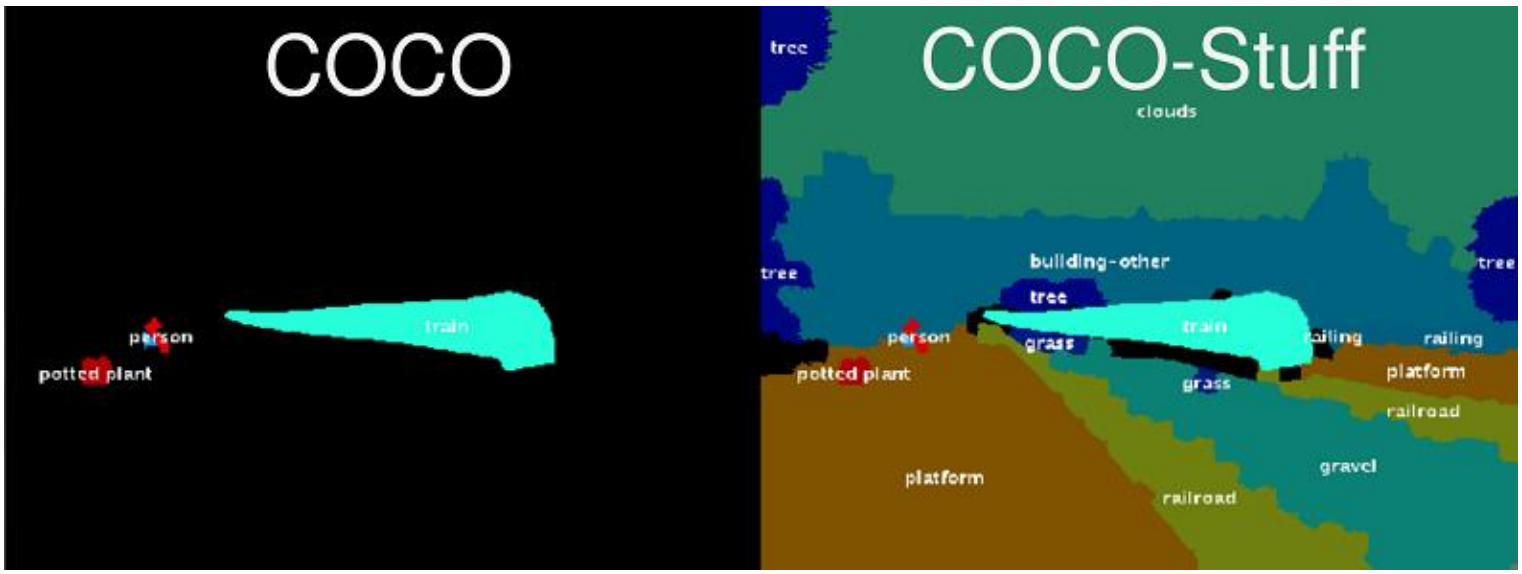
Holger thinks that currently computer vision is a bit incomplete. He says we only really look at objects and have to develop much better models to take context into account. He thinks this work is a good new benchmark to evaluate new methods for these tasks.

The most difficult part, he says, has been **scaling the work up**. It was easy when they did it for 10,000 images in a university with their own students annotating images. When they went up to 164,000 images, they had a company - **Mighty AI** - do the annotations for them, and once it's crowdsourced, it gets much more complex. You need to do a lot of verification to ensure that the annotations are good.

Finally, Holger wants to point out that they did not invent the stuff topic. It goes back decades. You can find it in psychology and linguistics and lots of other places. They are just one piece of the puzzle and many others are working on the same mission.



"Currently, computer vision is a bit incomplete!"



From left: Laura Leal-Taixé, Emanuel Laude and Jan-Hendrik Lange in front of their poster

The task can be formulated as a hard non-convex mixed-integer energy minimization problem.

Emanuel Laude is a PhD student in the Computer Vision Group at the Technical University of Munich under the supervision of Daniel Cremers. He spoke to us at his poster on June 19.

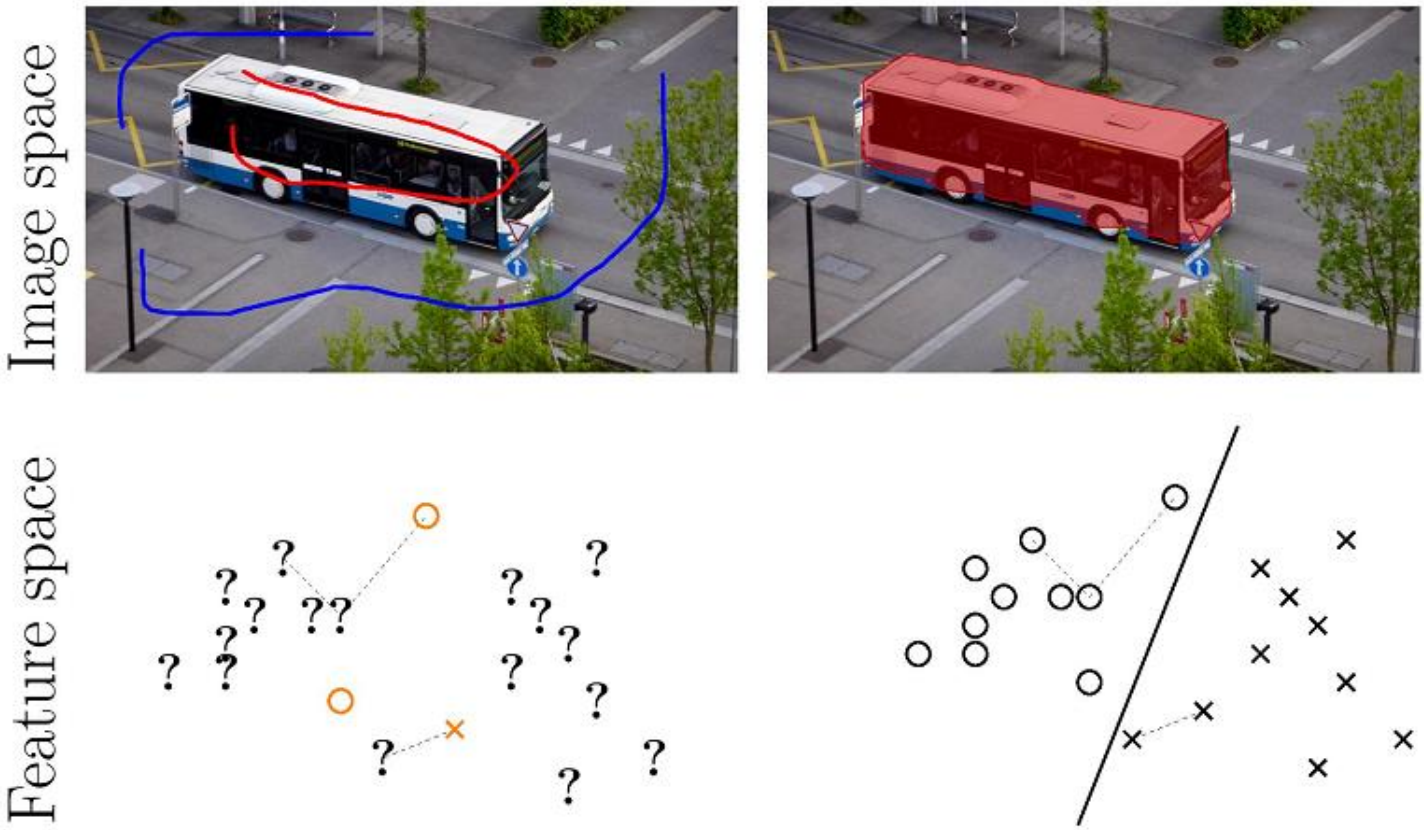
Emanuel tells us that they provide a general-purpose-solver for semi-supervised and transductive learning problems. In contrast to semi-supervised learning, in transductive learning there is no training phase. Instead the labels are inferred directly at test-time incorporating the given labelled training examples.

He says the task can be formulated as **a hard non-convex mixed-integer energy minimization problem**. With this formulation, for instance, they tackle the problem of interactive image segmentation and supervised video object segmentation.

He explains how this work helps: "*The main contribution is a provably convergent solver to tackle this difficult problem. We extend the* ***classical Alternating Direction Method of Multipliers (ADMM)****, very popular in traditional supervised learning (e.g. training of SVMs), to handle mixed-integer problems of the above form. In addition, we provide a theoretical proof, guaranteeing the convergence even under suboptimal label updates. In contrast to grab-cut resp. k-means, our algorithm is better suited to deep features and in video object segmentation can better cope with severe scene changes.*"

One of the work supervisors - **Laura Leal-Taixé**, also from the **Technical University of Munich** - tells us why she thinks this work is important: "*Essentially, it works on a completely different machine learning paradigm, so instead of the classic training and testing, we don't have a training step. With* ***transductive learning****, you have irregular and inference steps where you're trying to match all the data points that you know are labelled, but at test time, not at training time. This is the difference between inductive and transductive. The thing is that if you actually formulate this problem in this way then it's really hard to solve. This is where the paper comes in with a formulation that actually allows you to solve this problem.*"

The next step for this work, Emanuel concludes, is to try out **different optimisation algorithms**. The formulations are quite general and can be applied to **other computer vision problems** such as tracking and detection.



BEST OF CVPR 2018 SALT LAKE CITY • JUNE 18-22 Thursday DAILY

John R. Smith - IBM Watson

John R. Smith is a researcher at IBM and an IBM Fellow. He manages a research team at IBM T. J. Watson Research Center working on computer vision, language, speech technologies and knowledge interaction.

IBM has come in force to CVPR this year.

Yes, CVPR is a big conference for us. It is one of the big conferences in AI actually. We can see this year the attendance is tremendous. I think more than 6,000 people!

6,500 registered attendees…

Yes, really unbelievable. It just shows that computer vision is one of those core areas in AI and thanks to technologies like deep learning, the progress on many problems has been amazing over the past few years and we expect to see this trend continuing.

IBM is here, of course, because computer vision is so important to a lot of the work that we do in many different industries. We are working on problems in everything from healthcare – think of medical imaging – to retail. We are showing some work here, for example, around using dialog and image retrieval together for applications like fashion and shopping, to even fundamental technologies, like how do we improve the efficiency of inferencing from a chip point of view.

You have several papers that are presented here at CVPR; can you tell us about some of them?

Yes, sure, we are showing a bunch of things here. One important activity that I will highlight is that IBM has recently built a partnership with MIT – something that we call the MIT-IBM Watson AI Lab. This is a pretty deep partnership on many different areas of AI. It is very collaborative – we don't necessarily see a distinction between what the people at university do and what IBM researchers do. We are both rolling up our sleeves and working on everything together. One of the first things to come out of the partnership is the development of a unique and very large annotated dataset for actions in video. It's more than one million videos today and these are annotated at the three-second clip level, with a vocabulary of hundreds of different action categories. This Friday, we will have a workshop here and together with MIT organize the first open evaluation and challenge around this particular dataset.

I cannot talk about IBM Watson without asking: what have you got coming up?

Certainly, in the area of vision, there is so much more to do. I think we will see a lot more coming out. Just a hint of some directions – we are showing some work here which is around sports

highlights. We have deployed systems already at the Masters Golf event, Wimbledon, US Open. We are using computer vision combined with other modalities – sound, speech, language, and so on – to understand this kind of content better. The value of doing that is that it can create much more personalized products for the audience and draw their attention to those exciting moments. On the one hand, we will see more on applications, but on the other there is the potential for putting these pieces together to go after new problems. This was a big week for IBM because we unveiled a new system called Debater. It is not vision focused at the moment, it's more about language, but it's really a tremendous advance because it is about a system that can study an important topic deeply and find ways to argue pro and con about that particular topic. It's called Debater for a reason. Think of humans debating; here is a computer participating in a debate.

Can you tell us about some of the things that you are demonstrating at CVPR?

As we think about vision and the requirements, not only you want systems that are accurate, of course that can recognize what you want them to, but speed, bandwidth, power, low power – efficiency is something that is very important – all of these things are very important. One of the things that we are showing here is the development of a neuromorphic chip. What I mean by that is a unique system architecture that uses a spiking network to communicate. We can train this using traditional deep learning methods. We can train a convolutional neural net, we can use a tool like Caffe to do that, but then essentially this network is transformed and compiled for this particular neuromorphic chip. One of the demos we are showing here is pretty awesome. It is gesture recognition and it is real time, but everything is happening on this chip and it is extremely low power. We are talking milliwatts of power that it is able to perform this classification of gestures. The way this chip essentially communicates is using spikes. This is a very different approach compared to a traditional CPU, which might use 32-bit or 64-bit architecture, floating point precision and so on. The reason this is called neuromorphic is because it is

John R. Smith showcasing a real-time demo of a low power, high throughput, fully event-based stereo system. This gesture recognition work is authored by Alexander Andreopoulos, Hirak J. Kashyap, Tapan K. Nayak, Arnon Amir and Myron D. Flickner.

biologically inspired. Think of the way our nervous system works, the way neurons communicate and so on, via spikes.

Is there anything you can tell us about IBM that we might not already know?

One thing that is probably very important is that we are making tremendous progress in the field of AI today. Most of this progress is really what we would call narrow AI. Narrow tasks. Where the success is coming from is generally when we have a large amount of training data. We think of narrow AI as we have a small number of large data problems and deep learning is doing well. Where IBM is particularly focused is the enterprise space and industry problems. The nature of the problems in the enterprise space is we have a very large number of small data problems. When we apply deep learning just out of the box, it doesn't work the same way it does in narrow AI. It's putting new emphasis on learning more from less data. We need techniques of transfer learning –greater transferability, better exploitation of knowledge and reasoning combined with learning. We see this as very important as we go forward.

More than that, as we really want to apply techniques like deep learning for industry applications, the majority of them are decision support. There are people who need to make decisions. Think about a healthcare application – a doctor who is trying to use AI tools to make a better diagnosis of a patient or treatment recommendations. That model may be accurate, which is great, but if the doctor and patient cannot understand how the computer is reaching whatever decision, then this may be a big gap to really rely on its output. It's putting emphasis on explainability. We know that there are also issues about bias in how we train, and often human bias finds its way into bias during learning.

Security of models is also very important. How do we know the models aren't poisoned somehow or corrupted? This whole space of problems we contrast to the narrow AI field and we think of this as more a broad AI. Although that is not necessarily the accepted term, but it's saying that there's still a lot of gaps we need to bridge here to really make AI

successful in these industry applications.

You will have met many young scientists this week who are attending their first CVPR conference. What strikes you positively about them compared to when you were a student?

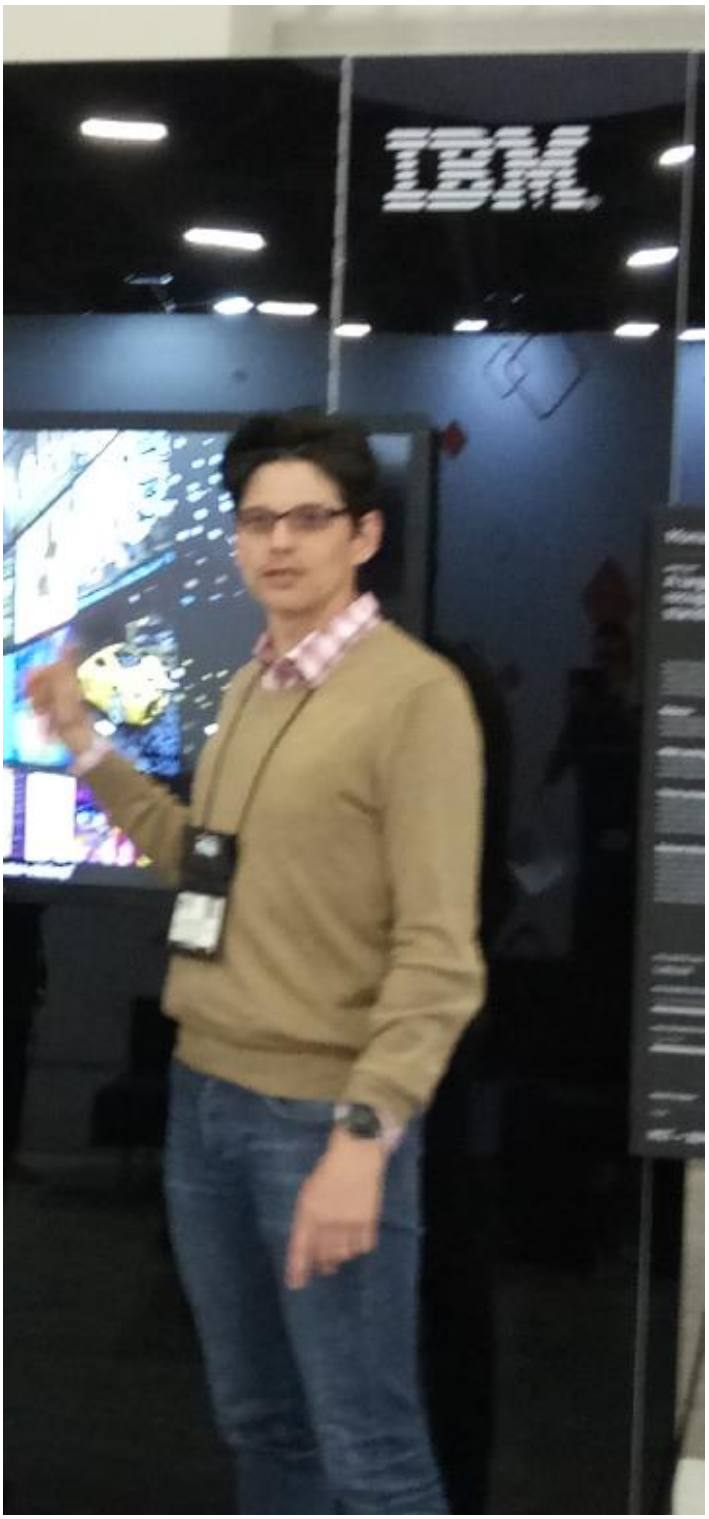
It's a big contrast. Having been in this field for a long time and having worked on image, video, multimedia, computer vision for multiple decades, what has really happened is the people here doing the work – the students, the researchers, the postdocs, the faculty – they're empowered now in a way that hasn't ever happened before. Certainly not in a long time. There's an excitement in their field.

From an application point of view, there are so many different outlets for the work that they do. The number of opportunities they have is tremendous – unprecedented in a way – and there is so much support for the work they are doing. I feel really good for all of them and think they should take full advantage of this opportunity. They are in the driver's seat.

What about the wide range of nationalities represented here today, compared to a few decades ago?

Yes, certainly the community is doing better in diversity, but it is a work in progress. There are many ways in which we need to improve our diversity. It's good to see positive steps here. For example, there's the Women in Computer Vision workshop happening on Friday. I attended that one last year and it was really great to see.

We need to address some of these diversity challenges at all levels. Certainly, we'd like to see an even better result sometime in the future.



Hailing originally from Russia, Ksenia Konyushkova is a final-year PhD student at CVLab in Switzerland at EPFL, working with Pascal Fua. She speaks to us ahead of her spotlight and poster at CVPR.

The work she is presenting is about reducing annotation load for bounding box annotations and was completed during her four-month internship at Google with **Jasper Uijlings**, **Christoph Lampert** and Vittorio Ferrari.

Ksenia explains that they wanted to train an object detector and to do this, needed to annotate bounding boxes. If humans have to manually draw bounding boxes, it can be a time-consuming task. However, previous literature has proposed a solution called box verification series. The idea being that a detector is trained -- maybe a weak detector that is trained only with image-level labels – then this detector is applied to an image and it generates some box proposals. Humans can then quickly verify the box proposals by looking at them and saying if they are correct or not.

The problem with that, Ksenia goes on to say, is that it doesn't always work. Maybe this object is not detected at all, or maybe, as you show the boxes in the order of decreasing score of the detector, the real box comes as the hundredth box, so then it will take a very long time to reach the box that you want to get.

In this work, she is trying to choose what kind of annotation method is the best. For example, a clear image with just one distinct object in the middle is very likely to be detected even by a weak detector, so by doing a box verification series, it will take very little time. However, to detect a small object in a crowded scene, with a weak detector it is unlikely that the object will be found, or it will come very late in the series of boxes. Ksenia is training agents that are trying to figure out which of these modalities of annotation should be used – either verification or manual drawing.

Ksenia tells us there are two ways to solve this problem. The first one is a model-based approach: *"I try to predict the probability that a given bounding box is going to be accepted by the user. Then if we know this probability, we can compute the time that it will take to annotate an image with any sequence of actions. For example, we'll say we'll verify the first bounding box and if it is rejected, we will ask the user to draw. Since we know*

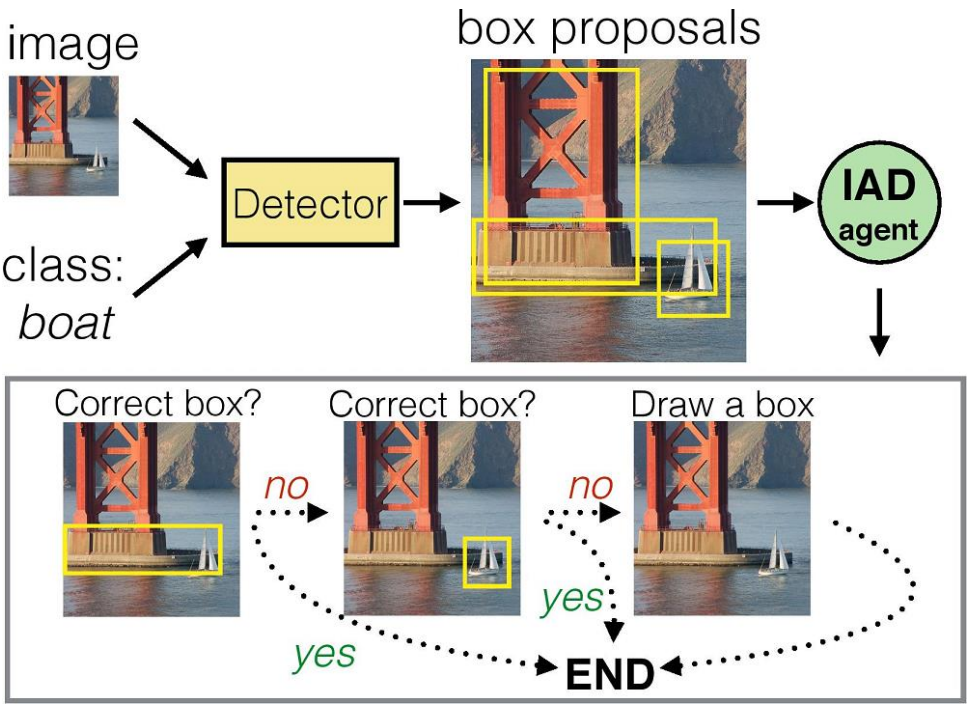
the probability that this box is going to be accepted, we can know what the expectation of the time is that it will take to annotate an image. We construct a provably optimal strategy – given some assumptions as always – that results in the minimum annotation time."

Her second solution is completely model free: "*We use reinforcement learning to train an agent that will act in the best way to minimise the annotation time. We have our environment, that is a human who is annotating images. We have an agent that can act, and the actions of the agents are what modality of annotation we are using. Then we receive a reward from the environment, that is the negative time of the annotation. When we reach the end of the episode that means that we obtain the bounding box that we wanted. Then the reward is zero. In this case, we maximise the reward and naturally we minimise the annotation time.*"

Ksenia says they were able to train such an agent using **DQN, which is Deep-Q Network**, an algorithm from a paper by **DeepMind**. They adapted it to their scenario and were able to train an agent that learns how to act in these annotation dialogs. She adds that the title of their paper is about learning intelligent annotation dialogs, because they construct this sequence of actions, either drawings or verifications, and they are their dialogs.

Ksenia concludes by telling us about their experiments. In one setting they have fixed settings and test many different scenarios. For example, they have a weak detector but need to obtain very precise bounding boxes, or they have a very fast drawing strategy but don't need precise boxes. They show that they outperform other methods there. The other setting is even more realistic. They retrain the detector and their agent that produces dialogs during the data collection. As they collect more data, they can retrain everything, and everything becomes better and better over time.

Ksenia presented her **spotlight** and her **poster** at on June 21 at CVPR.



"... we maximise the reward and naturally we minimise the annotation time."

Adriana Kovashka

Adriana Kovashka is an Assistant Professor in Computer Science at the University of Pittsburgh.

Adriana, you are organizing a workshop tomorrow.

This is a workshop on understanding visual advertisements, which is a very new and challenging task. There's not that much work in understanding advertisements so far. The point of this workshop is to bring together people who are, not necessarily working on this problem, but working on methods which can help in understanding ads. What I mean by understanding advertisements is to understand what the ad is trying to convince you to do, whether that's to buy a car or help protect the environment, and what arguments it provides visually to convey that message, which can be very challenging from a vision perspective, but also a reasoning perspective.

Is it an idea of yours?

Working on advertisements came from my background in media studies. I did a double Computer Science and Media Studies major in college - Pomona College in California. I've always been interested in art and how the media makes us think, how it plays on our beliefs and maybe biases. The workshop was kind of a bold idea, since - being the only organizer - I was afraid that it would be a lot of work. It ended up being fine. I had help from a great undergraduate student as well.

What take-home thoughts do you expect from tomorrow's workshop?

I don't know yet! [*smiles*] That's partly why I'm so looking forward to the workshop. We also have these brainstorming sessions, which I'm going to kind of moderate. Some of them are going to be game-like. Hopefully, we'll have good attendance so we can have great brainstorming sessions.

Did you invite anyone in advertising?

That's a great question! I haven't reached out to advertisers specifically, partly because I haven't thought of a good way to do it. I have been talking to someone in political science in my department. I think that would be greatly helpful. The reason I haven't done it so much is because I feel like the challenges are more in the realm of AI. I haven't had strong enough reasons to reach out to advertisers.

I take personal offense because I could have been one! [*both laugh*]

Maybe we can collaborate in the future.

What do you believe, ideally, will be the resulting effects of the workshop tomorrow?

There are several. On the one hand, I believe that understanding advertisements and persuasion in the media is important because people are exposed to images in the media which have implicit or explicit persuasive aspects.

I think we need to look at images from this point of view as well. I also hope that hearing the talks and what people have to say during brainstorming sessions will help me and others shape this idea in a more concrete, approachable fashion. Maybe we'll identify some tasks that we can work on that are actually approachable, because advertisements are really broad. The images, being very broad, require a vast range of techniques. This is something that VQA (Visual Question Answering) grapples with. I hope that the community has made progress in identifying approachable tasks. I hope we can do the same for advertisements.

All readers are invited [Room 150 DEF]. Let's go back to your work outside of this workshop. How long have you been in your current position?

I've been Assistant Professor at the University of Pittsburgh a little over three years. I got my PhD in 2014 from the University of Texas at Austin. I was working with Kristen Grauman.

What is your regular work? [*both laugh*]

My "real work"! So it's primarily research with some amount of teaching. My job is to train graduate students to become good researchers, which is very interesting and also very challenging. Some graduate students are made to be researchers. Some have great technical skills, but maybe don't have as much appreciation of novelty. You need to train them to be researchers, not just engineers. Working with graduate students has been a very interesting experience. I also do teaching, which is also very rewarding: in the past semester, I taught a graduate and undergraduate vision class. I very much enjoyed the undergraduate actually, because students were deeply interested in the topic. We really see huge excitement. It was very enlightening. In the past, I have struggled with students finding the subject hard. Somehow, this semester, I think we connected a lot more, and there was a lot of fun being had in class. It was a great experience.

Did you always dream of becoming a scientist?

Hmm… yes! I've always been interested in art as well, but I think I've been a scientist most of my life. I think that makes more sense for the kind of person that I am.

You are not originally from Texas…

No, I was born in Sofia, Bulgaria, but I've been in the US since 2004. I'm not really sure if I'm more European or more American at this point.

How can you no longer feel European? As an Italian, I cannot understand that! [*laughs*] Right, well, I'm probably, deep inside, European. That never goes away.

Where is home?

Home is Sofia, Bulgaria.

Will it always be?

[*smiles*] Yes!

Most of the attendees at this conference have never been to Bulgaria. Can you share something about Bulgaria that most people don't know? Don't tell me food, it's too easy!

[*laughs*] I was going to say that, and I realized that's too boring! [*both laugh*]

I guess this applies to Bulgaria as a whole, but we have great nature. The Bulgarians that I hung out with while I was still living there, mostly my friends from high school, were very open to ideas. Our favorite thing to do was to hang out in the park, discuss philosophical ideas, and drink wine.

Which was the best part?

All of them! That's how I remember Bulgaria, the combination of intellectual pursuits and just being Bohemian, I guess, and enjoying nature at the same time.

Can you recommend a place to visit in Bulgaria?

Despite what I said about possibly being more American, I've given several of my servers Bulgarian names. One is called Sozopol, which is a beach town where there is an annual culture festival with jazz and theater and all that. It's a great place to go in the summer.

Do you ever think that you did things better when you were a student? Or are students doing better things now? Do you see any remarkable difference?

I've seen both. I can comment on both of these.

Please do!

I think some of my students, many of them, are more productive and more focused than I was when I was in grad school. That's amazing!

What is the reason for that?

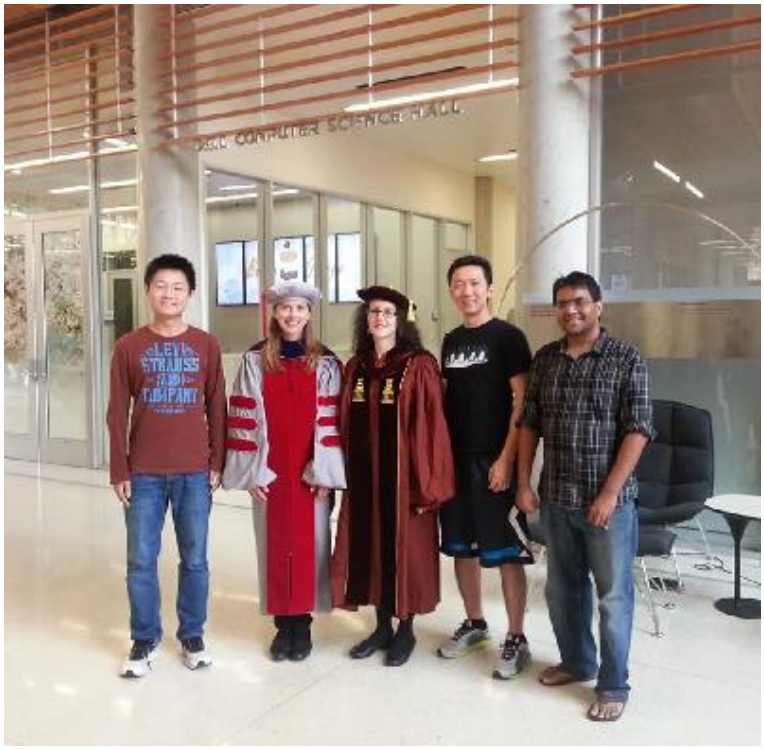
I don't know! I think they just are more driven than I was. Maybe I was exploring this space, and they always knew they wanted to do this in particular.

Did you become more driven with time?

I became more driven with time. That's why I am an Assistant Professor, but I think they are better than I was when I was in grad school. The second part is what they do that I don't think I would have done. I think I was a little bit more open to feedback and suggestions in terms of things like writing. I trusted my advisor all the time.

Let's go back to your drive. It has increased over time, and you are very much attracted to media and communications, which you even studied in college. You love to teach. You certainly love to learn and make research. Actually, research is a major component of your life. How many drives can you have? If you are not as driven as your students are now, to me you sound overdriven! [*both laugh*]

I am by no means overdriven. I do take on more tasks than I should, which ends up not always being a good idea. In a sense, this drive is because our community has changed in ways that I sometimes dislike, but also in ways that I like. I think there are more diverse ideas being pursued, which I like. I have



noticed that, when I was in graduate school, I used to only look at papers that were in my area. Whereas now, I find more topics interesting. I don't know why that is, but I like that.

"Not overfocus on performance"

Is there anything you would like to change in our community?

I haven't been in this game for long enough to be able to say authoritatively what I would like to change. There are people with much more experience than I have. One thing that we discussed at the Visual Question Answering workshop was that perhaps the community should encourage sharing novel ideas more and not overfocus on performance. That's one thing that I would like to see.

...which is not far from what Nikos Paragios said in his famous article about the "deep depression".

I think it's a complex issue, and I found what Jitendra Malik had to say at that workshop very interesting about how we, as a science community, need to be driven by empirical results. We don't need to only do that. We need to value novel ideas and give novel ideas a chance so that they can develop more..

I would also encourage people to attend the Good Citizen of CVPR workshop, which I think is a great idea by Devi Parikh and Dhruv Batra from Georgia Tech. There will be a variety of topics discussed there: how to be a good member of the CVPR community. I encourage everyone to attend that. That is on Friday as well.

"It was fun!"

So you are advertising other workshops than yours? [*both laugh*]

Actually, I have a conflict of interest because I am also speaking there. [*smiles*] I would also encourage people to attend Visual Understanding of Subjective Attributes of Data, which is also held on Friday. Thanks for the interview, Ralph! It was fun!

From left: Yongqin Xian, Zeynep Akata and Tobias Lorenz

Yongqin Xian is a PhD student at Max Planck Institute for Informatics, with Bernt Schiele and Zeynep Akata. We spoke to him before his poster session on June 20 at CVPR.

Yongqin tells us that in this work, they are trying to tackle the generalized **zero-shot learning problem**. They want to train their classifier to predict both seen and unseen classes at test time. This is different from conventional zero-shot learning, because in conventional zero-shot learning, the classifier only predicts unseen classes.

This problem is very challenging, he says, because it suffers from extreme data imbalance between seen and unseen classes. There is a lot of seen classes data, but there is no unseen classes data at all. This makes the classifier bias to seen classes. That's why in this work, he proposes to generate synthetic data of unseen classes to fix this data imbalance. However, instead of generating synthetic images of unseen classes like most **GAN** papers do, he proposes to directly generate **synthetic CNN image features** of unseen classes, conditioned on all class-level attributes.

The challenge in doing that is that they need to guarantee that the generative features are good enough to train on supervised classifiers. If it's unstable to train, they need to propose **GAN architecture** that is stable and can generate features with good quality.

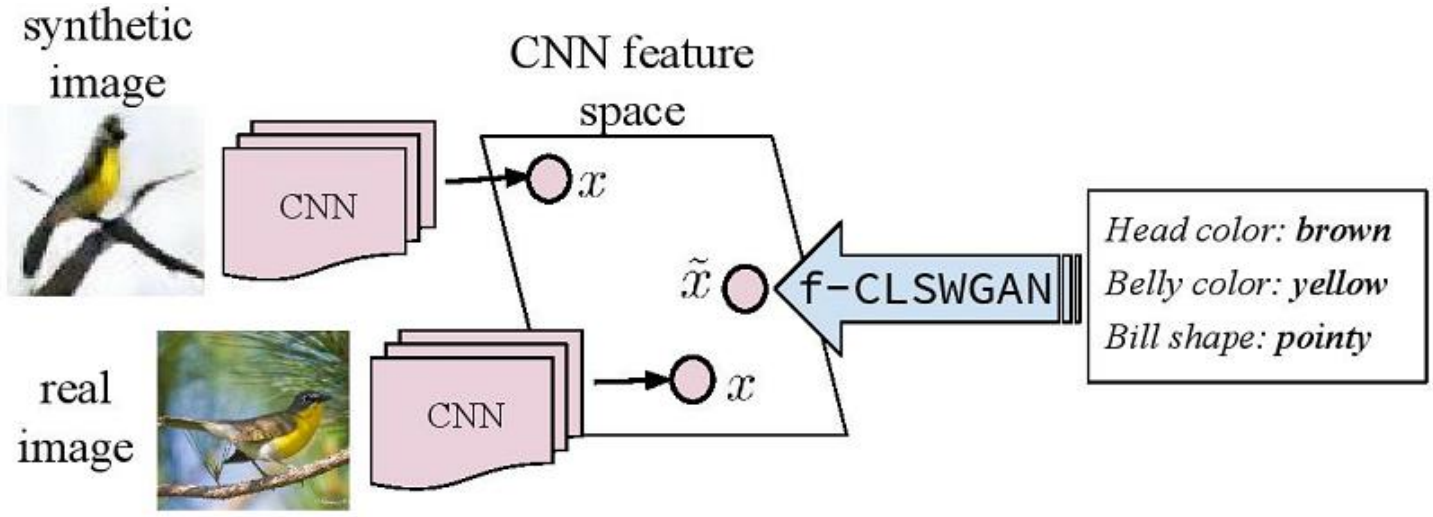
… you can use this model to generate more synthetic features of those classes for which you have very few examples.

Yongqin explains what algorithmic techniques they use to solve this: "*We proposed to use Wasserstein-GAN to stabilise the GAN training. To enforce the discriminative ability of the generated features we propose a classification loss which enforces the generated features can be correctly classified by a pre-trained classifier.*"

Zeynep adds that they make use of the fact that the data that is surrounding us is inherently multimodal. If you have text that accompanies images, and you don't have enough labelled images but you have text that you can use to associate different sets of classes, then you can use this model to generate more synthetic features of those classes for which you have very few examples.

She tells us: "*We show the capability of this model on **ImageNet**. ImageNet is one of the largest scale datasets that is available to us. It generalizes to the cases when we don't have any label training data from some of the classes, just because our model is able to associate different classes in a conditional generative adversarial net framework.*"

In terms of next steps, Zeynep says they have thought about making use of the existence of text better. At the moment, they are generating image features that correspond to text sentences or attributes, but from looking at it the other way around, they can also generate text for images. They could expand this framework to do explanations, for example, explaining scene understanding and the semantic image content.

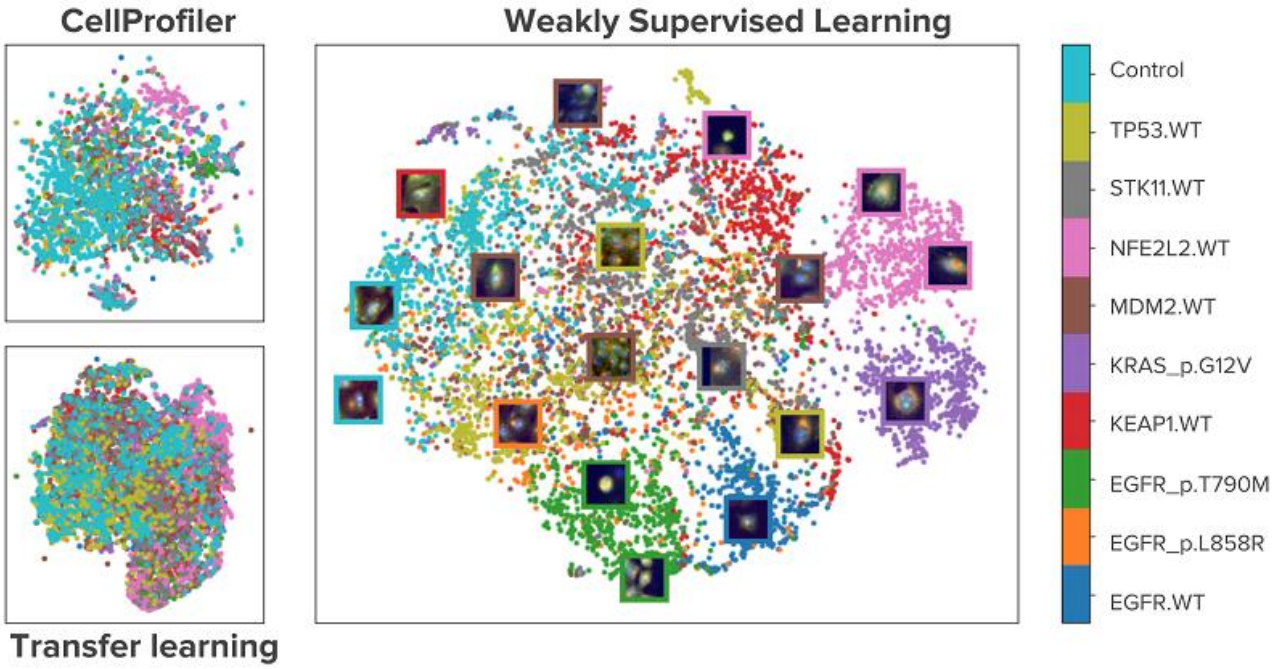


"We take the images and those images usually have multiple cells, so we use neural networks to extract each individual cell from the image."

Juan Caicedo is a postdoc researcher at the Broad Institute of MIT and Harvard. He works with a multidisciplinary team of biologists and computer scientists, led by Anne Carpenter, and the lab is mostly focused on analysing biomedical images. More specifically, microscopic images for understanding the effects of treatments, either genetical or chemical perturbations. He speaks to us before his poster presentation at CVPR.

The work Juan is presenting is about analysing microscopic images in order to learn representations for those images without any labels. The technique is called **weakly supervised learning**. He is excited to see so many different researchers working in weakly supervised learning at CVPR 2018.

Juan tells us that he wants to learn representations for these images in order to compare the effects of different treatments. Imagine a patient with cancer and you take some cells of this patient and capture images of those cells under the microscope. Then you want to know what happens if you apply certain drugs to those cells. Just by looking at the images, is the population of cells going to have a positive response or a negative response? For many problems in computer vision, including microscopic images, there are not nice and clean labelled datasets. Since they don't have labels for these specific patients, they want to make a system that can learn without labels or additional information.



Juan explains how they do this: "*We are using convolutional neural networks, which is the mainstream technique to analyse and process images nowadays. We take the images and those images usually have multiple cells, so we use neural networks to extract each individual cell from the image. Then we train another network which learns to recognise the differences with respect to other cells. That's the thing that we use in order to differentiate if a treatment works or if it doesn't.*"

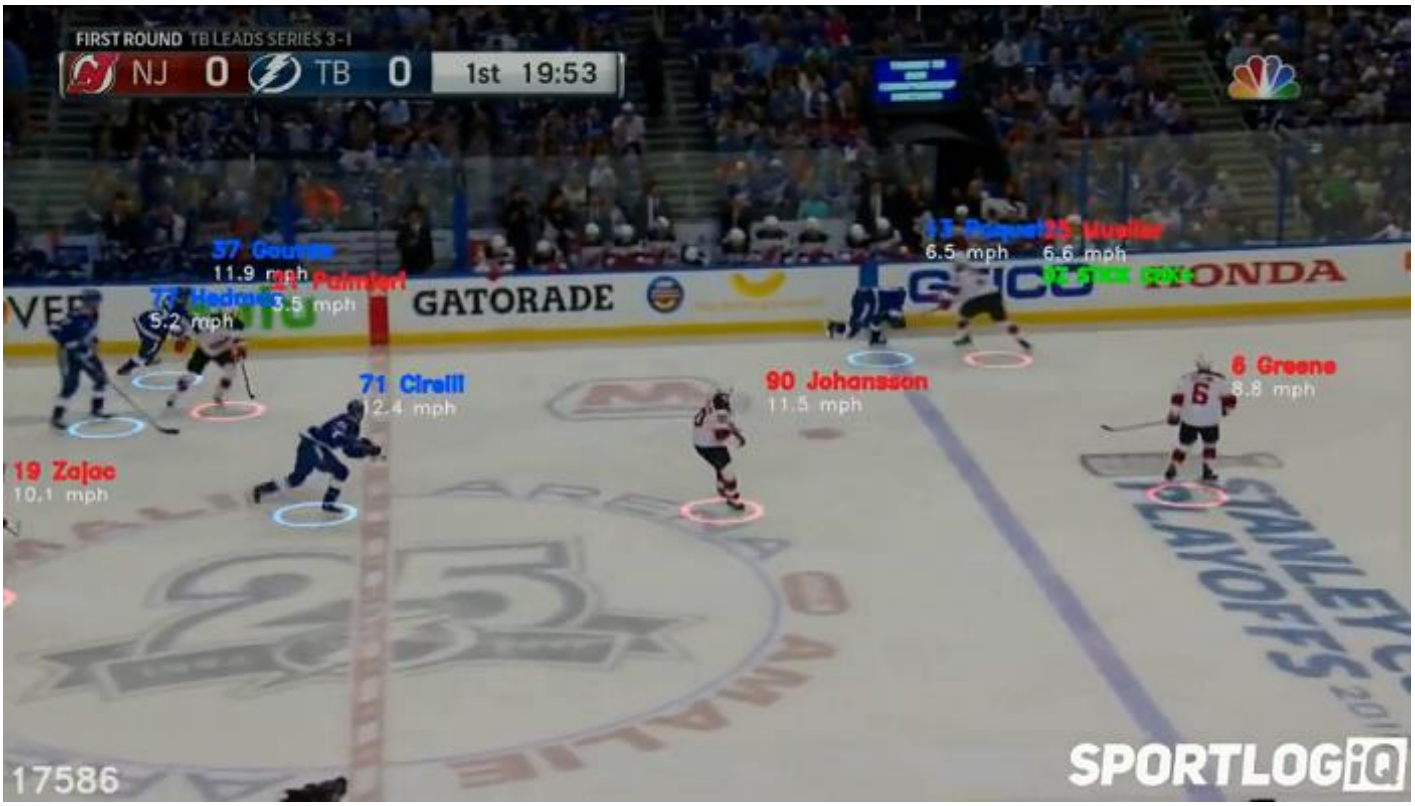
The work that Juan is presenting is the first time that they have used **neural networks and convolutional networks** for the analysis of this type of biological image. They are using it to understand cancer mutations. In the poster, he will present some examples of how they detect when a mutation has a difference with respect to healthy cells.

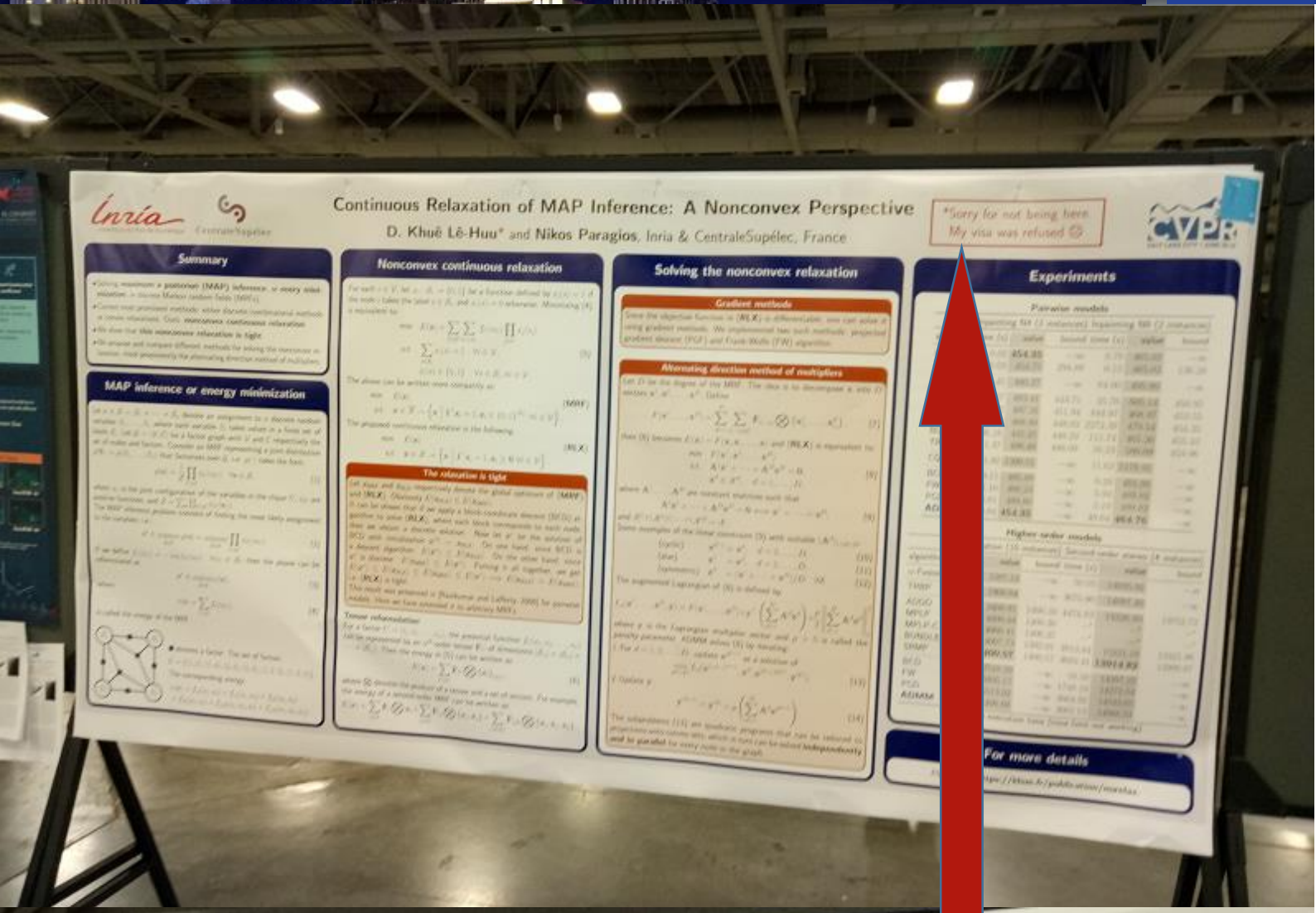
Juan explains that one of the main problems in cancer treatment nowadays is that a single tumour can have multiple mutations. We don't know exactly which of those mutations is the one that drives the cancer. We can identify all of those mutations, but we don't know which one is the problematic one. With this image analysis technique, they can detect which of the mutations are having an impact in the growing of the cancer or the type of cells that are present. He believes that with this they can unlock the difficulties of treating cancer at a much larger scale.

Thinking about next steps, Juan says that right now they are analysing even more mutations. It's something that was not possible before, because the techniques to understand when a mutation is impactful or not cannot be run at larger scale. It's the power of images and he says that's why they are presenting their work at CVPR, because they can run a larger scale study with millions of images in order to understand pretty much any mutation in the human genome. He hopes they will scale it up to many more mutations in order to prevent and treat cancer.

Juan presented his poster on June 21.

SPORTLOGiQ is an AI platform that understands the contents of sport videos and generates actionable insights for coaches and teams and stories for fan engagement with Media. Starting in hockey, SPORTLOGiQ has now expanded into soccer/football and beyond. Check these videos demonstrating the benefits.




Continuous Relaxation of MAP Inference: A Nonconvex Perspective
D. Khuê Lê-Huu* and Nikos Paragios, Inria & CentraleSupélec, France
*Sorry for not being here. My visa was refused
Summary
MAP inference or energy minimization
Nonconvex continuous relaxation
The relaxation is tight
Solving the nonconvex relaxation
Gradient methods
Alternating direction method of multipliers
Experiments
Pairwise models
Higher order models
For more details


ective
*Sorry for not being here.
My visa was refused ☹
Experiments
Pairwise models
Inpainting N4 (2 instances)

Facebook's Cristian Canton Ferrer (left) and Brian Dolhansky (right) showing their poster on June 21 at CVPR.

Matthias Nießner (right) and Angela Dai (center) presenting their poster on June 20.

Silvia Zuffi presenting her work on June 20 (co-authors Angjoo Kanazawa, Michael Black)

Dan Xu presenting his work on June 20 (co-authors W. Ouyang, X. Wang, Nicu Sebe)

The 4th Women In Computer Vision Workshop

by Ilke Demir

Computer vision has made tremendous progress in the recent years over a wide range of areas, becoming one of the largest **computer science research communities**. However, where are "*we*"? The percentage of female researchers both in academia and in industry is still significantly low. As a result, most female computer vision researchers feel isolated and the lack of inclusion creates unbalanced workspaces and biased products.

The workshop on Women in Computer Vision is a gathering for both women and men working in computer vision targeting a broad and diverse audience of researchers from both industry and academia to extinguish this imbalance. The first goal of the workshop is to raise visibility of women in computer vision. We accomplished this goal by inviting high quality research talks from junior and senior female researchers to present their work as keynote speakers and oral presentations. We also organized a panel discussion for inclusion and diversity topics in an open and friendly environment between female and male colleagues.

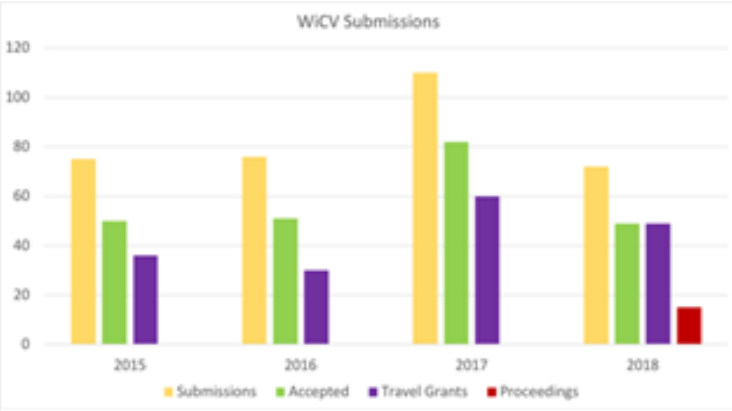
Our second goal of giving opportunities to junior female students and researchers was accomplished by enabling them to present their work via a poster session, and providing travel awards to ease and encourage their participation. We particularly encourage work in progress and work from junior graduate students so that the authors have a chance to hear feedback and suggestions on their work. The last major goal is maintaining and growing WiCV network where female students and professionals share experiences and career advice all over the globe. We especially organized a mentorship banquet to provide a safe and casual environment in which junior women can meet, collaborate, exchange ideas, and form beneficial relationships with senior faculty and researchers in the field.

...a safe and casual environment in which junior women can meet, collaborate, exchange ideas and form beneficial relationships with senior faculty and researchers in the field.

Organizers Viktoriia Sharmanska, Ilke Demir, Adriana Romero and Lyne P. Tchapmi. In the smaller picture, organizer Dena Bazazian who was not able to attend due to visa restrictions.

To begin with, we had an incredible set of organizers (see previous page), from geographically distributed institutions at 4 different time zones, with a maximum of 9 hours difference. It was an honor and a pleasure to work with those brilliant researchers, whom I feel close enough to call my academic sisters.

WiCV was held for the first time at CVPR 2015. Over the years, the attendance and quality of submissions to WiCV significantly increased, however having such work only in poster session was too ephemeral for our visibility goals. First, to have a permanent resource, we believe that this year we took a big step by having workshop proceedings. Second, this year we also **live-streamed the workshop on Facebook**, which is also a great resource for anyone who could not attend WiCV. Compared to previous years, we increased the duration of the workshop from half a day to a full day gathering, inviting more senior and junior researchers to present their work. Lastly, another advance is that the frequency of WiCV has been doubled: **the fifth WiCV workshop** will also be organized this year, collocated with ECCV.



We had great submissions on a wide range of computer vision and machine learning topics. Over all submissions, 8% were selected to be presented as oral talks, 21% were selected to be included in the proceedings, and 68% were selected to be presented as posters. We had a diverse program committee of 43 reviewers to evaluate and help improve the papers. Last but not least, we believe that we had a surprising upgrade over past editions, we were able to provide travel grants for authors of all accepted submissions who applied for a travel stipend. This is the first time that WiCV is supporting all presenters and we are greatly thankful to our sponsors. We also had a record sponsorship amount of 126,500 USD, almost twice as much as previous years' sponsorship amount.

Before the main workshop day, we had our mentorship banquet in the evening, sponsored by Facebook. It was also the first time that the committee was physically gathering and we were almost screaming by the excitement to see each other. We had over 100 participants to the dinner, with a queue at the door, including senior researchers, speakers, mentors, and WiCV participants. The first mentorship talk was given by **Xin Lu**, where she talked about her story including career choices, desire for impact, and advices for the new generations. I particularly enjoyed the discussions on my table, because I was able to ask difficult questions to those I academically admire (**Thomas Funkhousser**, **Jessica Hodgins**, **Octavia Camps**, **Ayellet Tal**, and **Jana Kosechka**), in this friendly setting. Afterwards **Dima Damen** shared her experiences and advices in a very sweet mentorship talk. Lastly, **Timnit Gebru** touched all of our hearts by her story including the realization of being double minority and how she kept coping with the world

Timnit Gebru sharing her story during the WiCV mentorship banquet

in difficult times. All mentorship talks were so special to everyone in the room: they encouraged us, inspired us, trusted in us, and prepared us for future.

The following day started early and with the awesome help from our friends at Facebook, we managed to prepare our precious WiCV bags before the workshop. I gave the opening remarks of the workshop including our motivations, statistics, improvements over the past years, and a glimpse of the program. We also added a section called "**Opportunities**" to summarize other events and communities that the audience can benefit. We prepared our workshop program to include **4 keynotes, 6 oral presentations, 43 poster presentations, and a panel discussion**. This year, we selected our keynote speakers to balance academia versus industry, as well as junior versus senior researchers ratios. Moreover, we put special effort to diversify speakers covering different research domains and backgrounds. We believe that diversifying the set of speakers is of crucial importance to provide junior researchers with potential role models with whom they can identify and who can help them envision their own career paths. This also helps us put all perspectives together in our panel discussion. This year it was an honour to host **Jessica Hodgins** and **Octavia Camps** as the role models that guide all of us in this journey. Jessica's keynote talk enlightened us about the history of analysis and synthesis of human motion in vision, graphics, and robotics, and Octavia presented an insightful keynote about using dynamic systems as encoding for video understanding. Our relatively junior keynote speakers were **Laura Leal-Taixé** and **Carol E. Reiley**. Laura presented their research about blending homography estimation with deep learning, and Carol introduced the autonomous driving technology of drive.ai

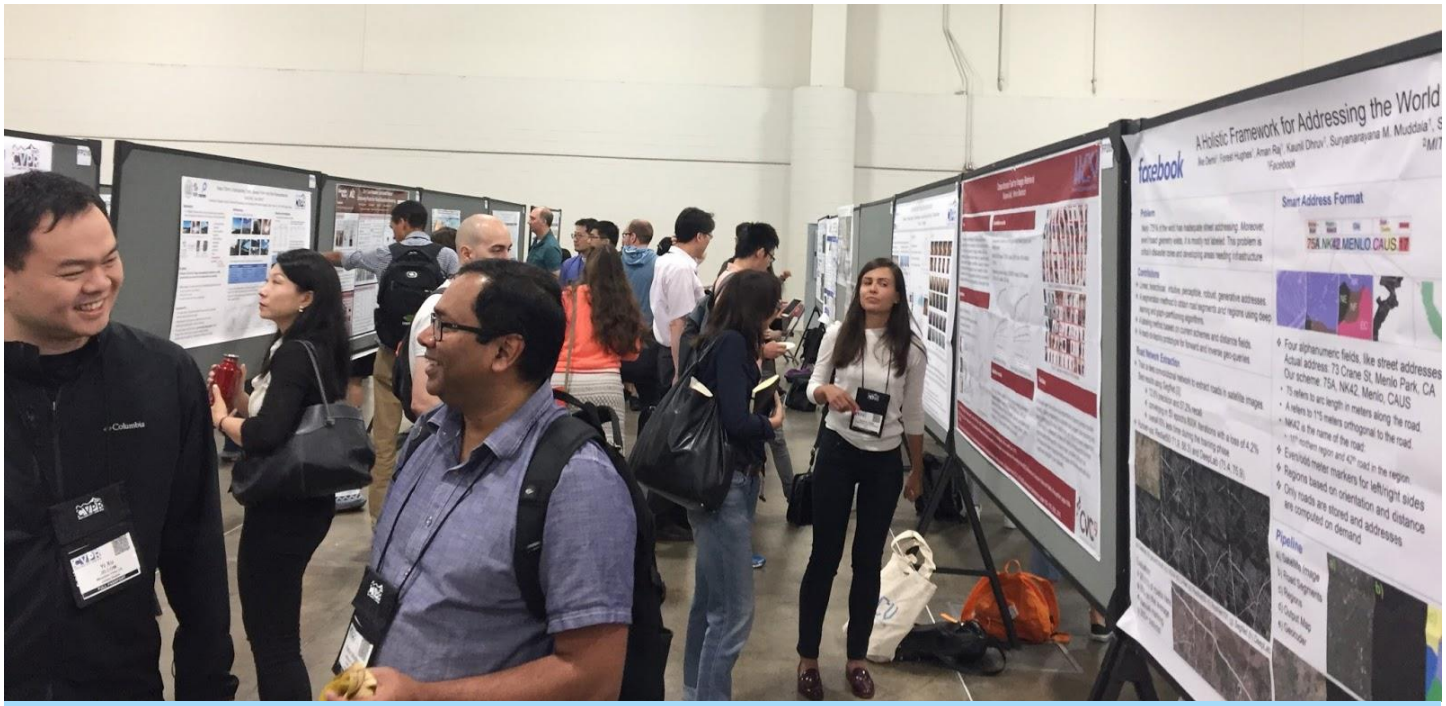
Jessica reminding us to have fun by showing how her robot jumped from a fire circle at 2am, before her dissertation defense.

with oracles about the future of the automotive industry.

In addition to the keynote speakers, we invited the best paper submissions with novel or recently published work to WiCV as oral presentations. The main goal of these sessions is to give female students an opportunity to give a research talk in a professional and supportive setting. You can reach the papers online, and check out the program for the list of oral and poster presentations. The poster session went really well with many researchers in the general poster area stopping by to ask questions, so we can assume that we achieved our goal to increase visibility of female researchers beyond the WiCV room.

The workshop also included a panel, where inclusion and diversity topics was discussed in an open and friendly environment between female and male colleagues. In addition to the keynotes and mentorship speakers as panelists, we had Michael Black and Jitendra Malik as our male panelists. We had many challenging questions from the committee and from the audience, such as the role of male allies, family pressure based on gender roles, the clique culture in CVPR, changes in hiring pipelines, and personal experiences of panelists regarding failures and impostor syndrome. In all cases the flow of the conversation was always friendly and joyful, under my moderation. You can watch, learn, and enjoy the panel here (I even send salutations to my family after the question of Andrew Fitzgibbon from the audience). I think such conversations should become more usual and less challenging if we want to bring and include everyone at the table. With the panel, we concluded WiCV 2018.

We believe that WiCV 2018 will be a great accomplishment for presenters, participants and organizers towards a more connected community to overcome the ongoing gender imbalance in the workforce and its side effects. With first-time proceedings, a



WiCV poster session with the participation of general CVPR attendees.

From left to right: panelists Jitendra Malik, Xin Lu, Laura Leal-Taixé, Jessica Hodgins, Dima Damen, Octavia Camps and Michael Black, with moderator Ilke Demir

record amount of sponsorships, a full day gathering, and talk recordings, we foresee that the workshop will proudly make its mark towards its goal of increasing visibility, providing support, and building community. We would like to thank all senior researchers that supported this initiative and we are expecting the future generations to carry this torch until we do not need it anymore.

See you at ECCV!

As a last word, we would like to send our thanks to all our sponsors, to Negar Rostamzadeh for the incredible help and initiative at the beginning, to previous organizers for the information flow, and to CVPR 2018 Workshop Chairs. Finally, we would like to acknowledge the time and efforts of our program committee, authors, mentors, speakers, submitters, and attendees for playing an active role in building WiCV. **See you at ECCV**!

Above, WiCV in a nutshell. Thanks to all contributors!

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.

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

ECCV 2018- European Conference on Computer Vision
Munich, Germany MEET US! Sep 8-14 Website and Registration

CGVC2018 - Computer Graphics and Visual Computing
Swansea, UK Sep 13-14 Website and Registration

CSCS 2018 - ACM Computer Science In Cars Symposium
Munich, Germany Sep 13-14 Website and Registration

MICCAI - Medical Image Computing and Comp. Assisted Intervention
Granada, Spain MEET US! Sep 16-20 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:

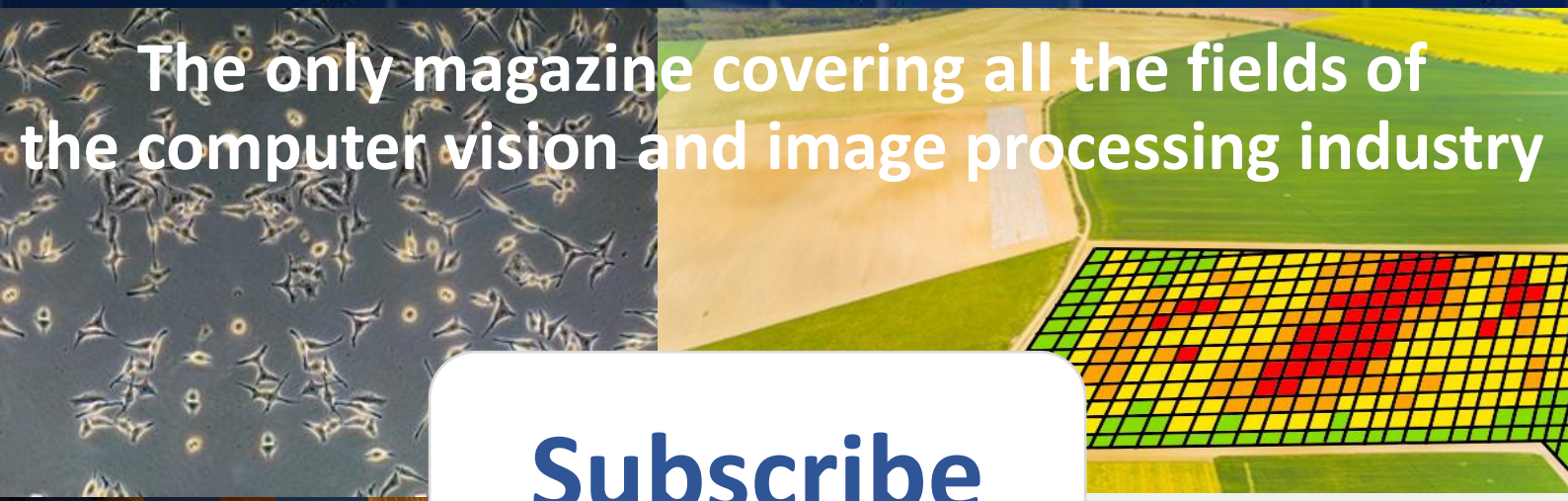
Give us feedback, please (click here)

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

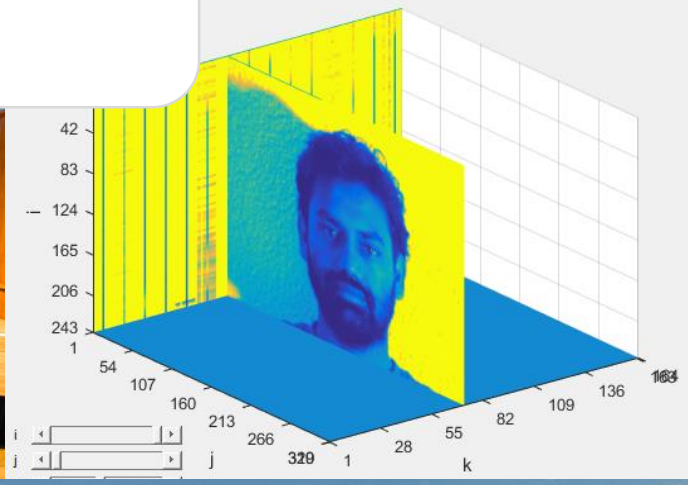
The Magazine Of The Algorithm Community



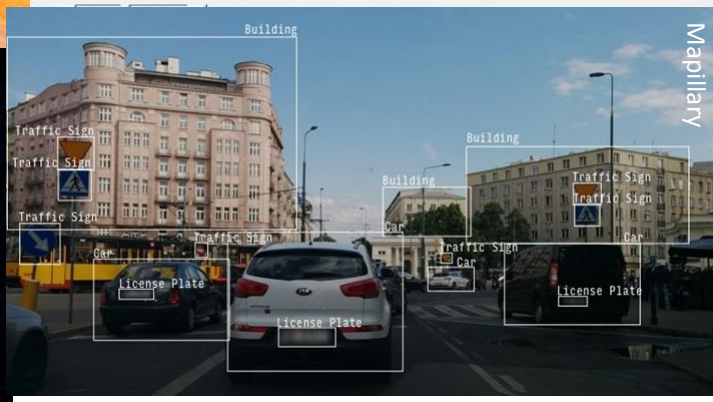
The only magazine covering all the fields of the computer vision and image processing industry

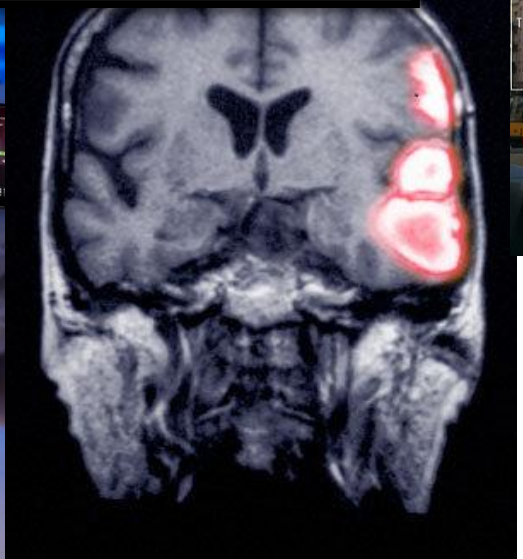
Subscribe

(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