ECCV'20 ONLINE
23-28 AUGUST 2020

2020 DAILY Thursday

Women in Computer Vision
Xin Lu (Adobe)

Oral Presentations
Marcel Geppert

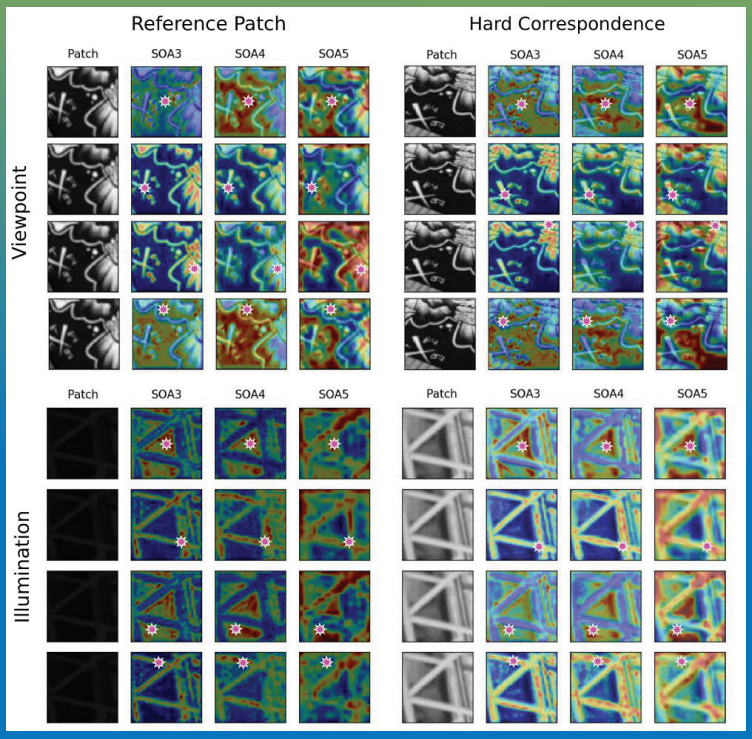
Poster Presentations
Tony Ng

General co-Chair
Bob Fisher

Picks of the Day
Robby Neven

Spotlight Presentations
Dario Pavllo

Workshop
BioImage Computing



In Cooperation with
Computer Vision News
The Magazine of the Algorithm Community

A Publication By
RSIP VISION
Global Leader in Computer Vision and Deep Learning

For Today, Thursday 27th

*My name is **Robby Neven**. I am a first year PhD student at the **EAVISE lab at KU Leuven**. My research focuses primarily on embedded deep learning, enabling models to learn on resource constrained devices. This is becoming more important as AI is being integrated in everyday devices from autonomous cars to smart refrigerators.*

*Tomorrow I present my research on feed-forward training at the **Embedded Vision Workshop**. As it is my first conference, I'm excited to share my work with fellow researchers. Feel free to virtually hop by.*

*Too bad we can't go to **Glasgow**. As a whisky enthusiast, you can imagine I was excited to visit some famous whisky distilleries. Ah well, to all the people I now have to miss out on I'll end with a **Scottish saying**: Lang may yer lum reek!*

Robby's Picks of the day:

Orals:

- Bailin Li - EagleEye: Fast Sub-net Evaluation for Efficient Neural Network Pruning **#2170**
- Jun Fang - Post-Training Piecewise Linear Quantization for Deep Neural Networks **#4328**
- Hongwei Yong - Gradient Centralization: A New Optimization Technique for Deep Neural Networks **#4092**

Spotlight:

- Jingwei Xin - Binarized Neural Network for Single Image Super Resolution **#2055**

Poster:

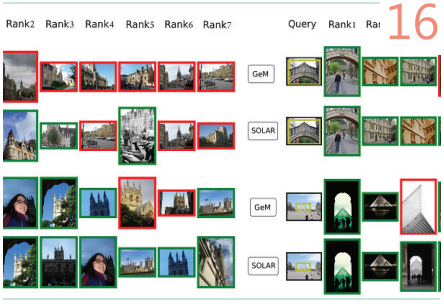
- Kunyuan Du - FTL: A universal framework for training low-bit DNNs via Feature Transfer **#4186**
- Su Jong-Chyi - When Does Self-supervision Improve Few-shot Learning? **#4229**

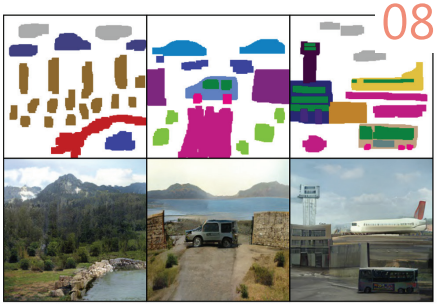
Do not miss our own Thursday picks. Read about them, starting on page 4!

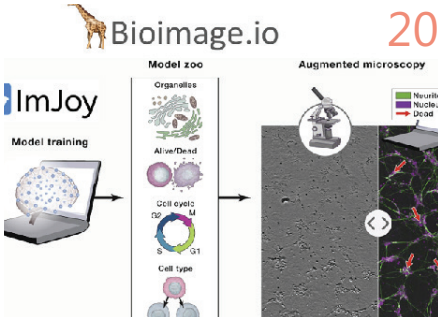
ECCV Daily

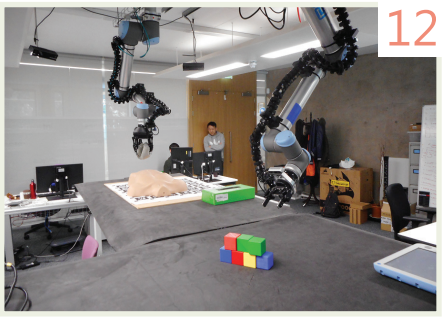
Publisher: **RSIP Vision**

Designer: **Rotem Sahar**









Good morning ECCV!

This is the last **ECCV Daily** magazine for this year.
These days, when researchers cannot meet in person, it is key to consolidate ties and improve communication. This is what we do at [RSIP Vision]: keep in touch with the research community and subscribe now to [Computer Vision News], it's free.
See you next year, **hopefully in person**, at **CVPR** and **ICCV 2021**.
Enjoy the reading and have a great day at **ECCV 2020 online**!

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

Privacy Preserving Structure-from-Motion

Marcel Geppert is a PhD student at the Computer Vision and Geometry group at ETH Zurich, under the supervision of Marc Pollefeys. His work addresses the significant privacy concerns that arise as localization and mapping solutions become more cloud based. He speaks to us ahead of his oral presentation today.

Current structure-from-motion (SfM) methods use a lot of image data, and with new methods for mobile devices processing images on a server in the cloud, **private user information could be revealed**. This work aims to remove as much of this information as possible, while maintaining the same results in the form of a point cloud or map.

"Usually, we would not transfer images directly, we would do feature extraction like SIFT," Marcel explains. *"We would use a deep learning method by Francesco Pittaluga that takes those features and their keypoint positions and gets surprisingly good results by just reconstructing the original image. In our method, instead of keeping the exact 2D keypoint position, we add a degree of freedom and write it as a line that includes the original extracted keypoint. The position information is missing, so it is not possible to reconstruct the input image anymore."*

The inspiration for this work came from **co-author Pablo Speciale**. He was first author on two papers: **Privacy Preserving Image Queries for Camera Localization** and **Privacy Preserving Image-Based Localization**. The first proposes to protect the map with respect to the user. Instead of having a point cloud, it does the same thing in 3D, **but replaces each point with a line**. With this method, it is still possible to estimate the camera position based on point features in the image and lines on the map. The second paper has a point cloud with line features – which is also used in this work – and localization of that. The problem with this method is that to create the map, you still need to use **standard SfM methods**.

"That is basically what we added here," Marcel tells us. *"The single modules that we use were known before. The localization was done by Pablo. Point triangulation is not that hard. You need more views than for standard keypoints, but with lines it is still possible. The*

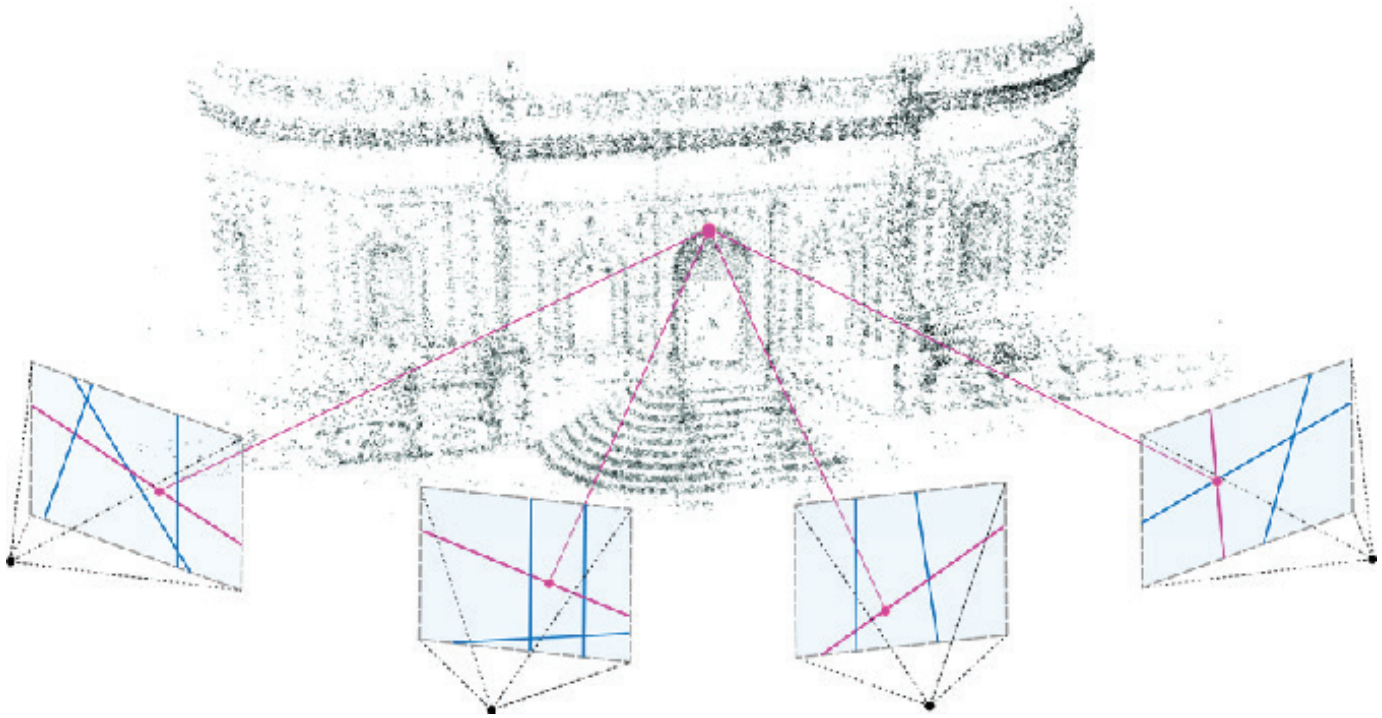
*difficult part is initialization when you need to estimate the relative pose of the cameras. The line features give only a single constraint mathematically compared to two from a point feature, so you need more images in order to triangulate the point in the first place. Also, you need additional constraints for pose estimation. Line features cross so quickly that it is not really feasible to do a full relative pose estimation. We need four images and I think it is 17 correspondences on all four images. Then running this in **RANSAC** is just not feasible."*

To resolve this, this method assumes that **gravity direction is known** – which is a reasonable expectation with modern devices – and with gravity direction known, it is much easier to perform initialization because **features** can be aligned with the gravity. Part

"This work aims to remove as much of this information as possible, while maintaining the same results..."

of the problem then becomes the 2D version of the relative pose, but by doing a relative pose estimation for three views in 2D, it can upgrade that to full 3D photos with four views.

You can view the team's supplementary video to see an example of a **reconstructed point cloud**. They try to reconstruct original images from the data and compare that to traditional SfM methods. You will see that it is mostly dynamic content with people walking around who almost completely disappear because it is not possible to reconstruct that part anymore.



"The difficult part is initialization when you need to estimate the relative pose of the cameras."

Marcel says it is the new possibilities that arise from this method that excite him the most. He has worked with several SfM and SLAM methods before and is excited to know that by adding weaker constraints, it is still possible.

We put it to Marcel that he is very fortunate to be working in this field at this time when it is positively booming.

"Yes, I am very lucky that I am here at this point when there are so many resources dedicated to this work. I can learn so much because every year there are so many new papers and ***the state of the art is always improving****. Just think about all those things that did not work one or two years ago – now, they do.* ***This work is never boring****!"*

Readers of our magazine will know Marcel's supervisor, Marc Pollefeys, and his work at **Microsoft** on the **HoloLens**. How does it feel to be working with such an impressive researcher?

"It is exciting! From his work with Microsoft, we have learnt some things that we probably would not have learnt otherwise. They have very smart people who we can collaborate with. When you talk to Marc, you get the impression he immediately understands what you are telling him. His mind is already one step ahead of you and the ideas are popping up everywhere!"

"His mind is already one step ahead of you and the ideas are popping up everywhere!"

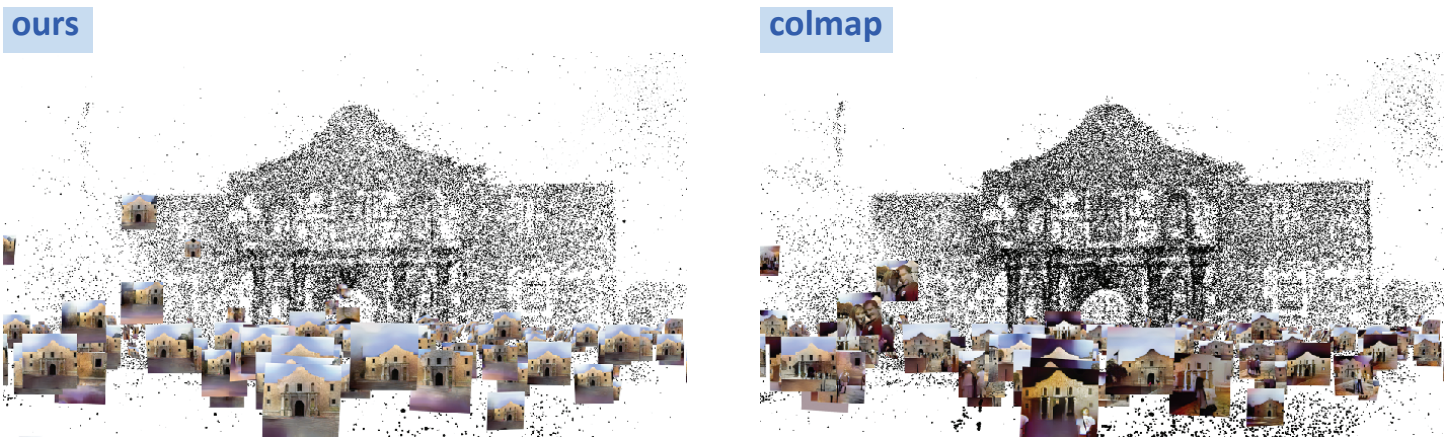
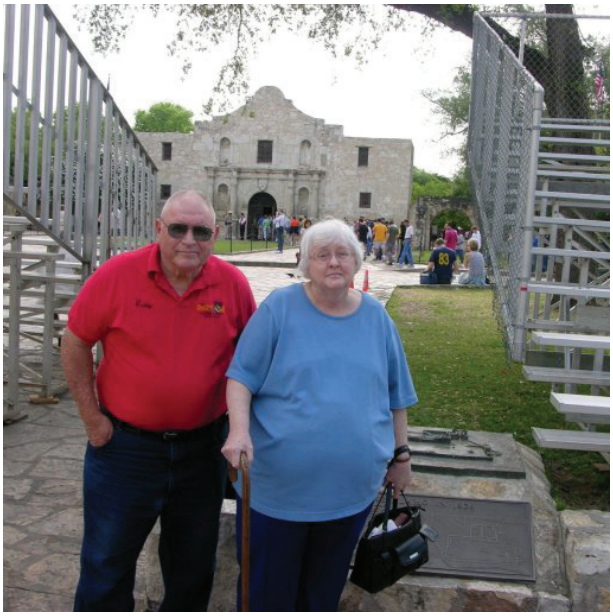


Image reconstructions, displayed with their estimated poses and the triangulated point cloud.



Comparison of the feature inversion with InvSfM on the Alamo dataset.
Top Left: Original image
Top Right: from COLMAP reconstruction.
Bottom Left: from privacy preserving reconstruction.

Would Marcel consider becoming a teacher himself?

"I have not decided yet. I am usually excited by seeing the application of our work, so I would probably move more towards industry. Right now, there are so many opportunities. If you look at the ***Microsoft lab*** *or other companies like* ***Facebook*** *and* ***Google****, they are doing more research in an industrial setting. That is the most interesting thing for me. Teaching is nice, but I still prefer research and development."*

To find out more about this work, visit Marcel's oral [#1031] and Q&A today (Thursday) at 13:20 (UTC +1).

Controlling Style and Semantics in Weakly-Supervised Image Generation

Dario Pavllo is an Italian PhD student at ETH Zurich, under the supervision of Professor Thomas Hofmann and Senior Researcher Aurelien Lucchi. His paper is about generative models and he speaks to us ahead of his spotlight presentation today.

The goal of this work is to **improve the control of generative models**. For downstream applications, it is important to think about what you want to use a generative model for. If you want to generate a dog and use that in a video game or animation, you must tell the model that. While some recent work generates high-quality, almost photorealistic results, you do not have very much high-level control. This model complements those other works but is about improving that control.

*"Specifically, what we do is **generate complex scenes**," Dario tells us. "Not a single object, but multiple objects in a way that we can control the style of those objects. Say you want to generate a car near a tree, for example, but you also want to specify the colour of the car and the style of each object in the scene. We call these **attributes**. Another thing we want to do is use text to control the style of the scene. You can generate a landscape, for instance, and then describe the weather with text."*

"Everyone should jump on the 3D train!"

This approach is **weakly supervised** and does not require labeled data. It contributes a scheme that can be attached to other architectures and the explicit control over style has not been done so extensively before. Recent models for generation of complex scenes use a mask, or a semantic image that describes what is there in every position of the image, to improve the results. This model does not require anything to be hand-labeled by humans. Instead, it uses another model that was trained for object detection to prepare the labels.

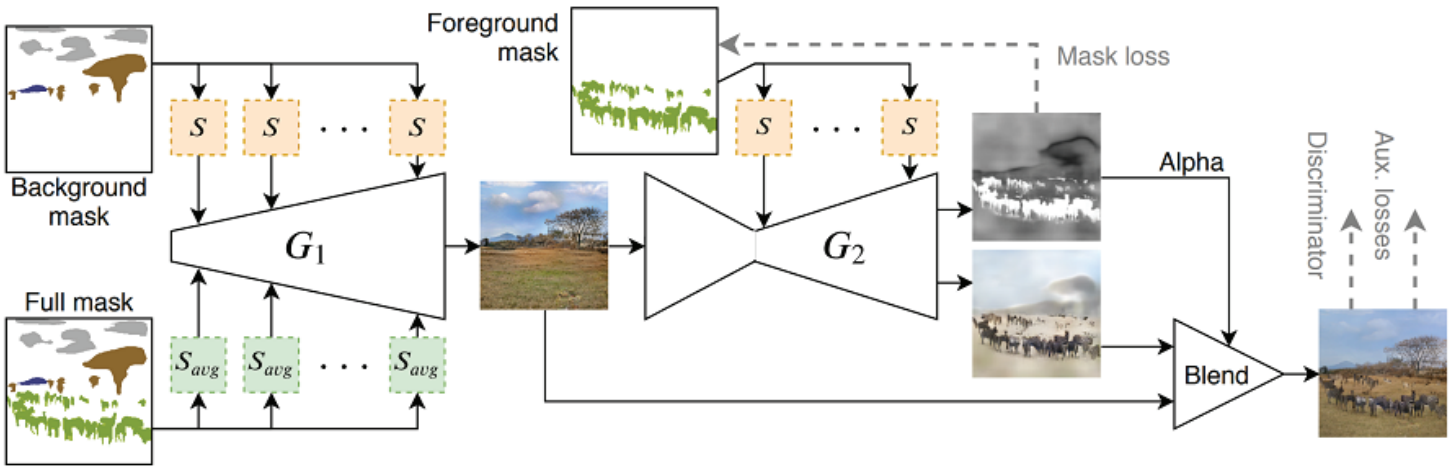
Dario tells us they have observed some limitations on **mask-based approaches** – including this model – and the problem is the generated results stick too much to the mask. *"Right now, if you want to*

generate a car, you have to draw the outline or shape of it," he explains. *"That's fine for simple objects but if you want to draw animals or more complicated objects it presents a limitation. To improve on that, we need to add a step of 3D reasoning because the 3D images we see are actually 3D projections of 3D scenes. Currently, I'm working on integrating this 3D reasoning into generative models so that would be the next step for this work."*

He adds that he has seen plenty of examples where models have failed or produced unexpectedly funny results, as you can see in the images of giraffe, zebra, sheep and dog…

In terms of **computer vision**, the model uses **GANs**, which are the state of the art for generative models and achieve sharp images. The work uses **an architecture called SPADE**, which was proposed at **CVPR 2019** and is also the state of the art, although other architectures could be used with this approach.

"We need to add a step of 3D reasoning because the 3D images we see are actually 3D projections of 3D scenes."



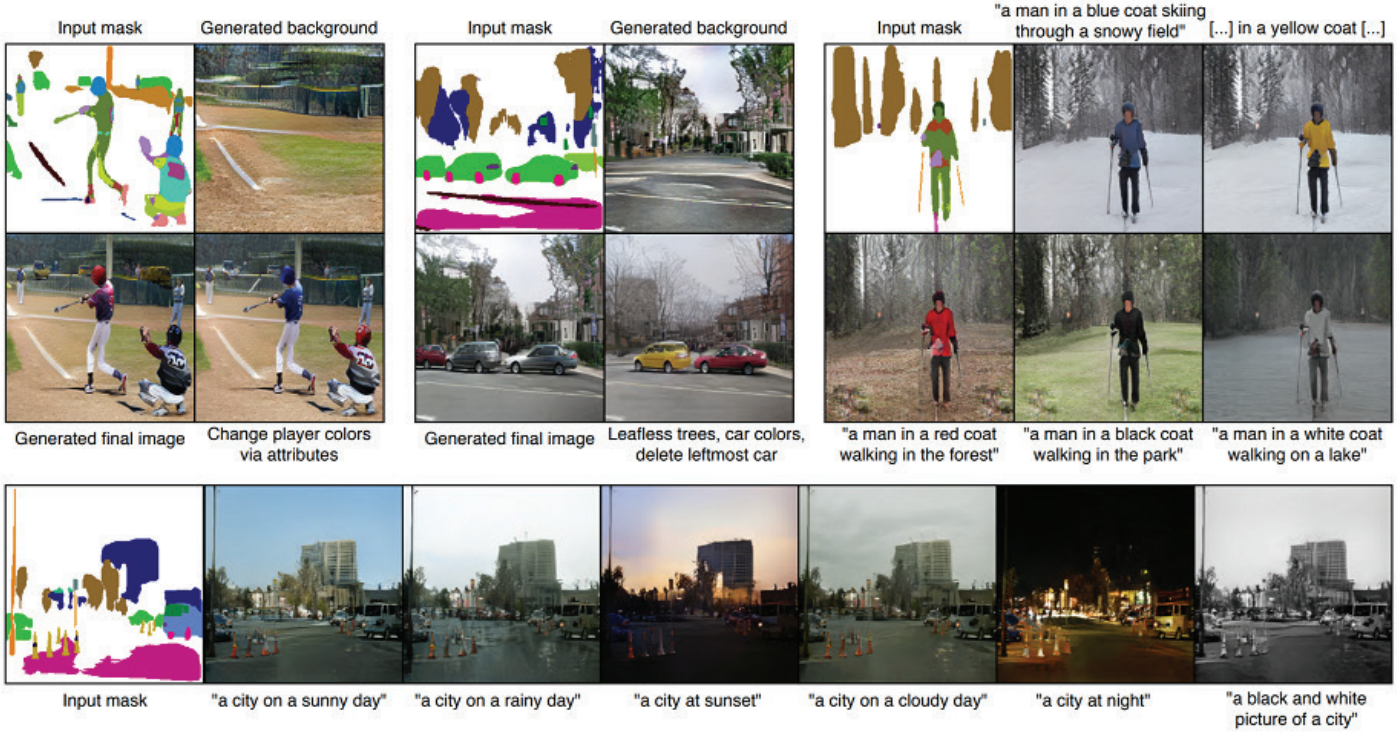
In the real world, this work could be used for **artistic purposes and content creation**. It is not quite at a level where it could generate movies, but certainly simple animations and video games, and Dario expects this to improve in the next few years.

Dario is entering the field of deep learning at a point where it is booming. How does he feel to be a researcher in 2020?

"I think I started at exactly the right time because this field is pretty recent. It started in 2012 with AlexNet – the ImageNet recognition challenge. I started my PhD three years ago, but I had already been doing research one or two years prior to that in machine learning. If you think back 10 years, you could never have imagined models that generate photorealistic images or that can recognize objects. I am so excited to see what is going to happen in the future. There is still a lot of work to do in this direction. ***Everyone should jump on the 3D train****!"*

"... models have failed or produced unexpectedly funny results ..."

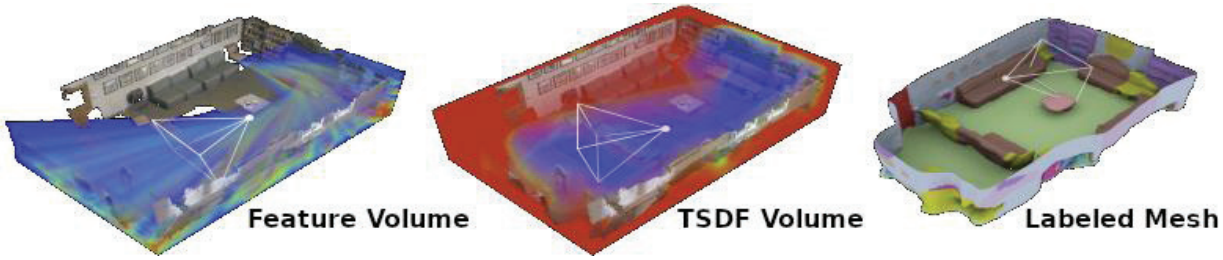
To find out more about Dario's work, visit his live spotlight session [#3219] and Q&A today (Thursday) at 12:30 (UTC +1).



Atlas: End-to-End 3D Scene Reconstruction from Posed Images

We would like to suggest you visit the poster of this paper we liked. This work from **Magic Leap** reconstructs a **3D mesh of a scene from a sequence of RGB images**. While most prior work tackling this problem first predicts depth maps for pairs of images and then fuses the results, **Atlas** simplifies the process by removing the intermediate depth map representation. Instead, they fuse features from all the images into a single voxel volume and then directly regress the 3D geometry. This allows the network to do more reasoning about structure in 3D as well as for end-to-end training. We were impressed by the **visual quality** of their results, particularly in their video below

Author **Zachary Murez** told us that he is excited about the possibility of combining Atlas with ideas from another **ECCV2020** paper that we reviewed on Monday: **"NeRF Representing Scenes as Neural Radiance Fields for View Synthesis"**, allowing for increased resolution, reduced memory, and self-supervised training.



*"The network architecture bears a lot of resemblance to prior **cost volume based stereo depth networks**,"* Zachary added *"but the key differences are that features are backprojected into a canonical voxel volume instead of planes of a view frustum, and voxel occupancy is regressed instead of depth probability."*

Robert Fisher is the Professor of Computer Vision in the School of Informatics at the University of Edinburgh, where he has been on staff for almost 40 years. This year, he is a General co-Chair at ECCV 2020, alongside Vittorio Ferrari, Cordelia Schmid and Emanuele Trucco.

Bob, thank you for speaking to us. As one of the General Chairs for this year's event, can you tell us a bit about the process that has got us here?
Yes, it has been keeping me awake days and nights for much of the last year! It was about four and a half years ago that we started the process of bidding for the conference. Originally it was intended to be in Edinburgh. At that time, ECCV typically had about 1,000–1,400 delegates, and the Edinburgh Conference Centre could take 1,800 delegates. Then it grew a little bit. Then there was that overflow room that we could use with a video connection and that could get us up to about 3,000.

Since then, the field of computer vision has really exploded. Particularly because of the success of deep net technology and the many ways one could apply it. Observing what happened at first Amsterdam and then Munich, we decided to move the conference to Glasgow, which has an Exhibition Centre that can take up to 6,000 delegates. We were all set for that then coronavirus came along, and the Scottish Government turned the Exhibition Centre into an emergency hospital. Ultimately, we are still in the middle of the pandemic, so delegates would not have been able to travel anyway.

How many participants do we have this year?
Currently registrations are at 4,600.

Do you have a message for participants who are attending their very first conference, but are missing out on the in-person experience we all treasure?
That is an interesting question. We do have a couple of innovations that they would not have experienced had it been a traditional live conference, so there are some benefits. One of the important things about these conferences is meeting people and making connections. For junior researchers, it is a chance to meet with some of the more senior

people in the field and get to know them a little bit. For senior people, there is a lot of engaging, talking about potential projects and proposals, and teamwork activity. The online platform works well technically and we have tried to replicate the feeling of a live conference and how dynamic the field is, but it is not exactly the same as being in a room with a few thousand people and getting excited about the papers being presented.

Over the course of your career, you have done a lot of research. Can you tell us more about some of that?
Lately, I tend to work more on practical applications of computer vision and a little bit of robotics technology. My most recent project was TrimBot2020 where we built the first outdoor gardening robot that could trim bushes and prune roses. Before that, I had the Fish4Knowledge project, where we acquired a database of over a billion coral reef fish. Part of what we were trying to do was environmental monitoring, and part was working out what you can do with really large image databases, like recognizing the species of the fish.

These have been European Commission-funded projects and something else that I have worked on lately is building consortia. It has been exciting to help build teams, connect people, and leverage the individual skills of all the team members to do something that is bigger and better.

2 textured 3D scans from a scanning project investigating body shape monitoring during diets

Not all my work has been about the research. I have put a lot of effort into CVonline and now into collecting datasets, which I hope has been valuable to the community. Some great researchers have been collecting and publishing datasets – I have created a few myself – and you can see how being able to compare performance has really driven improvement in the quality of computer vision research. It is not all because of deep learning, although that has been a huge factor as well. It is exciting to play a little part in the gathering and the promotion of these datasets to the community.

Do you find it easier now to connect teams and promote projects than it was when you were a young researcher?
There are certainly more opportunities for that now. The funding agencies recognized that there was too much silo-based research that did not make it much out of universities. There are more

opportunities for projects now that link research to the application. That does not mean there is always going to be a commercial product at the end of the project, but that it is heading in the right direction for something like that. It has bigger impact.

Do you find it easier now to connect teams and promote projects than it was when you were a young researcher?

There are certainly more opportunities for that now. The funding agencies recognized that there was too much silo-based research that did not make it much out of universities. There are more opportunities for projects now that link research to the application. That does not mean there is always going to be a commercial product at the end of the project, but that it is heading in the right direction for something like that. It has bigger impact.

When I interview researchers, I often ask them what the application of their work will be in the real world. Some say that what happens afterwards is not up to them. Do you think something needs to be done to help facilitate this relationship between research and application?

Certainly, at least in the UK, the government feels there is a gap there. There are several different funding schemes which help with that translation from the lab to commercial work. With the European Commission funding programmes, part of what your proposals are reviewed upon is potential impact. That does not necessarily mean commercial – it could be societal or healthcare impacts. I do not have a

3D reconstruction of a rose plant with cut points identified

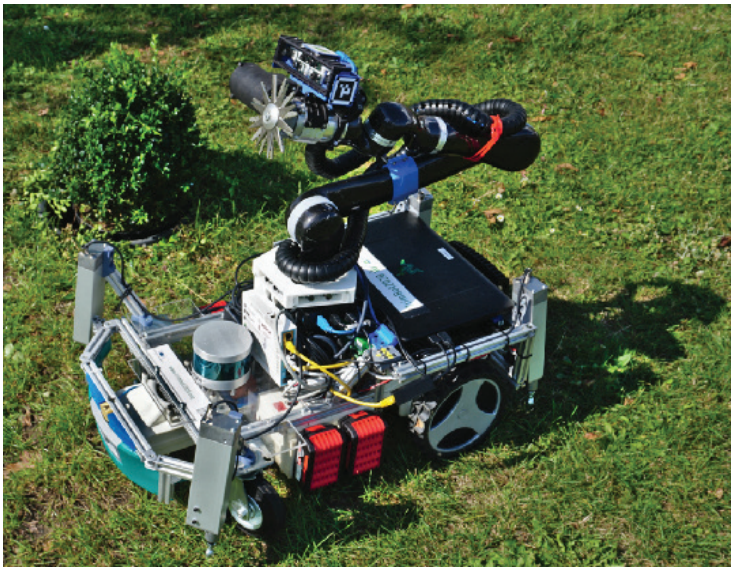
problem with theoretical research, but personally, I have always found practical projects lead to better theory and new ideas. I am not the kind of person who sits down with a piece of paper and says, *"What is there new to work on?"* I find it more stimulating to work with real-world problems and identify gaps in the research field that way.

You have mentioned the importance of funding. Is money the only thing that is needed?

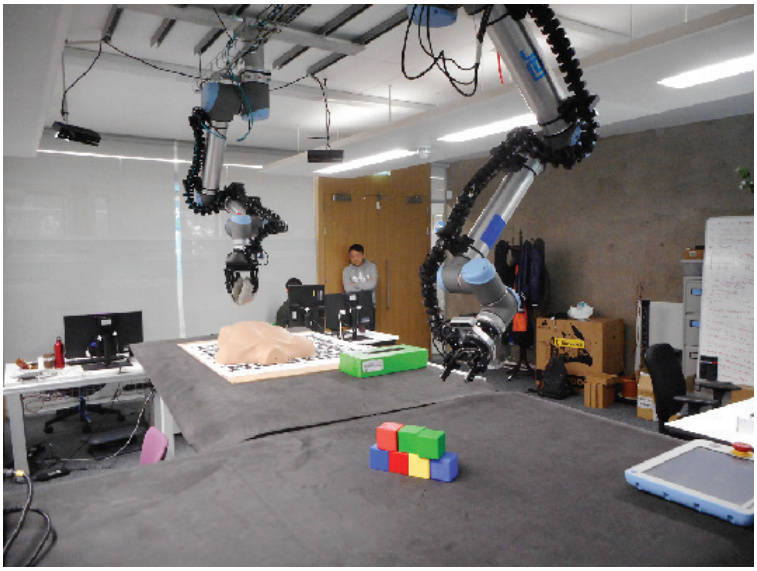
Good question! Money is a very important part, but teamwork and collaboration are too. I am constantly surrounded by bright people but somehow it feels like there is a multiplicative effect. When you get a couple of these bright people together, it is more than just a sum, it is like the great work. product of the ideas. These consortia-based projects have been exciting in terms of cross-fertilization of ideas. By integrating the work together, each partner sees a need for something new and that stimulates their own research. That is not to deny there are individuals and individual groups who are doing great work.

Thank you, Bob. Before we go, do you have a final thought for the community?

Here is some advice that I would give to young researchers: find things that you really like to do. Maybe you will be rich and famous one day – I think there is a lot of chance and luck involved in that – but if you find something that you enjoy doing, it means every day you get up and say, *"Today I am going to work on this great thing that I love!"* That feeling can sustain a very long and happy career.



The TrimBot2020 gardening robot, with the topiary trimming end effector



A dual arm ceiling mounted UR10 robot workcell for surgical assistance robot research

SOLAR: Second-Order Loss and Attention for Image Retrieval

Tony Ng is a PhD student at Imperial College London, supervised by Dr Krystian Mikolajczyk. His work called SOLAR, which stands for Second-Order Loss and Attention for Image Retrieval, explores two second-order components in deep learning. He speaks to us ahead of his poster presentation today.

Tony's PhD research topic is **visual-based localization**. The first step to do this in many pipelines is image retrieval. In image retrieval, database images are sorted according to how similar they are to a query image. A way to do that is to transform the image into an embedding through a **convolutional neural network**, **such as ResNet or VGG**, so it can do the sorting. The final product after the transformation from the image to the embedding is called the **global descriptor**.

This work focuses on **landmark retrieval**. Landmarks present some specific challenges. They can all look very similar, or very different, so it can be hard to tell which class they belong to. Sometimes you can have more than one landmark in a single image. Also, the image could have very different conditions, extreme zooms, or viewpoint changes.

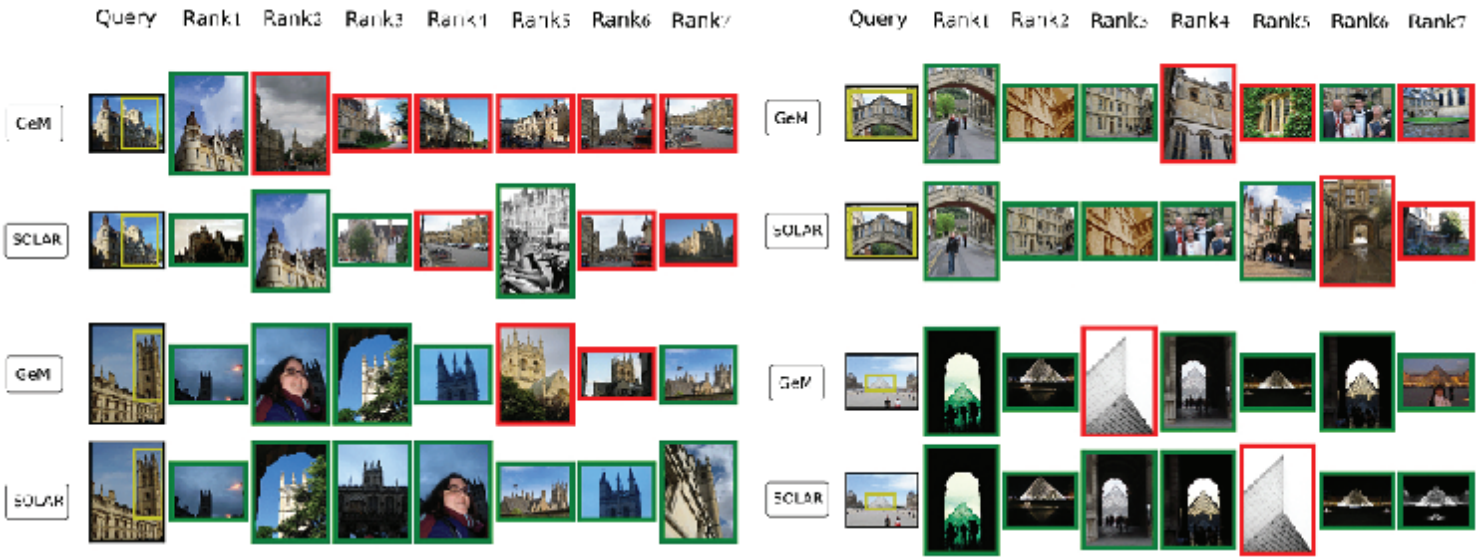
The method described in this paper and shown in the video below uses the self-attention mechanism from **Natural Language Processing (NLP)**, which calculates the second-order correlation between two feature maps. It was recently introduced to the computer vision community through the non-local block.

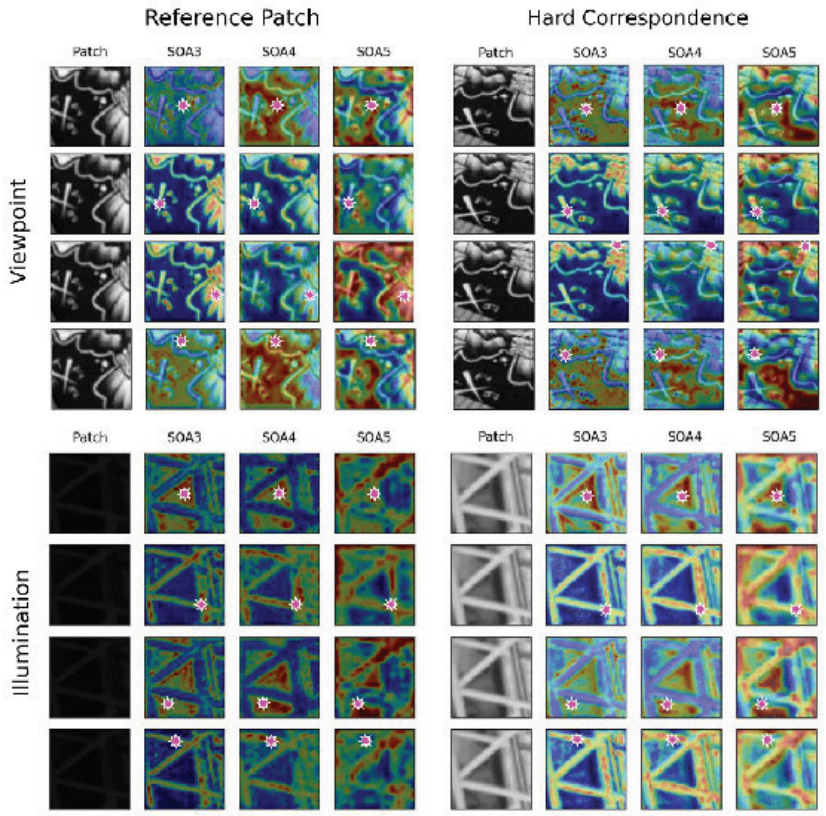
Tony explains how this could help resolve some of the challenges mentioned above: *"In landmark retrieval, there are two major camps for calculating the global descriptor. One camp uses local features detected on an image and they do aggregations. That method is more accurate than the method I'm trying to tackle – which uses one single CNN and then pooling on the last feature map to get the global descriptor – but it's slower and requires more memory. In **visual localization problems**, speed*

and applicability are important, so I am focused on improving these global pooling methods."

One major problem Tony found with these methods is that they only calculate the first-order measure in terms of how to do the global pooling. Each feature in the feature map does not know about the other features and it is taking an average of the feature map to get the global descriptor. When he reviewed papers on self-attention and non-local blocks, he found it interesting that each feature could have the information of other features, without adding too much computational complexity. He incorporated this second-order attention in this work, alongside second-order loss, which was recently introduced in local descriptor learning by co-author **Yurun Tian**. Local descriptors work only on a small patch, instead of a full large-size image.

"The statistics between large images and small patches are quite different, so I thought it would be nice to also try it on global descriptors for large images and for landmarks," Tony tells us. *"What I found in this paper was that adding this second-order loss on top of my second-order attention gives the best results. It greatly improves upon the global methods and even matches some of the local methods that require **10 times the running speed, 10 times the running time**, and **three or four times the memory cost**. This is a very promising result. I'm sure that the communities of image retrieval, descriptor learning, and visual localization will be very interested to hear about it."*





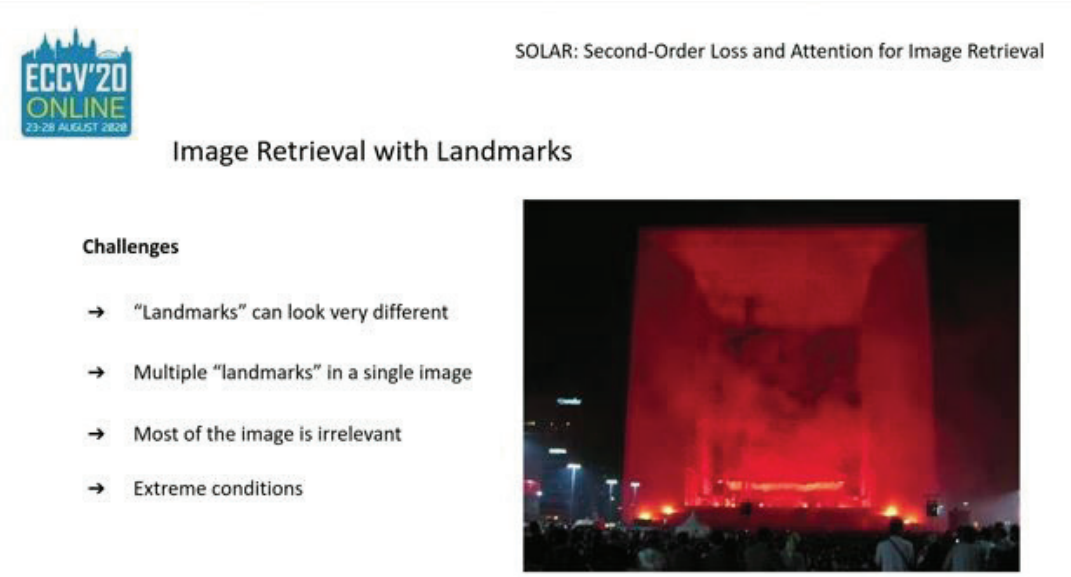
Thinking about adapting the work in the future, Tony says he is currently working on including camera poses and some geometry into the descriptor part so that he can make the descriptor more optimized for the camera poses, not just retrieving from a database of images. He describes his work on the geometry part so far as "tricky" because it is hard to find representations of geometry that are invariant to a lot of constraints seen in this kind of image retrieval dataset. He originally assumed a model that worked well on image retrieval would also work well on camera localization, but this is not necessarily the case, as the two are very different.

"In camera localization you have lots of different parameters to think about," he explains. "You have the camera models. They can have different focal lengths. Images can be taken in different conditions and from extreme viewpoints. To get the global descriptor from one image to just a single factor, you have to do some sort of averaging across the feature map, so ***you lose all spatial context****. Imagine for normal human perception, for human localization, you have to know exactly which part of the building belongs to top-left of the image or bottom-right of the image. You have to know those things, but when you turn them into numbers, they're all lost. We need to find a way to make sure that* ***the output descriptor that we have includes that information****, so that we actually know whether these two images come from similar poses or not."*

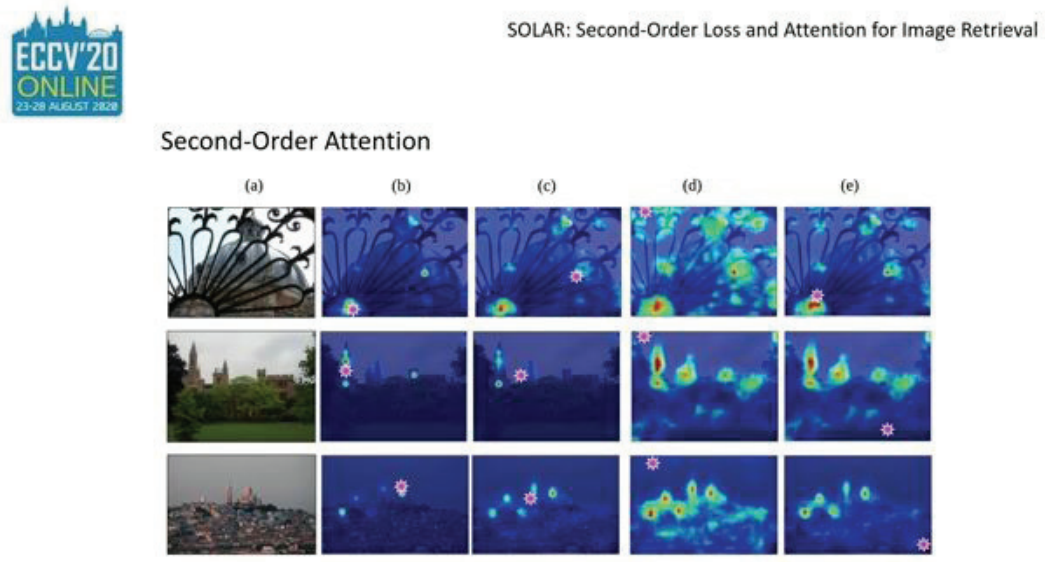
Tony has a final thought for the community: *"I think we should embrace more of the techniques that are being used outside of this quite limited field. Embrace the innovations from across other computer vision and machine learning communities. Although I'm probably too young to make an observation like that! I've had my first paper accepted at ECCV and still have a lot more work to do. But in general, I think any sort of innovation and breakthrough in the field of descriptor learning and visual localization would be very nice."*

To find out more, visit Tony's poster presentation [#4388] today (Thursday) at 14:00 and tomorrow (Friday) at 00:00 (UTC +1).

Long Video



Short Video



ECCV'20 ONLINE Thursday DAILY

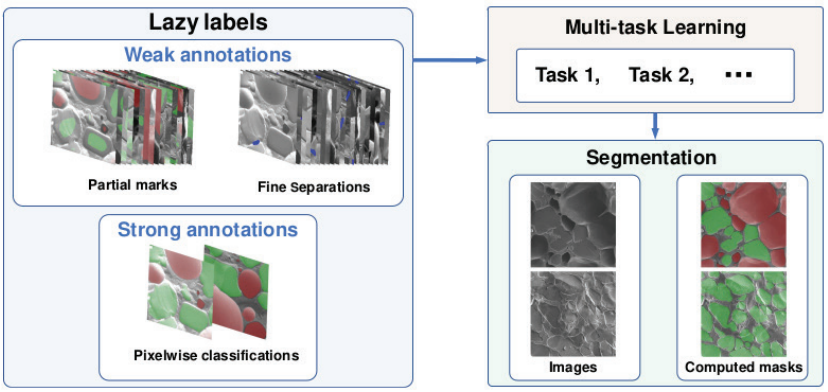
by Jan Funke

BioImage Computing

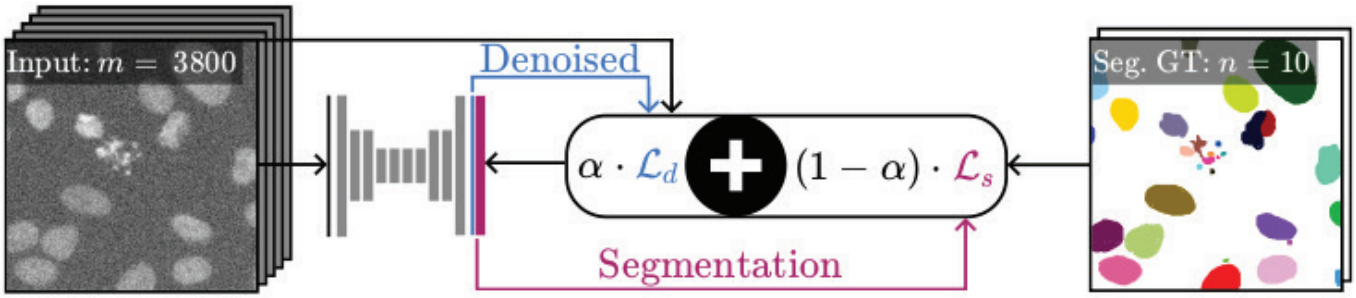
The 6th edition of the BioImage Computing workshop (BIC) was held virtually last Sunday, August 23rd. Although relatively young, BIC has already become a high-quality, well-attended workshop bringing together researchers working in the intersection of computer vision and the life sciences. This year, BIC was organized by Jan Funke (HHMI Janelia, USA), Dagmar Kainmueller (BIH and MDC Berlin, Germany), Florian Jug (CSBD and MPI-CBG, Dresden, Germany) Anna Kreshuk, (EMBL Heidelberg, Germany), Peter Bajcsy, (NIST, USA) Martin Weigert, (EPFL, Switzerland), Patrick Bouthemy, (INRIA, France), and Erik Meijering, (University New South Wales, Australia).

Following an open review process through **OpenReview**, the organizers selected **17 submissions** for presentation at the workshop. Those submissions are representative of the wide range of **exciting computer vision challenges we encounter in the life sciences**, e.g., computational imaging, denoising of microscopy images, bioimage segmentation and classification, multi-modal registration, and tracking of cells and animals in videos. In addition to those submissions, the BIC workshop featured six invited speakers this year. The following highlights some of our **best picks of BIC 2020** (all images belong to the authors of the papers).

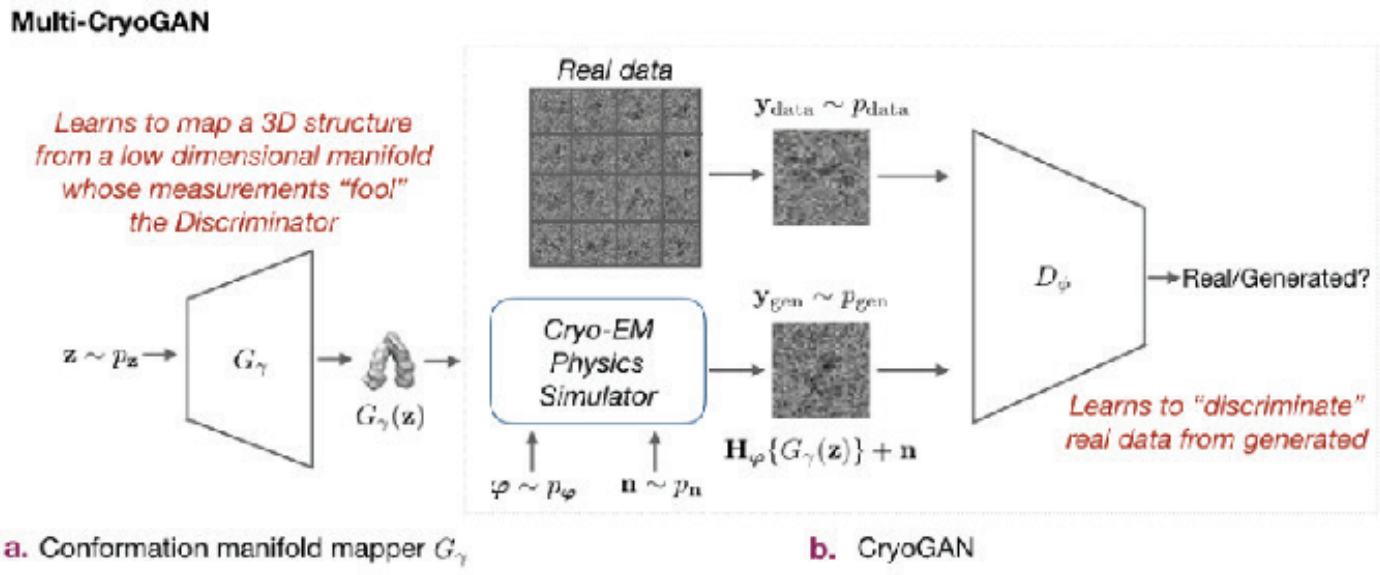
BioImage Computing is as diverse as the questions that experimentalists like to answer. Consequently, machine learning methods rarely transfer between domains and human generated ground-truth is needed. This is a time-consuming and limiting step in many experimental labs, and the question how to minimize manual efforts without sacrificing accuracy is of paramount importance. This BIC saw two submissions that take this question head-on:



Learning to segment microscopy images with lazy labels by **Ke et al.** explores the question of how much can be learnt from "lazy" labels, i.e., rough strokes to detect and separate cells instead of precise ground truth segmentation. Using a multi-task learning objective tailored to those labels, the authors show that the use of additional lazy labels greatly reduces the need for precise labels: their method achieves **at least the accuracy of a conventional supervised method while using only around 10% of the available precise labels**.

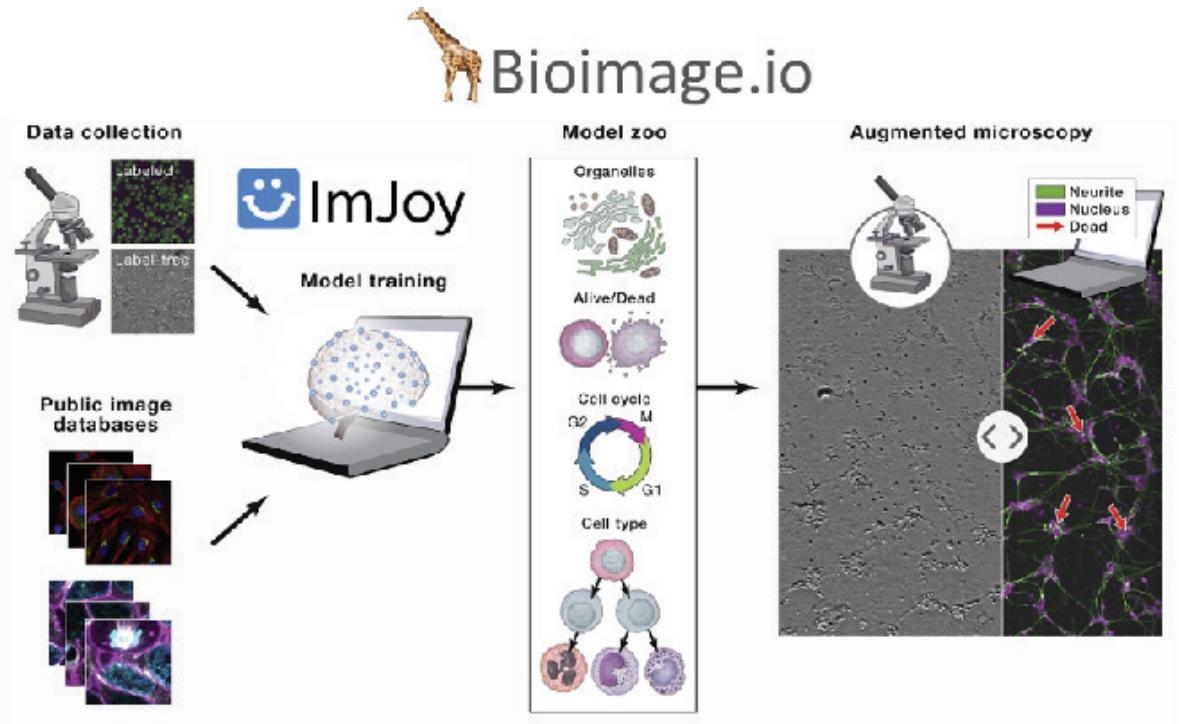


The lack of ground-truth can also be compensated for using learnt priors from the images alone, as demonstrated by **Buchholz et al.** DenoiSeg: Joint Denoising and Segmentation is a method **to jointly train a network on a self-supervised denoising task** and a supervised segmentation task. Extensive experiments demonstrate that this co-learning allows the network to make better use of the available ground-truth.



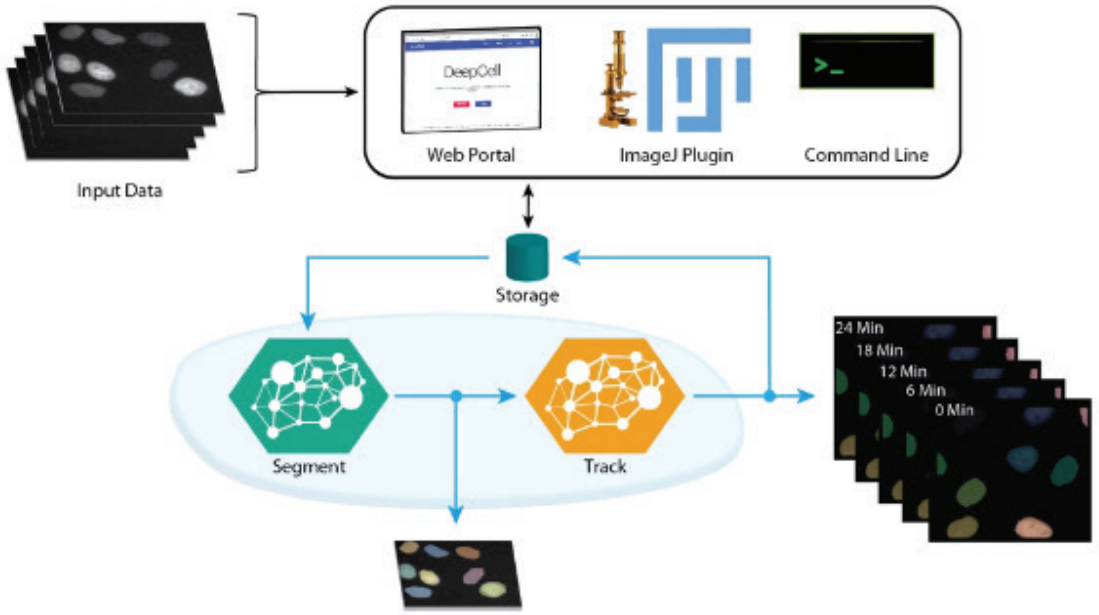
The same philosophy of **leveraging intrinsic statistics in bioimages to compensate for missing ground-truth** annotations is taken to the extreme when it comes to Cryo-EM. Here, the task is to infer the structure of extremely small particles (e.g., individual proteins) from many noisy electron microscopy (EM) projections. **In Multi-CryoGAN: Reconstruction of Continuous Conformations in Cryo-EM Using Generative Adversarial Networks**, **Gupta et al**. use a GAN to transfer the classical pose and conformation estimation into a distribution matching problem: A distribution of latent variables is learnt that represents pose and conformation of the imaged particles. Synthetic 2D images are then produced through a forward model to **produce realistic projections that are indistinguishable to a discriminator from real measurements**. The whole model is trained end-to-end in an **unsupervised manner**.

Being in the intersection between computer vision, machine learning, and life science, contributions to the field of BioImage Computing are not just academic proof of concepts: successful methods are not only improving over the previous state of the art, but are also easily usable by experimentalists. This point was clearly made by several invited speakers:



Wei Ouyang from the KTH Royal Institute of Technology presented (amongst many other topics) **ImJoy**, a **computational platform for the deployment of deep learning solutions**. With ImJoy, experimentalists have access to **state-**

of-the-art deep learning solutions and can apply them to their own datasets without the need to be a machine learning expert. This is an idea that is shared with a similar tool, **Ilastik**, which is already well established in the community. In his invited talk, Constantin Pape from the EMBL in Heidelberg shared the future plans around Ilastik: not only will **deep learning** be part of it, but Ilastik will also support models from the platform bioimage.io, a model zoo for bioimage analysis, to bring progress made in the community to the hands of experimentalists.



In his invited talk, **David Van Valen from the California Institute of Technology** gave us insights how his group is addressing the eruption of large datasets and the consequent computational demands to process them. Using **scalable pipelines with Kubernetes**, David's group designed DeepCell, a platform for the segmentation and tracking of cells in real-world-sized datasets.

Finally, **Kristin Branson from HHMI Janelia** and **Virginie Uhlmann from the EMBL-EBI in Cambridge** went full circle and showed us how to use computer vision and machine learning techniques in the context of downstream analysis. Kristin presented her group's work on the tracking of animal poses in videos and the subsequent characterization of behavior, while Virginie reminded us that not everything has to be deep learning: Although the analysis and classification of shapes has a long history in classical computer vision, their application to bioimages is of great value, non-trivial, and new efficient and applicable methods are needed for this domain.

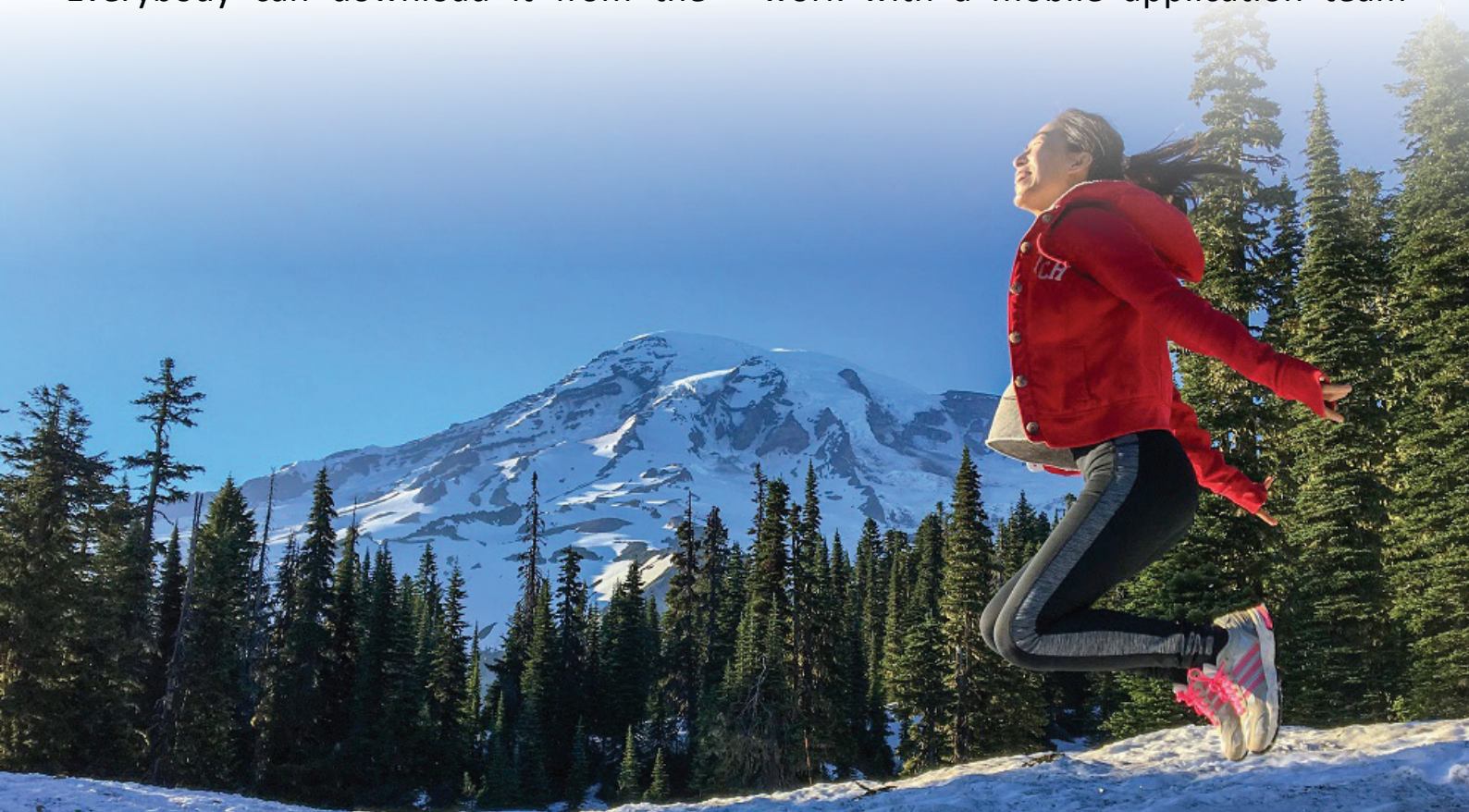
Xin Lu is a senior engineering manager and Scientist at Adobe. With a PhD in Imaging and Machine Learning Applications, Xin has been invited to attend the first international workshop on Bodily Expressed Emotional Understanding (BEEU), where she will be speaking about her own experiences and developments in the software industry.

Xin, what are you doing at EECV?
I participate in the Bodily Expressed Emotion Understanding workshop. I am a speaker there and participate in a panel discussion.

Can you tell our readers about your work?
Yes! For the past couple of years, I have been working on a mobile camera app that has just recently launched. Everybody can download it from the Apple store or Google Play store. It is called 'Photoshop Camera'. I hope this does not count as an advertisement [*laughs*]. It includes a lot of my research and development work. It has a lot of interesting features. Basically, it is an AI-based camera app for pictures that everybody can use in their daily life.

You can do as much advertising as you want: this magazine is made for advertising your work!
[*both laugh*]

How did a young scholar graduated from her studies, move onto a PhD, and then evolve into developing this app?
Well that is a very fair and interesting question! After I graduated from Penn State, I joined Adobe. Soon after that, I realized there was an opportunity to work with a mobile application team



to develop interesting mobile features for users. Initially the development target was for Photoshop Mix and Fix. Since I'm coming from a background of developing machine learning applications in imaging and computer vision, and the team is good at mobile app development. So naturally, I am thinking whether there is anything that I can contribute to the team by matching the two different expertise. I'm thinking, maybe I should start working out efficient neural networks and work out on-device runtime to do model inference, so that different models can run on mobile devices without the need to transport users' private data to the cloud. Running neural networks on mobile devices is beyond most people's imagination in 2015. We started there, and then a couple of months later we had an initial prototype! After that, there was a chance to do a new camera application on mobile devices. With a camera app, the most important thing is to make it real-time, and to make it AI-assisted, smart, and intelligent. My early work of running neural network on device actually became one of the fundamental parts of this mobile camera app.

It's one thing to know how to do it, but it seems that you really like what you are doing.

Can you share what it is like working for Adobe?

Adobe is a very interesting company. It has a lot of people that are passionate about imaging. One of the special aspects about this camera app is actually to try to help people to get creative. We provide a lot of creative effects that, in the past, only professionals could do with Photoshop. But now everybody can share this creativity in the capture space easily. They can take pictures with professional-level effects with just a couple of clicks.

Which kind of devices support this app?

It is available to both IOS users and Android users.

Xin with Adobe founders Charles Geschke and John Warnock

Your story with Adobe didn't begin here.

You're right, I did three summer internships with Adobe. During the internships, I have been working on a couple of different projects related to imaging, low-level image denoising, and high-level image curation. All of those were very interesting projects! Adobe has a very large group of researchers, and we all participate in computer vision events and top conferences regularly.

So first, you did a summer internship with Adobe, and you enjoyed it. The next summer you do another internship at Adobe, you enjoyed it. Then you go back the next summer for yet another summer internship with Adobe, and... you enjoy it! [*both laugh*] Is that true? And when you finish your studies you ask yourself: *"Where am I going to work?"* At Adobe, of course! [*laughs*] What attracted you so much to this company?

Well, if you want me to name one thing, I will say it is the artistic and creative aspects of Adobe. The first time I entered the Adobe building, I saw the decorations on the wall are all really nice - very artistic, clean, and very fascinating. I like those decorations.

I would like to make a game with you.

Okay! [*laughs*]

Let's say you have a little sister, and you want to convince her to come and work at Adobe after she finishes her studies. What would you tell her?

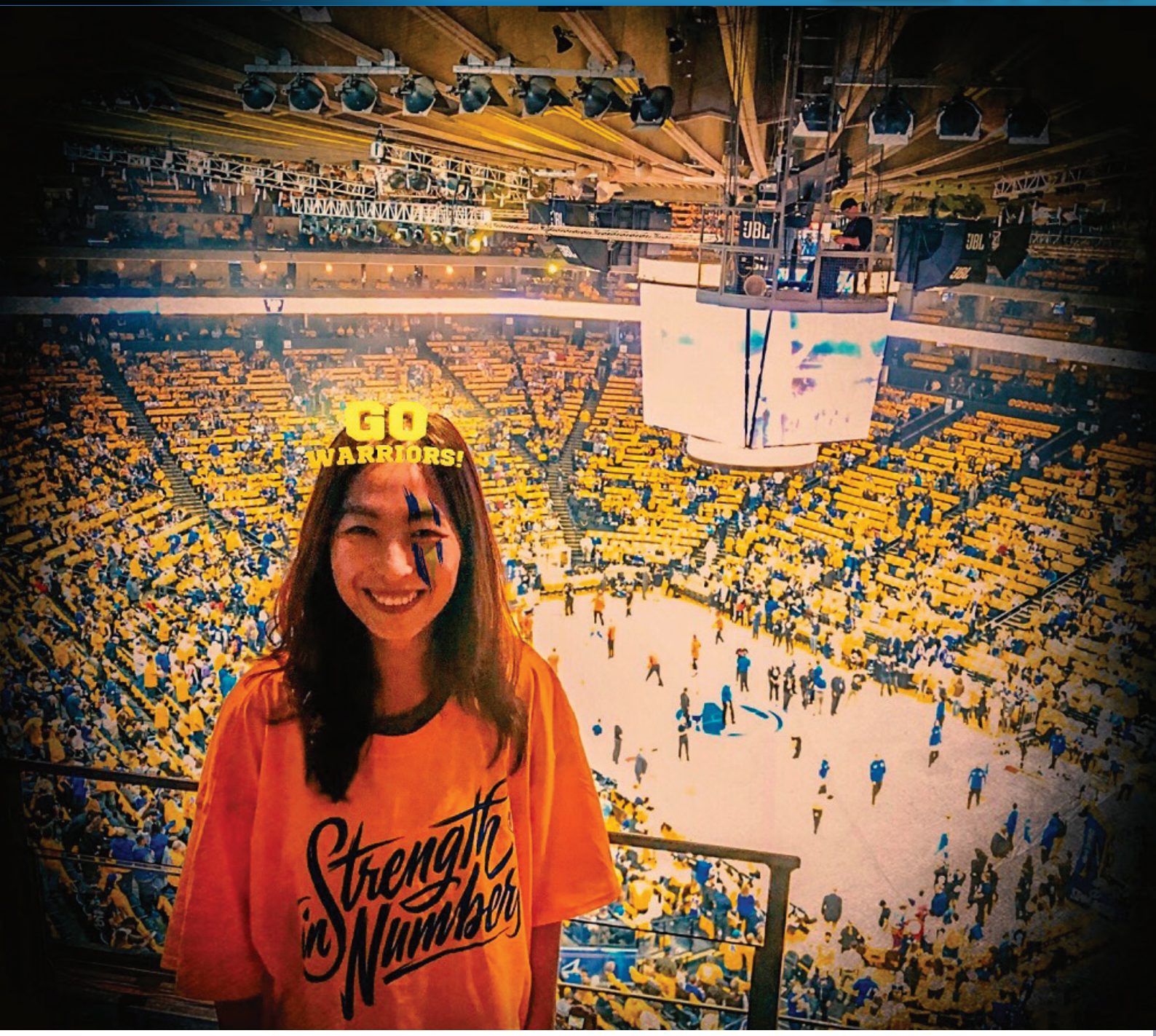
That's definitely a fair question! I would say people have different

choices, and the whole industry in software development is a very fascinating, innovative area. There are a lot of opportunities for individuals at Adobe. People here are very nice and collaborative, and the company culture is very good. Every individual has the chance to share and communicate their ideas in a friendly and open manner. On the other hand, we have the chance to continue learning new things and explore different types of work. Opportunities are always open so that individuals can freely change their routes and choose different types of work - A sense of rotation, that kind of thing. It's just very open and friendly!

I would like to play another game. If you didn't choose Adobe, which company would have been the second-best for you?
[*laughs*] That's a very interesting question!

Thank you! That's why I'm here... [*both laugh*]
The answer is... Well, I don't know! [*laughs*]

Well, the main thing is that you are excited about where you are! That's more important than choosing the second-best. You are not originally from California. Is that right?
No, I am not. I grew up in Tianjin, China. I studied electronic engineering there, and one day in my senior year I got a chance to talk to a professor who was working on a computer vision project. In those days, computer vision was really at an early age. So, my first project was related to face recognition. Immediately, I felt interested in that area. I decided to spend more time there. Then I got an internship opportunity in Microsoft Research Asia, and I did a 10-month internship there working on large scale data mining and research-related projects. This was also related to imaging. I felt interested in research. I wanted to do a PhD, which is what brought me to Pennsylvania. I spent 5 years in Pennsylvania working on very challenging research projects in imaging, computer vision, and image aesthetics and emotions arising from those things. That's part of the reason



why I'm a speaker in this workshop for bodily expressed emotions. After that, I moved to California, and that's it! [*laughs*]

Most of our readers are not from China. Can you tell us what it was like to grow up on that side of the world with very different cultures, habits, and styles of life?

Yeah sure! Firstly, I will say that in my generation the gap between East and West is not as big as you may imagine. People in my generation, especially Asian students, embrace science. We embrace engineering and are passionate about it. Then we go to graduate schools in America to study. Actually, I don't feel that there is a super strong difference when I study in American schools.

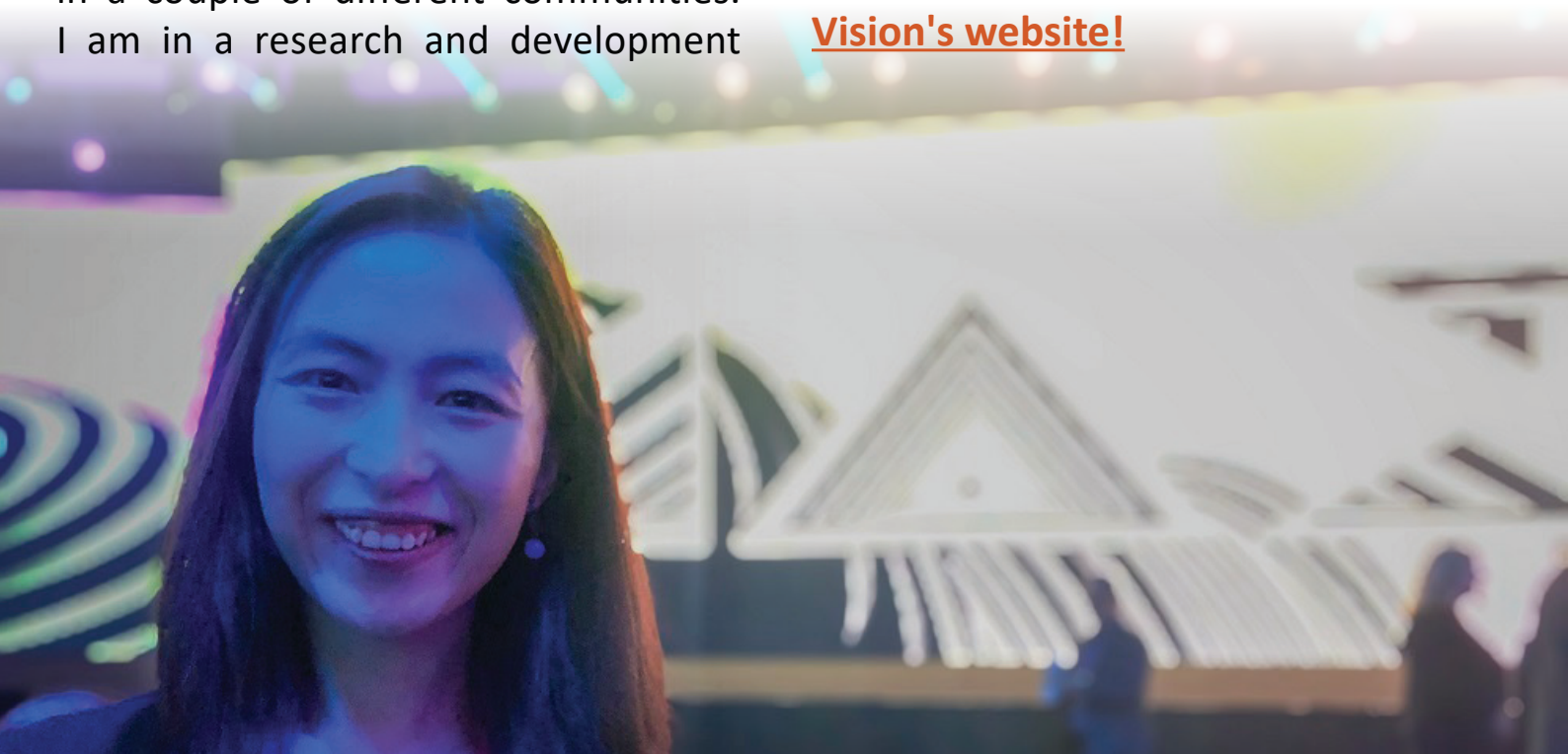
We have spoken a lot about the past. Can you speak a little bit about the future?
Sure!

What do you dream to achieve by the end of your career?
As I work in the engineering and scientific area, one of my goals is to contribute to the community and contribute to all the individuals with what we can do scientifically, to deploy and share the scientific learnings through some products to a lot of individuals in the world who have no idea about science. They may not have much idea about technology, but they can make use of this and enrich their life.

What should the community do or change to make it even easier for you? Do you see anything that needs to be improved?
Yeah, I appreciate the opportunity to be in a couple of different communities. I am in a research and development community, and I also participate in product development. So, I think in general, the world is starting to become more open and friendly to the younger generation. It would be great if the community could share some thoughts from senior people based on their experience and their own life, sharing their life learnings with some of the younger generations. That would be very nice because it would help us, and it may even inspire them as well.

I have one last game for you! I will interview you again in ten years when you will be a senior scientist, and you will inspire all the new, younger generation of grads. Do we have a deal?
Yeah, definitely! What an interesting idea! I hope that what I do in my life could potentially inspire the next generation of people.

Find out more than 100 interviews of Women in Computer Vision on RSIP Vision's website!



Keep in touch with ECCV also after ECCV!

RSIP Vision is distilling the best of the best for you in the BEST OF ECCV.

It will be published on Computer Vision News in September.

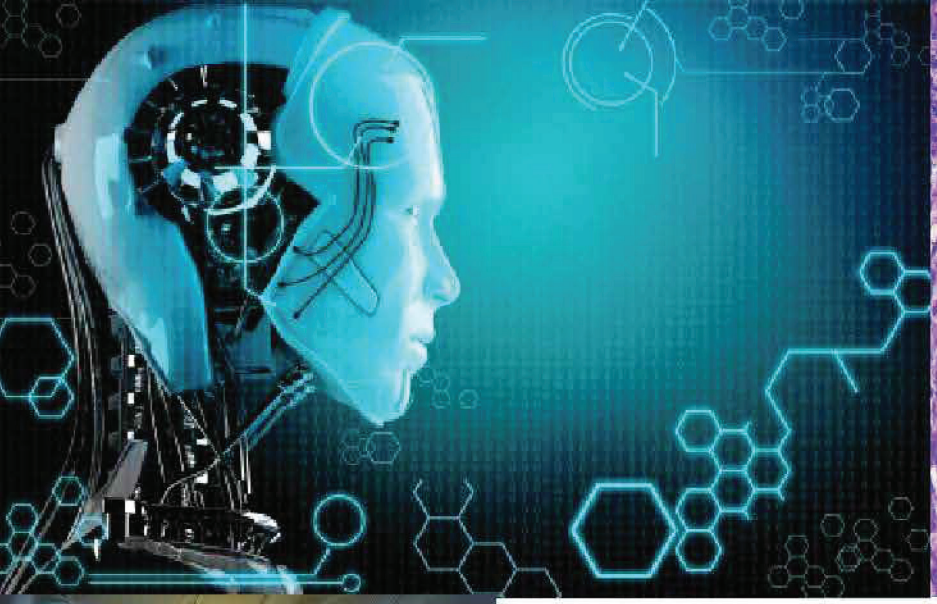
Subscribe now for free and get it in your mailbox as soon as it is published!

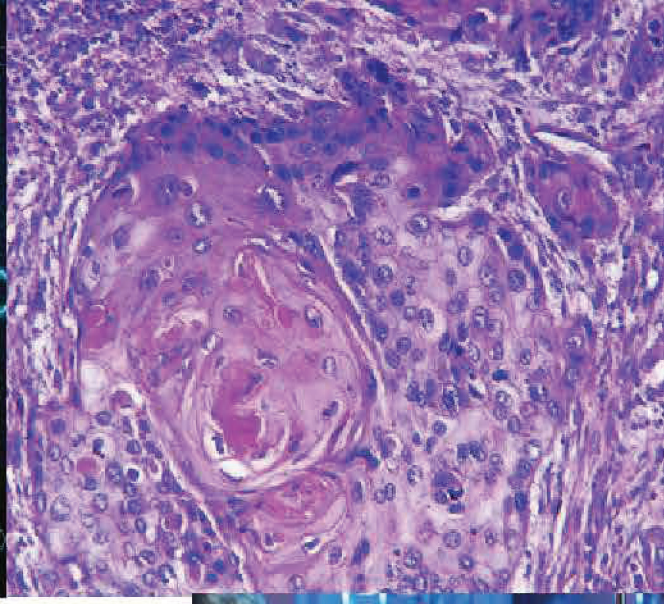
RSIP VISION
Global Leader in Computer Vision and Deep Learning

Take us along for your next AI projects!

ICCV 2021
See You There!

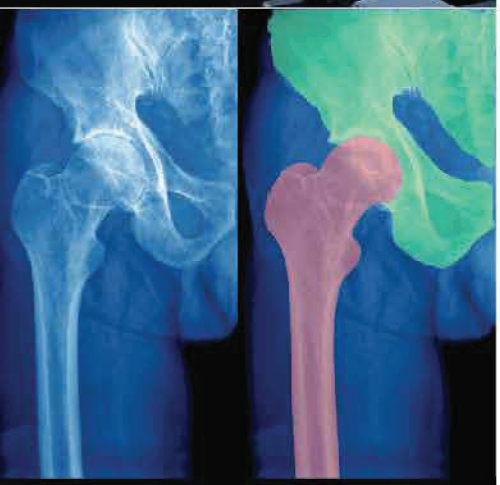
... in person ...





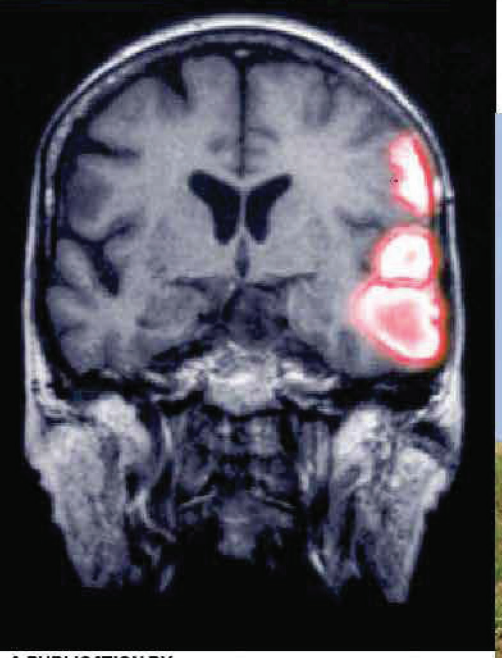
IMPROVE YOUR VISION WITH

Computer Vision News



SUBSCRIBE
CLICK HERE, IT'S FREE

The magazine of the algorithm community



A PUBLICATION BY

RSIP VISION

Global Leader in Computer