DECEMBER 2019

Computer Vision News

The Magazine of the Algorithm Community

Best of ICCV 2019
36 pages

Exclusive interview
Marc Pollefeys

Women in Computer Vision
Georgia Gkioxari

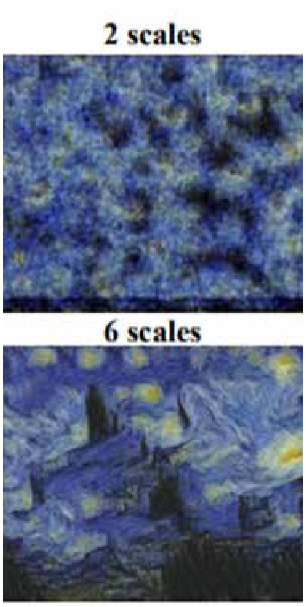
We Tried for You
Open3D for 3D processing

RSIP Vision Projects
Medical Segmentation

Project Management
Challenge-Driven Project Management

Computer News Events

Research Paper Review:
SinGan - Award Winner at ICCV

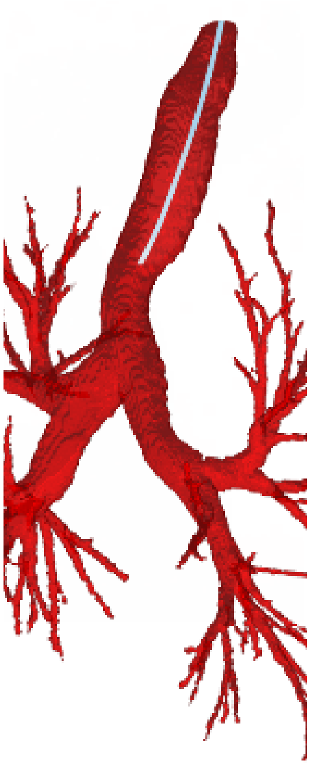


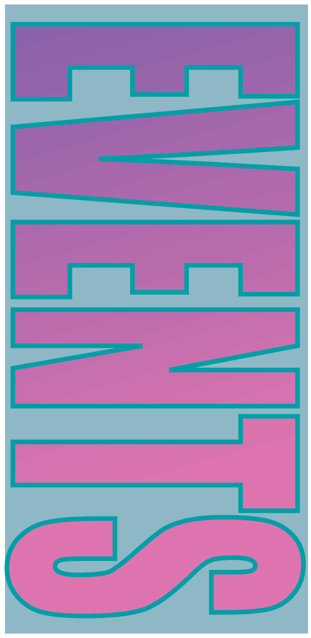
Editorial

Dear reader,

We proudly present a special edition of **Computer Vision News**, including our choice for the **BEST OF ICCV 2019**. We open with Amnon Geifman reviewing **SinGAN**, the recent winner of the best paper award at ICCV 2019. Along some of the very best works from the conference in Seoul, you will read our exclusive interviews with general chair **Marc Pollefeys** (about the Microsoft Hololens, ETH Zürich and more) and with **Georgia Gkioxari** (about her fascinating work with her highly-praised team at Facebook AI Research).

Computer Vision News represents just one of many projects by **RSIP Vision** in our ambitious quest to share the best AI knowledge with our community. Be part of it! Subscribe to the magazine, share it with colleagues and friends, join our Meetups





and Webinars, meet us at the main conferences and, most importantly, take us along on all your deep learning projects!

Enjoy the reading!

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

Computer Vision News

Editor: **Ralph Anzarouth**
Engineering Editor: **Amnon Geifman**
Head Designer: **Rotem Sahar**

Publisher: **RSIP Vision**

SUBSCRIBE

Follow us:

by Amnon Geifman

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen SinGAN: Learning a Generative Model from a Single Natural Image. We are indebted to the authors (Tamar Rott Shaham, Tali Dekel, Tomer Michaeli), for allowing us to use their images. This paper won the Best Paper Award at ICCV 2019, last month in Seoul.

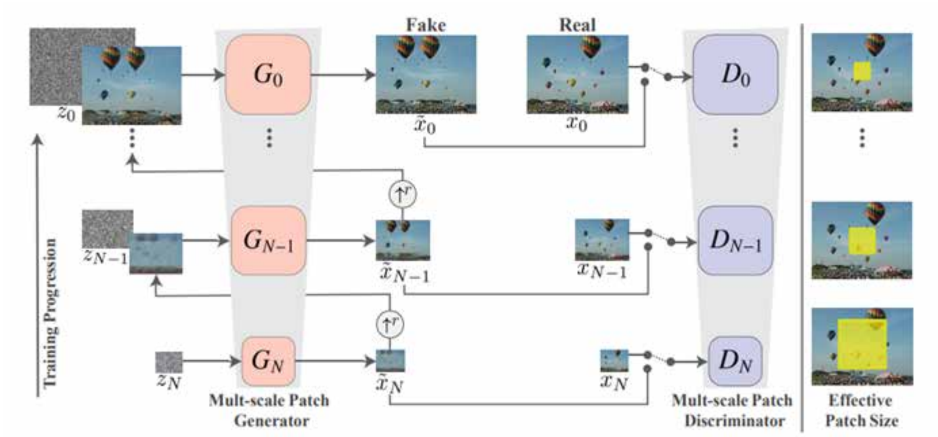
In the last few years, two subfields of neural nets have developed: the first is internal learning which is a learning paradigm that exploits the internal distribution of patches inside a single image to perform several computer vision tasks. The second is GANs- generative adversarial networks which, as we all know, is a framework that mimics a distribution of a given dataset by training a generator that fools a discriminator by generating good looking images.

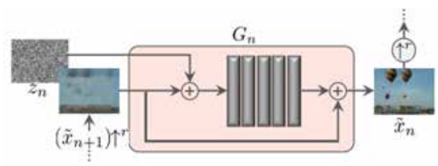
Today's paper incorporates these two fields to formulate a novel framework that trains a generator from a single image. This generator learns to capture the internal distribution of patches in a single image by training on different scales of the same image, against patch discriminator. This enables the model, at test time, to generate high quality diverse samples based on the original image. The paper demonstrates the applicability of the framework to many computer vision tasks such as super resolution, harmonization, editing, paint to image and more.

SinGAN Model

The goal of the model is to learn a generator that captures that internal distribution of patches in an image. To this end, the authors suggest a generative framework consisting of a hierarchy of patch-GAN, each applied on a different scale of the image. Specifically, the model consists of a pyramid of generators $\{G_0,G_1,..,G_N\}$: each generator is trained against a different scale of the input image $\{x_0,x_1,..,x_N\}$, where x_n is the downsampled version of the image x by a scale factor r^n>1. At every scale, G_n learns to generate an image-sample in which all overlapping patches cannot be distinguished from patches in the original down sampled image x_n.

The input to every generator G_n is a random noise image z_n, and the generated image from the previous scale $\widetilde{x_n}$ is upsampled to the current resolution. The coarsest level generates an image from a pure random noise. The model is easily explained by the following figures:





To train the model, the authors suggest a sequential scheme that goes from the coarsest level to the finest one. Once a level was trained, it is kept fixed during the training of the other levels. The loss of the SinGAN is composed by a linear combination of two losses, reconstruction and adversarial loss, so the objective at each level takes the form of:

$$\min_{G_n} \max_{D_n} L_{adv}(G_n, D_n) + \alpha L_{rec}(G_n)$$

The adversarial loss L_{adv} penalizes the distance between the distribution of patches in x_n to the distribution of patches in $\widetilde{x_n}$. Specifically, when an image is generated, a Markovian discriminator D_n discriminates between overlapping patches in the image, i.e. classifies them being real or not. The final discriminator score is the average over the patches score.

The reconstruction loss Lrec is a bit different. This loss enforces the generator to generate images that are similar, in some sense, to the original image. This can be done by ensuring that it exists an input noise map that generates the original image at each level x_n. Specifically, for $n<N$ the generator needs to generate x_n if $z_n=0$, and for $n=N$, it needs to generate x_N for a specific noise map z^*. Mathematically speaking, when $n<N$ the penalization has the form of:

$$L_{rec} = ||G_n(o, (\tilde{x}_{n+1}) \uparrow^r) - x_n||^2$$

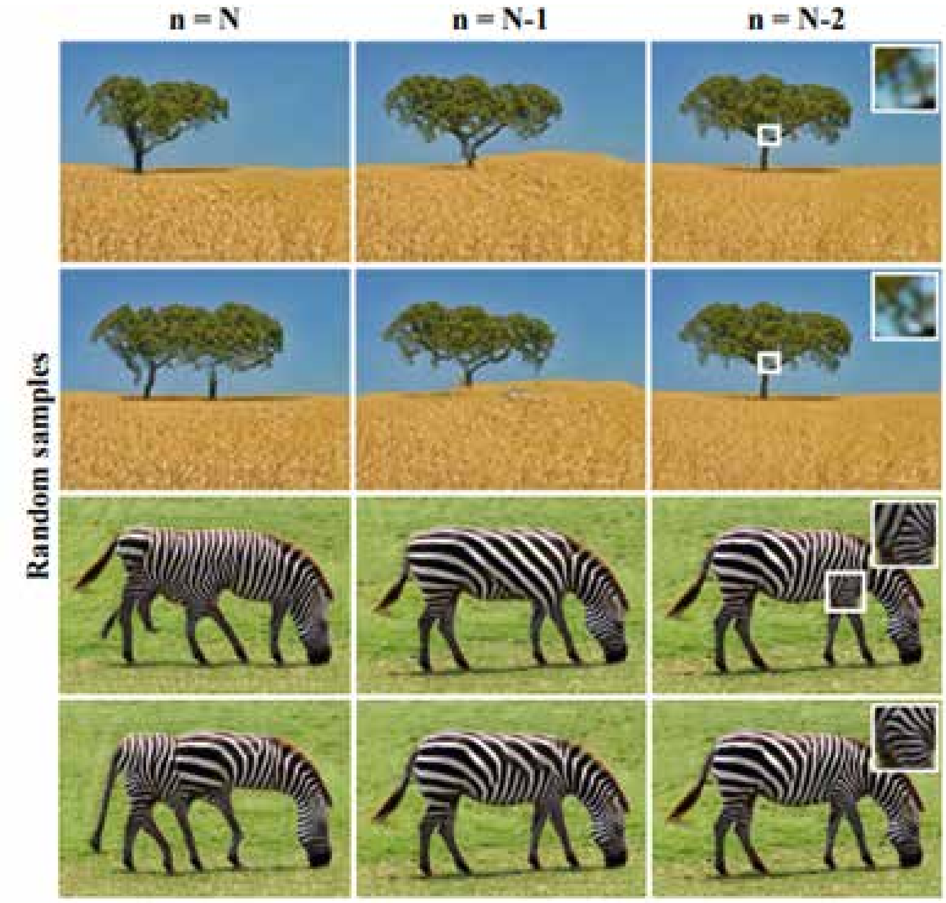
And for $n=N$ it is simply

$$L_{rec} = ||G_N(z^*) - x_N||^2$$

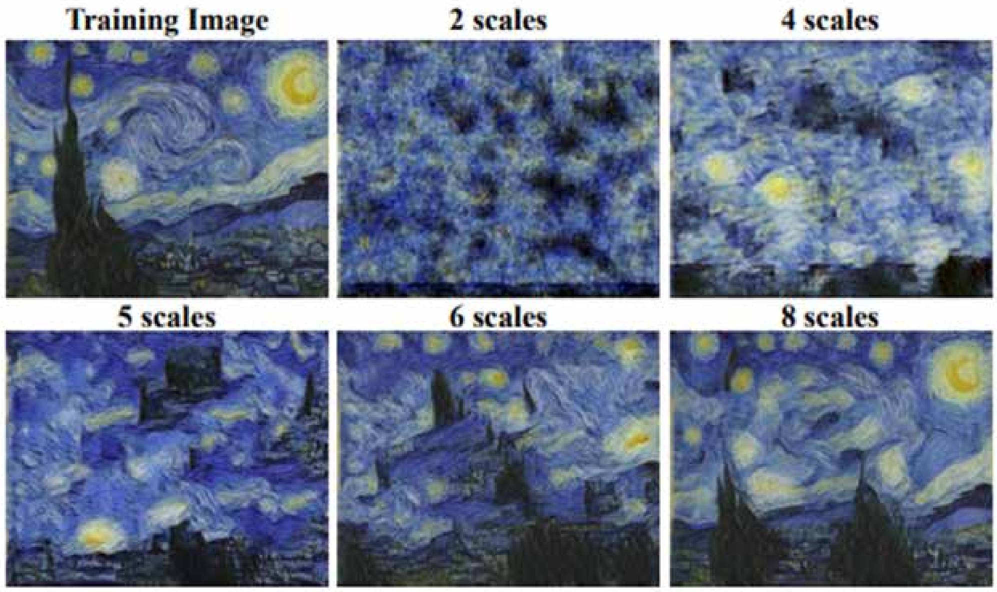
Where z^* is a fixed noise map.

Results

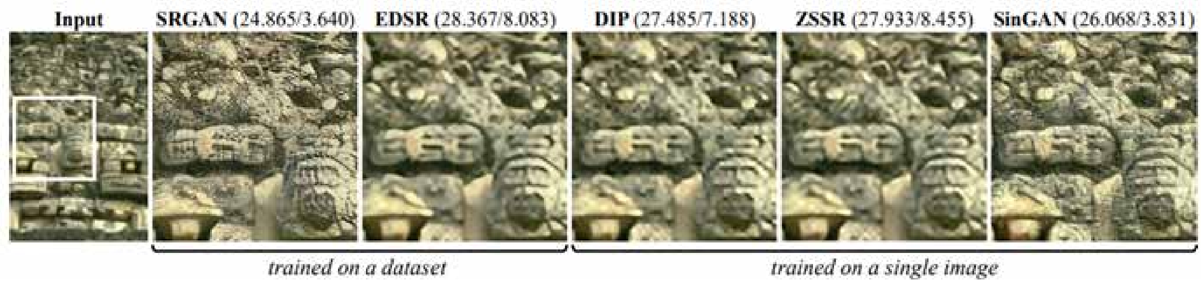
The paper shows high quality results in generating random images from a single image. The authors also explore their model on several computer vision applications such as super resolution, paint to image, harmonization and more. We start by showing an empirical study presented in the paper, about the different components of the model. Below we can see the generation of images in different scales, captured at inference time. The figure shows the coarsest level, and its consecutive two up-sampled levels. It can be seen how the model refines the solution and, as it goes to the upper levels, the generated images become more realistic. Moreover, the shape and pose of the object are preserved while the texture is improved at each level.



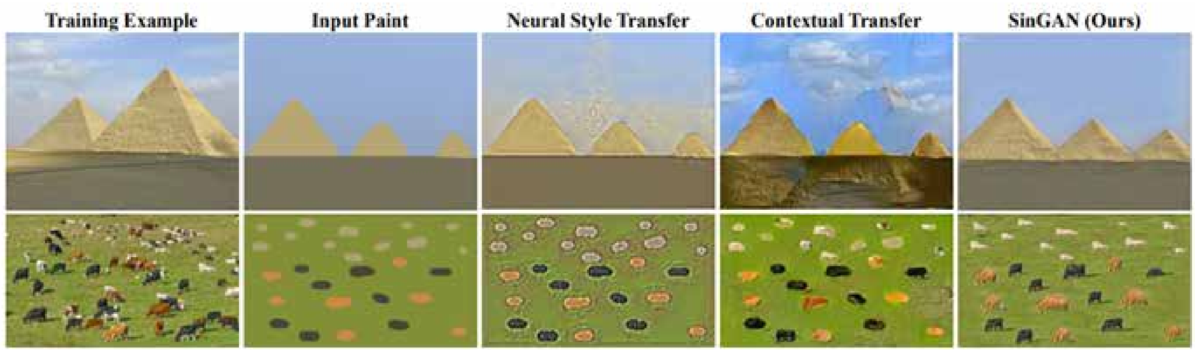
Another nice demonstration in the paper is the effect of training with a different number of scales. It can be seen that the number of scales in the SinGAM strongly influences the results. A model with a small number of scales only captures texture, and as the number of scales increases, the model manages to capture larger structures as well as the global arrangement of objects in the scene.



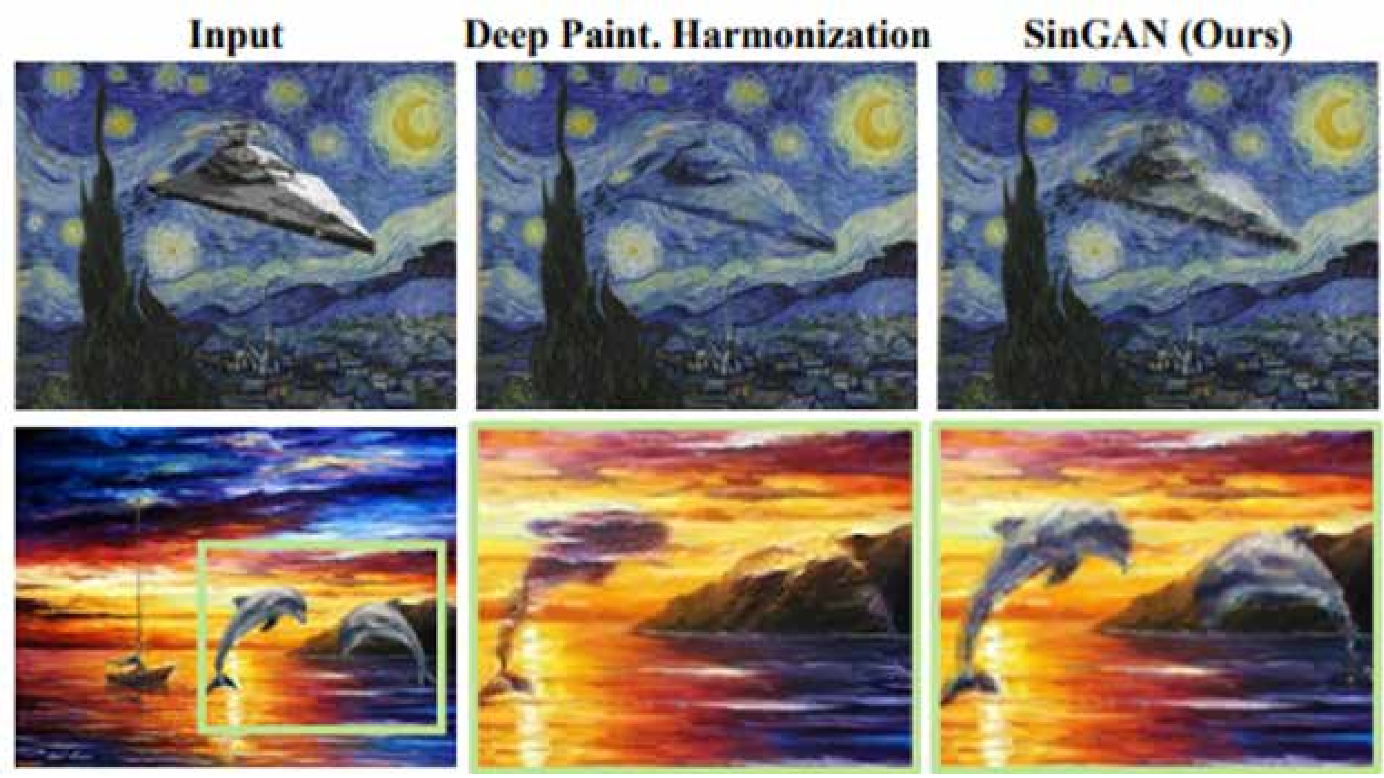
We now dive into the different applications of the framework. SinGAN model is also able to perform super resolution from a single image. This can be done by iteratively upsampling the image and feeding it to the SinGAN's finest scale generator. The authors compared their results to five states of the art super resolution methods and demonstrated comparable results to SRGAN.



Paint-to-image is another cool application of SinGAN. This is done by downsampling the clipart image and feeding it into one of the coarsest levels of the model. As can be seen below, this is a task that is well suited to this model. The SinGAN outperforms two other methods and generates an image that preserves the layout and general structure with realistic texture and fine details that match the training image.



The last application we show is the harmonization. This is the task of realistically blend a pasted object with the background image. To perform this task, the authors suggest to train the model on the background image, and then feeding the naively pasted object image into the coarsest level of the image. The results can be seen below (compared to Deep Paint Harmonization).

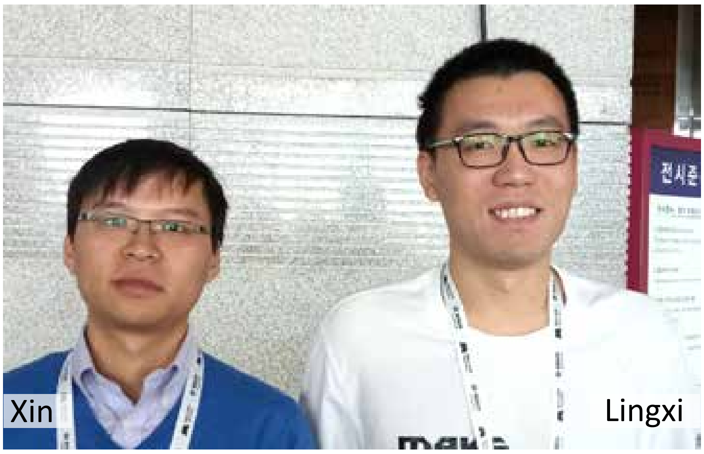


Conclusion

SinGAN is a novel framework which is able to generate images from a single image training. It uses a pyramid of generators that begins by generating an image from a random noise and ends in generating globally consistent image. The model learns the image's patch statistics across multiple scales, using the multiscale scheme. The authors demonstrate the applicability of the model to several computer vision tasks and show remarkable results. For this novel approach, the paper won the best paper prize at the latest ICCV. For more results and information we highly recommend you to read the paper.

Progressive Differentiable Architecture Search: Bridging the Depth Gap Between Search and Evaluation

Xin Chen is a PhD candidate from Tongji University and Lingxi Xie is a senior researcher at Huawei Technologies. They spoke to us ahead of their oral and poster presentation.



Xin and Lingxi propose an efficient and progressive method for **differentiable neural architecture search**. This method has achieved state-of-the-art performance on a benchmark dataset.

Lingxi tells us that there are two main challenges in neural architecture search. The first one is how to make it fast. Conventional approaches involve using reinforcement learning or genetic algorithms which need to train a basic network again and again. This process is quite slow. Now, there is a new technique named differentiable architecture search that can make this process faster, but it raises another challenge which is stability. Currently, when you search the network several times, you can get a different architecture every time and these architectures can produce quite different results and levels of accuracy.

He explains how their work helps to solve this: *"Our paper involves two techniques. The first one is that we propose a gap between the search stage and the evaluation stage. When conventional approaches search on the network, the search stage often involves a shallow network, but the evaluation stage involves a deep network, so there is a gap between them. We use a progressive way of searching architectures that allows us to gradually increase the depth of the search network so that we can close this gap and achieve better results. The second technique involves improving the stability of search. We add some regularization after the skip-connect operation so that the operation can gradually gain the weight rather than they grow up very quickly to surpass other operators. We also make it so that the network is able to preserve a fixed number of skip-connections after the entire search so that the final network will become much more stable."*

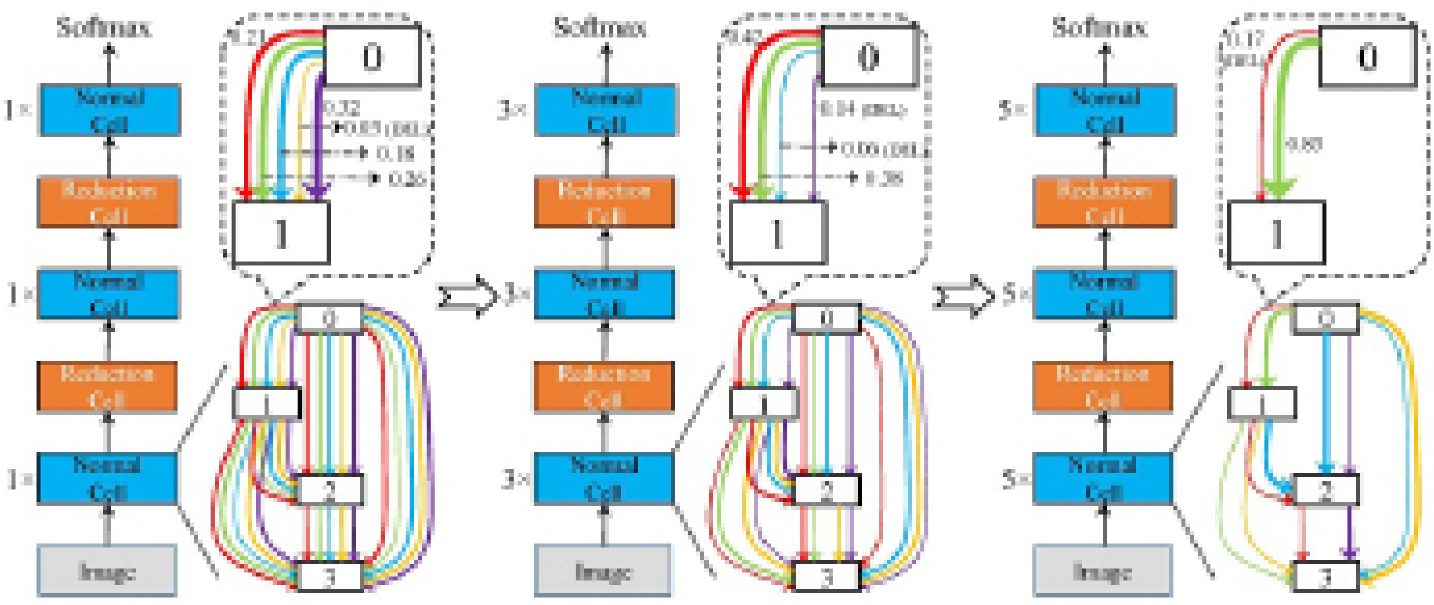
Lingxi is sure that neural architecture search or a bigger perspective of the problem called AutoML will be the future of artificial intelligence, computer vision and machine learning. However, he says there are still some challenges to be solved before AutoML can be applied to a wide range of applications.

As well as speed and stability, the scale of architecture search – meaning that the architecture can be searched in an even bigger space – is an issue. Currently, many problems limit the architecture search in a relatively small space. This is not ideal but has to be done because when this constraint is freed, most architecture search methods become either too slow or less stable. Also, many people apply hardware constraints because the search network is not so friendly to hardware.

Another challenge is the difference between the search stage and evaluation stage which means that they need to be two separate stages. In the future, Lingxi hopes that they can design a method that can search directly, and after finishing the search stage does not need to be retrained so that the model can be applied directly to real-world problems. Exploring solutions to these challenges will be future steps for this work.

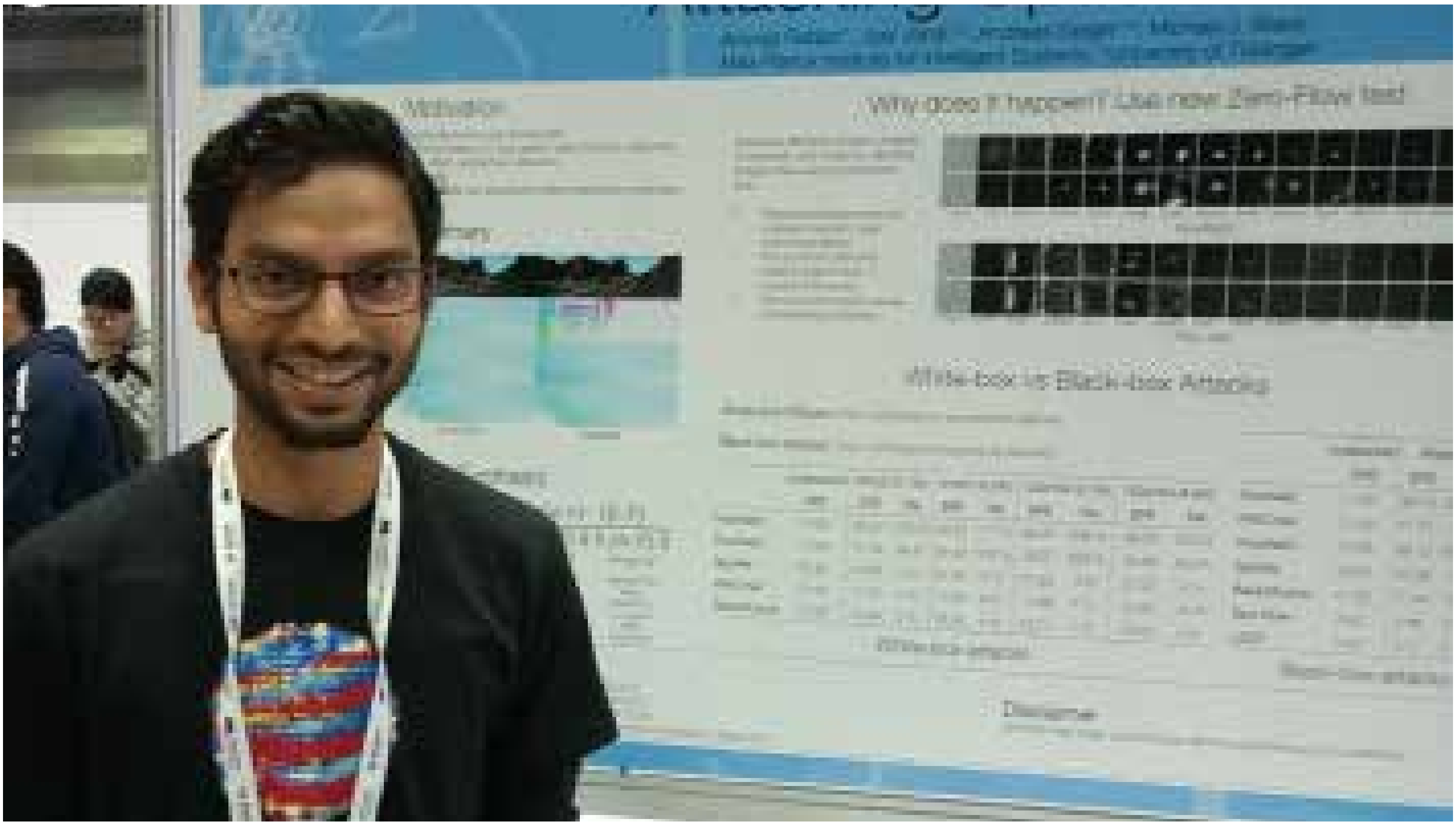
How can this work be practically applied in the real world? Xin tells us: *"One of the applications is on the image classification task. Just to search a network for the classification model and when you get a picture, the model can immediately tell us what the object in the picture is."*

In conclusion, Lingxi says neural architecture search and AutoML are very promising research subfields and will have a significant impact on artificial intelligence, computer vision and other fields such as computational linguistics. It is still in a preliminary stage and much work remains to be done, but he hopes that their work will inspire future researchers in the field.



(a) Initial Stage (b) Intermediate Stages (c) Final Stage

Attacking Optical Flow



Anurag Ranjan is a PhD student at Max Planck Institute for Intelligent Systems with Honorary Professor Michael Black. He spoke to us following his poster session.

The work is studying the **vulnerabilities of optical flow networks** and comparing them with classical methods and other vulnerable networks. Anurag says that in this case, they synthesized a small patch attack which was less than 1% of the image. When putting this patch in the scene and computing the optical flow using these networks, the optical flow failed miserably. The errors were about 400 per cent!

These optical flow networks have activations when they're not supposed to, with and without these attacks. This work investigates the problems with these neural networks to learn how to fix them.

"This work investigates the problems with these neural networks to learn how to fix them."

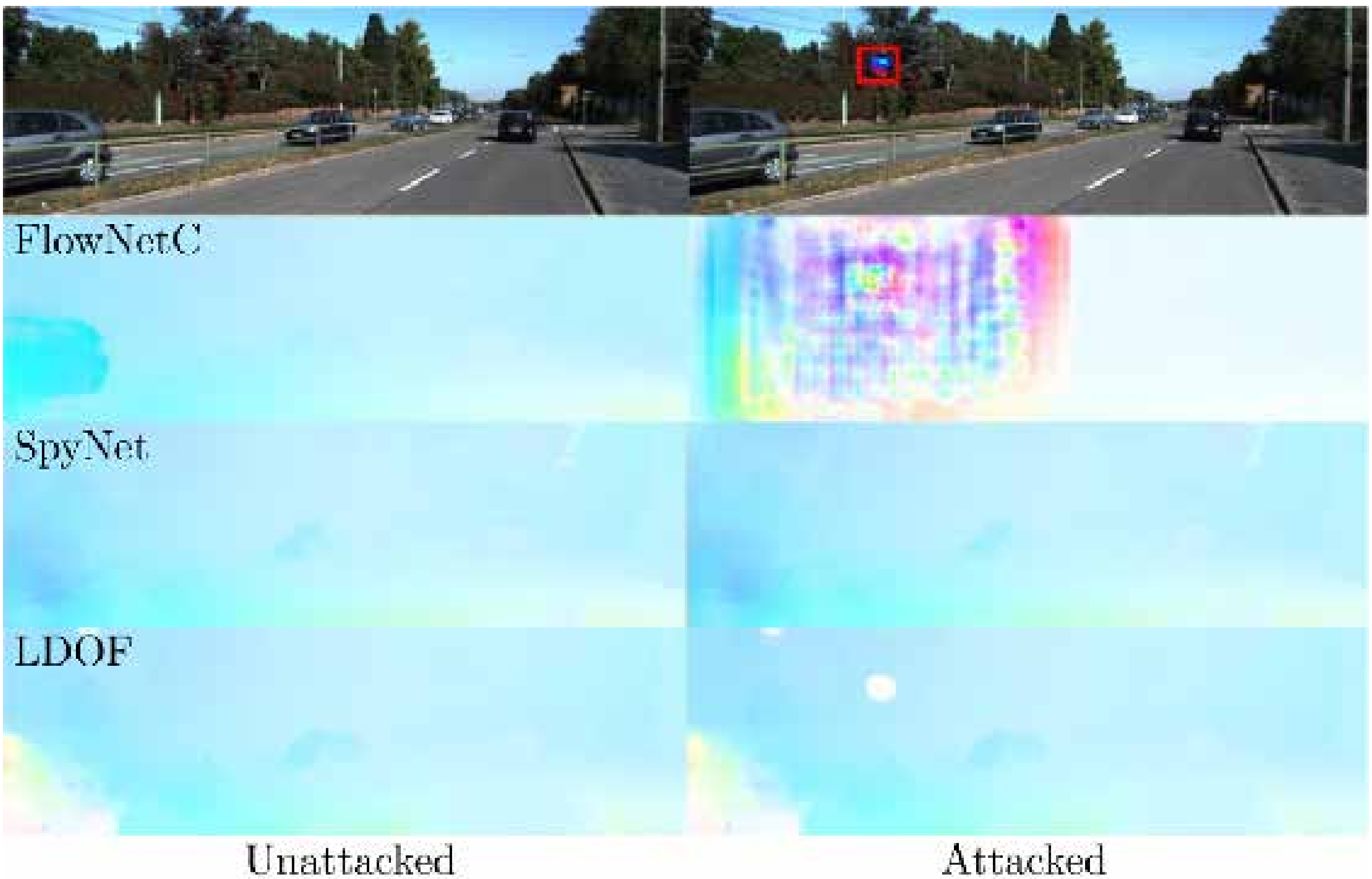
He says what's really surprising is that the optical flow predictions from the network are a lot worse than he thought: *"One of the biggest research questions is why this happens. We still don't know exactly, but we proposed some heuristics of what might* *be happening. One of the things is that the future activations in these networks are not spatially invariant even without the attack. You see some activations even when the activations should be zero. The other problems are that in these pyramid networks you see that they predict flow even when there is no motion and even when there is no attack. What happens during the attack is these factors get amplified and the predictions become really random."*

"but we proposed some heuristics of what might be happening."

In the real world, Anurag tells us this amounts to vulnerabilities in several applications where optical flow is used. For example, **self-driving cars use optical flow to compute motion**. You could have a patch on the street and the optical flow predictions **could be really bad**. In that case, the decision taken by that self-driving car would be incorrect.

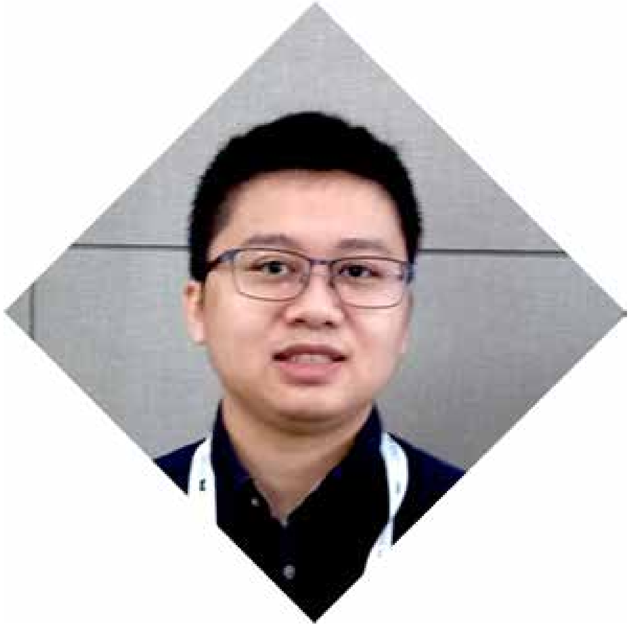
He is modest when we suggest that this could be life-saving work: *"It could be. It's promising, yes."*

Anurag says their work has given them several ideas about what these attacks can do. In future, they will be exploring other ways to attack these systems to reveal problems with the networks. They can then **propose defenses to make the systems more robust** so that they can be deployed safely in the real world.



Seeing Motion in the Dark

Qifeng Chen is an assistant professor at HKUST. He spoke to us ahead of his oral which he has presented on behalf of main author, Chen Chen, who was unfortunately unable to attend due to visa issues.



The work is about generating high-quality video in a low-light environment. Their model is trained on low-light raw sensor data from static environments but can be well generalized to dynamic environments where some objects might be moving. Qifeng says that the model produces much better video than popular cameras like Sony DSR and those on iPhones.

The key idea in their approach is they start with the raw video sensor data. The raw sensor data is not the RGB images usually used in computer vision tasks; it is what the camera actually sees. The camera will produce the RGB images or videos.

"The raw sensor data is not the RGB images usually used in computer vision tasks; it is what the camera actually sees."

"In the real world there are multiple applications for night-time computer vision tasks."

Qifeng explains: *"We start with the raw video data and then train a convolutional neural network to reconstruct a high-quality video out of it. We created a dataset specifically for this task with ground truth and our model can be generalized to generate a nice video. The algorithmic part of this project is that we are proposing a new type of loss. The loss we use in the model is we randomly pick two frames and we want to make sure these two frames will be consistent. We also make sure all the frames will be consistent with the ground truth, so we can actually make multi-consistency in our generated video."*

Qifeng is proud to be pushing the boundaries of low-light video. He tells us that in the real world there are multiple applications for night-time computer vision tasks. For example, autonomous driving at night so that we can see better, and for surveillance cameras to produce better video out of the sensor data.

Beyond Cartesian Representations for Local Descriptors

Kwang Moo Yi is an assistant professor at the University of Victoria in Canada, having made the leap from developer to faculty. He was previously a postdoc researcher at École Polytechnique Fédérale de Lausanne (EPFL). He spoke to us following his poster.

The work is about learning local features. Trying to figure out where you will be able to align two images together or find the camera pose difference between images. The state-of-the-art way to do this is by using **SIFT features** developed 20 years ago by **David Lowe**. This work has tried to do a learned counterpart for that. It's all about the last part of the pipeline which is to describe the points that are selected in a robust way.

While at EPFL, Kwang Moo worked on learning local features, resulting in the LIFT paper. A downside they found was that the detectors were making a lot of mistakes in the **scale and orientation estimation part**, meaning how you would look at these local patches.

What does this work do differently? He explains: *"This work is trying to use a different representation for looking at these local patches and to have them so that they would be represented in something called the log-polar representation which turns images into something looking like circles. When you rotate or scale the images it becomes translation in this representation, and we all know deep networks are very good at learning to compensate for translational errors."*

"The most challenging part is getting the benchmark right."

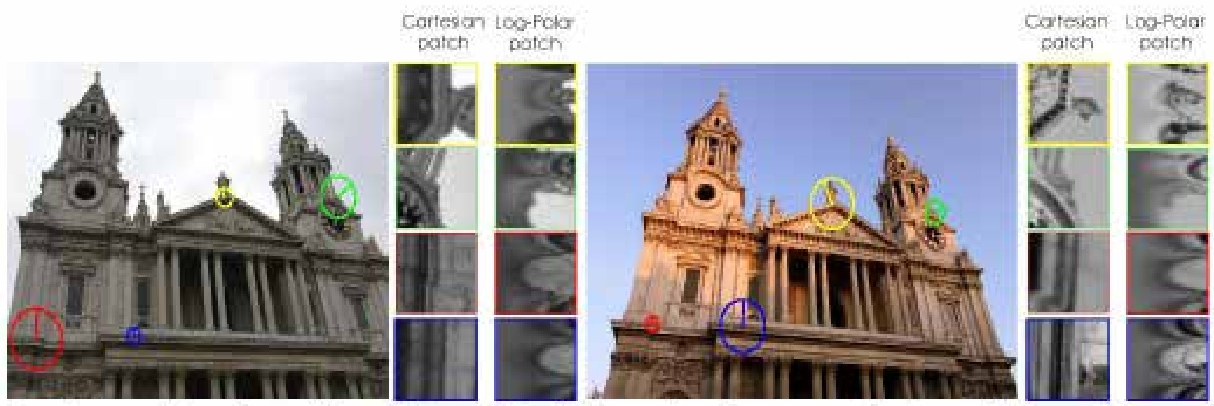
Kwang Moo tells us the most challenging part is getting the benchmark right. A lot of the baselines have their own default settings which might not be the best for a given dataset. They had to do a lot of under-the-hood work to sort that out which took a long time. From an algorithmic point of view, he says solving the challenge is simple because the core thing is just representing images as local images and the rest of the pipeline is the same as other existing methods. This is really a case of a simple idea, well executed, that does make a difference.

He adds that he is proud that they have revived something that was a direction that people were trying to take in the past. He tends to do a lot of things drawing inspiration from past works and believes log polar is one of those. There's also something called Daisy that was developed by CVLAB at EPFL. It's similar but there are some differences: Daisy used to create statistical features from normal typical images and then did the aggregation in log-polar domain, whereas in this case, they do it from the beginning because deep networks allow this.

Kwang Moo is not actually the main author of the paper. That is **Patrick Ebel**, who has moved on to a different institution now. Also, **Anastasiia Mishchuk**, the second main author, is unavailable to talk to us today, so it falls on the third guy! He modestly tells us that they deserve more credit than him, but we can attest to the fact that the poster session had a full house on Tuesday, so it all worked out well in the end.

He points out: *"One of the things that was really good for our poster was that we didn't have to take long to explain because* ***it's a single idea****. It's really nice that we're going to continue investigating this because it fits perfectly into what we've been doing for three or four years now: learning local features. This will be an opener for another chapter for this."*

Finally, Kwang Moo says the principal supervisor of the work, **Eduard Trulls**, has been a great partner for him in this journey of local features. Together with **Pascal Fua**, they are determined to continue working on this, even if they are all in different places.



Two images taken from different viewpoints with four pairs of corresponding keypoints.

Learning Trajectory Dependencies for Human Motion Prediction

Wei Mao (left) with supervisor Miaomiao Liu.

Wei Mao is a first year PhD student at Australian National University (ANU) under the supervision of Miaomiao Liu. This is a joint work with Mathieu Salzmann from EPFL and Hongdong Li from ANU. He spoke to us ahead of his oral and poster.

The work is about **human motion prediction**, which means generating where a human is going to move in the future after you have seen their movement in the past. Human motion prediction is key to the success of applications such as self-driving cars, human-robot interaction and social robots.

Wei asks us to imagine that we want a **robot to collaborate with a human worker**. If the robot can't predict the human movement, then the direction of the robot will always be delayed compared to the action of the worker. To collaborate with the worker seamlessly, the human robot needs to prepare for the future movement of that worker.

One key insight of the work is that the past trajectory of human joints is closely related to the future trajectory. That's why Wei says they first encode the past trajectory trying to find a relation between the past sequence and the future sequence. Also, the movement of one joint is closely related to that of other joints, so they leverage the graph convolutional networks to capture the joint trajectory dependencies.

The challenge is that **humans all move quite differently from each other** and predicting the future is difficult, but Wei says their method is simpler, smaller, faster and more effective than previous work:

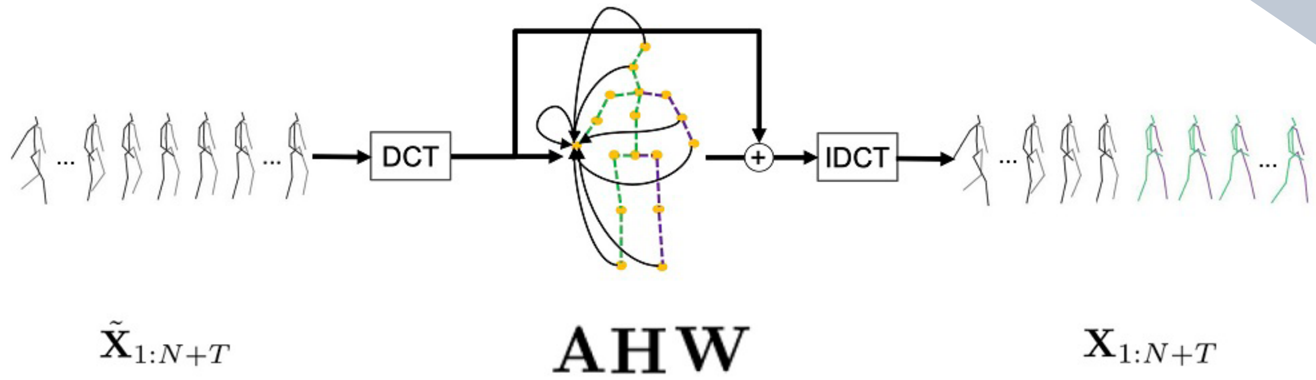
"Previous work used recurrent neural networks (RNNs). Using RNNs is a natural choice for handling sequence data, but there are a few drawbacks. The past information will be forgotten after a long propagation because you

need to pass the latent variable to the next step and then do it again and again. After several steps, the past information may vanish. Our method is a feed-forward method so that we don't have that problem. Also, we're trying to capture the dependencies between different joints using a fully connected network which means every joint is correlated to all the other joints, and that's different from previous works which use a fully connected layer or convolutional layer to capture the dependencies between different joints. If you use a convolutional layer, then the dependencies will depend on the kernel size. Also, it depends on how you arrange your data. Our method is more flexible because it captures the dependencies between all the other joints."

Their work so far has **been looking at predictions based on the human skeleton and stick-based human representation**, but thinking about next steps, Wei points out that humans are 3D. We are tall, fat and skinny, and we interact with the world with our skin, not our bones. That's why in the future they will be trying to estimate human shape as well. In conclusion, Miaomiao tells us why she thinks this work is crucial for **self-driving cars**:

"If we want cars to drive autonomously on the road, they need to understand not only the behaviour of other cars, but also the behaviour of humans."

*"If we want cars to **drive autonomously on the road**, they need to understand not only the behaviour of other cars, but also the behaviour of humans. We need to predict what their future actions might be. For example, whether the human will cross the road or not. **Colleagues working in the autonomous driving industry have confirmed that this is the next step that they will be working on**, so it's an important and interesting project for students."*



Marc Pollefeys

Marc, thank you for agreeing to talk to us. I have wanted to have a chat with you for a long time! We have many interesting topics to cover, including your work at Microsoft, HoloLens and ETH Zurich. Last but not least, you are one of the General Chairs here at ICCV 2019. Where shall we start?

Let's start here. I'm one of the General Chairs but, of course, the person that really made this happen is Kyoung Mu Lee who is the local General Chair. I know because when I was the local chair in Zurich for ECCV 2014 that's where I did most of the work. Here, clearly Kyoung Mu has done an amazing job. This is a fantastic facility and I have heard only good things from people who have really enjoyed the conference, including myself!

He has been able to scale by a very big order of magnitude.

Yes, when we bid for this four years ago, we were expecting maybe 3,000 or 4,000 people. That's what we planned for. He was able to keep scaling this to 7,500.

With no visible glitches.

Exactly. Between Kyoung Mu and the local team, they've been able to smoothly scale things and it has worked really nicely. I like the format. I think the awards session at the beginning was very nice. I liked the longer talks in that session. It is sad that we can't do that for all the orals anymore. There are advantages and disadvantages to having our community grow this fast. How will this evolve in the future? Will it slow down? We don't know. It's a very interesting time in the computer vision community right now.

With 4,000 papers submitted, it's certainly a huge task.

It scales with the size of the community, but it does bring some challenges. This was our program chairs' responsibility and I think they did a fantastic job.

It's impossible to ask you about ICCV 2019 without mentioning the fact that 3,000 participants are Koreans. Do you have any reflections about what the local community has been able to do and how they have grown they have

grown in the last few years?

There are many strong local communities in many countries which we don't always see enough of internationally, but that's an aspect that I really like about ICCV. The local community in the host country typically rises to the challenge. If we do ICCV in China at some point, I'm sure there will be a huge local presence too. In any country you go, you see the leaders of the local community really investing in organizing the best possible event.

Having planned such a successful event this year, do you have any recommendations for the organizers of the next ICCV?

Foresee enough space!

What are you the proudest of about this ICCV?

Over four years, attendance has grown by a factor of five and as an organization we've been able to cope with that. I think it's fantastic that the event and the community have not been disturbed by it.

Let's move on to HoloLens and Microsoft now. You are probably the most popular figure identified with this Microsoft product and have invested a lot of yourself into the challenge. How did you come to bet so much of your time and career behind the project?

On the computer vision side of things, I care a lot about HoloLens and invest a lot of effort in that. The main person is of course Alex Kipman, who is the person leading the HoloLens effort. He started that many years ago before I got involved.

So, his bet was even bigger?

Yes, he literally bet his whole career on it! Within Microsoft, there are many other people very much invested. Jamie Shotton in Cambridge is playing an important role in developing the computer vision capabilities for HoloLens. When Alex approached me in 2015, it took me a while to decide to dive in, but I realised that having been in academia for 20 years and having done a lot of work to try and make algorithms practical, that this really was an opportunity to do that for real. Also, at CVPR 2015, I noticed that the industry presence suddenly grew tremendously, so it felt like the right time to be thinking about exploring opportunities to contribute in a meaningful way. The obvious possibilities were in self-driving cars, augmented reality, mixed reality, robotics and so on. Looking at all of these opportunities, I decided to go for HoloLens and mixed reality. It will be a long journey, but HoloLens is already at a place where it can really make a difference. It's not a device that everybody is wearing out in public today, but this is of course the ambition and what we are working towards.

The current form factor can already do a lot of very useful things. In particular, in industrial scenarios such as in a factory where a worker is repairing a machine, or a surgeon working with their hands to do surgical procedures on someone. So, where people have to do complicated tasks, there exists relevant digital information that can help them do their task better, more efficiently or more safely. If that information can be brought in front of their eyes in an easy way so that it's natural for them to absorb it while they do the task, that can make a huge difference. That's what we're working towards and what I'm really excited about.

To do that, there's a ton of computer vision problems that need to be solved. You want the device bringing information naturally to the person aligned to the real-world problem in front of them. That requires a hologram that visually looks like it's rigid and attached to the world. It requires you to continually know very precisely, and even actually predict slightly ahead of time, where the device is. Just to show a static hologram in the world it needs SLAM (simultaneous localization and mapping) to run all the time, and

really high-quality SLAM, which is a combination of visual-inertial odometry and then larger mapping capabilities. All of this has to run in real time all the time while you use the device. In addition to that, just to place holograms in a way that they interact with the world – for example, placing a hologram that can actually sit on a table instead of somehow being half pushed through it when I'm placing it – requires a 3D model of the world at the same time. We use our depth camera for that, which is the same camera that is now in the Azure Kinect.

"ETH is indeed a fantastic place. "

There's also a need to be able to track what the user is doing. In HoloLens, there's computer vision that will continuously track the hands in real time. That's something that's done mostly in the Cambridge lab. There's also eye tracking. First, we look at the eyes to do high-quality biometric authentication by having cameras read your iris. At the same time, those cameras can also track your eyes. The most important direct outcome of that is that we can calibrate the display to where your eyes are and optimise the viewing quality for the position of your eyes. A practical use for this is an auto-scrolling feature. As you read text, it naturally keeps scrolling, following your gaze. Beyond that, it can also be used in combination with hands as an input modality of where you're looking. There are many things like this built-in to HoloLens. As we go forward, we want to go even further. We want HoloLens to be able to understand what the user is doing in the scene at the semantic level. So, not only track the hands in detail, but go beyond that and recognise the scene and objects that are in front of them. Those are all really important things to make the device even better suited to assist the user in solving whatever problem or task they're trying to address.

"Think of it like Post-its that you can attach to the world."

In Zurich, one of the key things my team focuses on is being able to share experiences across multiple users and devices. Both at the same time, but also so that I can place information in the world and then later I can either retrieve it myself or have other people with their devices retrieve it. In the same way that computers became a lot more useful once they were networked and once I could place something on the internet that somebody else could read - think Wikipedia - this is the same thing but for annotating and placing information in the context of the real world. That becomes a lot more interesting when I place information somewhere in the world and I can have another person access that information who is entitled to do so. That requires that my devices always build up a 3D SLAM map of the environment that I traverse, and I need to be able to share that map with other devices. The natural way to do that is to build up this map in the cloud. If I map part of the space and then somebody else walks through the space with their device and so on, all of us together can share and build it up. Essentially each of those is a small puzzle piece describing the world. If I assemble all those puzzle pieces, I can build and maintain an up-to-date map of the world that allows everybody to share information that's aligned and attached to the world. That's really the ambition.

The first step in doing that is a mixed reality service called Azure Spatial Anchors. That's about being able to attach things to the real world by uploading a spatial map of an environment to the cloud and having other devices able to align to it by uploading some information from their device to the cloud. The cloud can compute the relative alignment between the coordinate systems so that my understanding of space can be shared with other devices and their coordinate system can be aligned to mine. Therefore, wherever I place information they can retrieve it in the context of their representation of the world. That's currently in public preview.

The team in Zurich have been very much focused on contributing some of the key computer vision work to enable that. It's a service that wilbe one of the fundamental services under Minecraft Earth which is a new augmented reality game currently in early access release. In it, you can have multiple people play together in the same spatial context. So, I would see some Minecraft elements somewhere in the world, and someone else that is in the same experience as me can see that in the same place in the world on their device. This service works on iOS, Android, as well as HoloLens, so it's cross-platform. Azure Spatial Anchors enables real world sharing of these holograms and experiences in games, in the same way that HoloLens and phones allow people to communicate and share information. Think of it like Post-its that you can attach to the world.

You are clearly very passionate about this, and the community is very passionate about HoloLens. I think we can safely say that your gamble succeeded! Before we move on, do you have a funny story or something that we wouldn't know that you could tell us about HoloLens?

Before I joined the project actually, as they were ramping up production of the first HoloLens device, one of the machines in the production process had problems and the experts needed to help were far away in Redmond. The person in the factory was trying to fix it and was on the phone with people in Redmond but it was difficult. They suddenly thought: 'Hey, we have a few HoloLens's here, I'll put one on and then you can see what the problem is through Skype.' This enabled them to solve the problem quickly without needing anyone to travel. This is actually now a built-in feature on HoloLens 2 called Remote Assist which can be deployed right away for enterprises! It is a tool available now in Teams, which is our modern communication platform for enterprises. The remote assistance capability can be integrated directly in your company communication platform so that people out in the field or in the factory can contact the experts to help them resolve a problem.

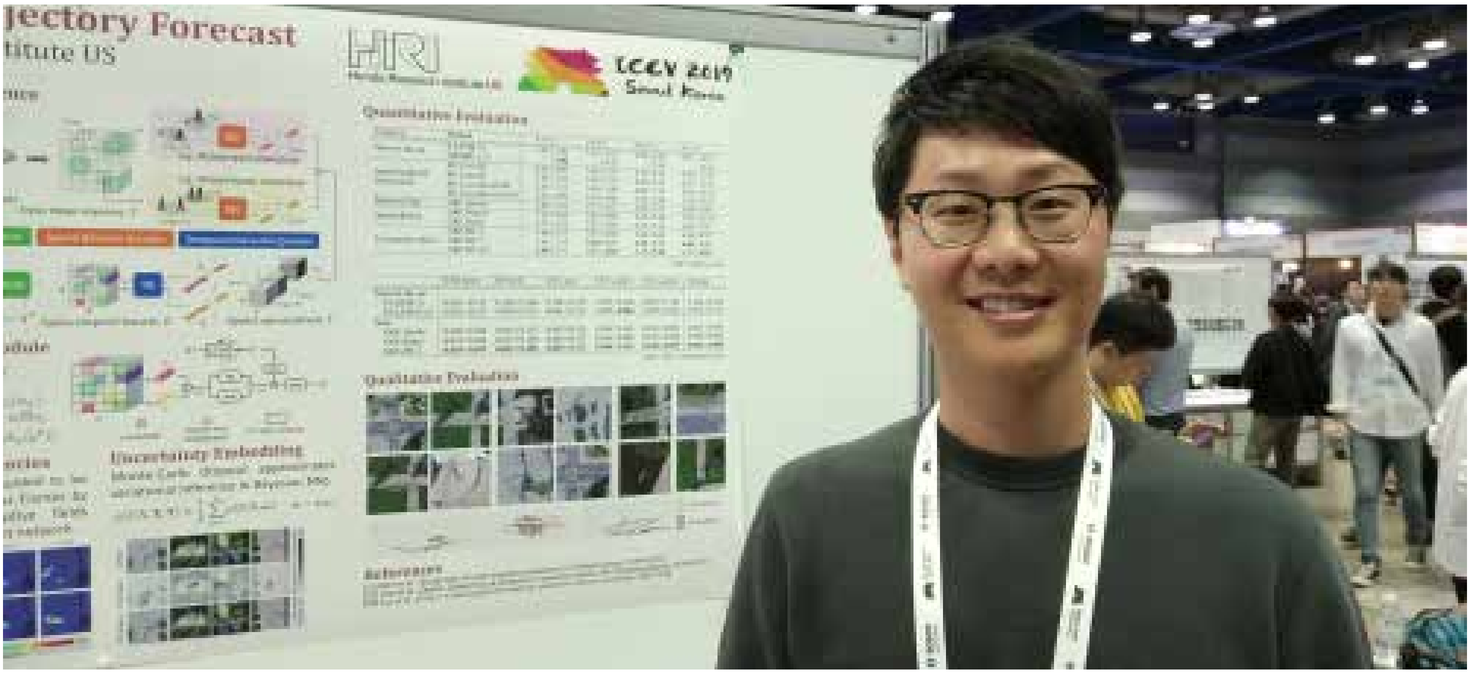
Marc, this has been a really fascinating interview. I knew that we would run out of time to talk in detail about ETH Zurich. It's such an important institution that it deserves an interview all of its own. Briefly though, what can you tell us about your work there?

ETH is indeed a fantastic place. Outside of the Anglo-Saxon world, it's systematically the best university, and in computer science is doing fantastically well. We have an amazing department. In the Shanghai Ranking for computer science we are No.5 after CMU, Stanford, MIT and Berkeley. In the Times Higher Education rankings, we are currently No.2 in the world. In large part, this is down to the sizeable resources that we get for every professor to invest, and we have a lot of freedom in research. That freedom means you can get on and start new things on your own initiative; exploring new areas without first having to prove yourself. Also, it's easy to attract top people at all levels as ETH is such a great place and Zurich is such a great city with a great quality of life.

"In the Times Higher Education rankings, we are currently No.2 in the world."

We have excellent students, but we also bring in excellent PhD students, Master's students, postdocs, and of course faculty. Combined with healthy resources, it is an amazing mix. When I joined Microsoft, I told Alex that I would only accept if he understood that after two years I would go back to Zurich. The only thing that changed is that I decided to stay 50/50 between ETH and Microsoft. I saw that I could do something that would be a win for ETH and a win for Microsoft. On the one hand, there's no better place than ETH to be able to explore these longer-term challenging topics and more open research problems. On the other hand, a company like Microsoft has resources, has a lot of very interesting problems, has interesting devices and capabilities, and bringing those two together is a win-win. The students get super interesting problems to work on that are inspired by the challenges that we see to realise the vision of HoloLens and mixed reality. Having a Microsoft lab that's closely associated with ETH, where we have ETH PhD students working with us, puts me in a unique position and brings a lot of additional value to both of my roles.

Looking to Relations for Future Trajectory Forecast



Chiho Choi is a scientist at Honda Research Institute USA. He spoke to us following his poster session.

Chiho's work is about a relation-aware framework for future trajectory forecast which uses a set of images to extract relational behavior between road users and their surrounding environments. It uses images as input and tries to model both human-human interactions together with human-space interactions, as well as relational behavior between agents.

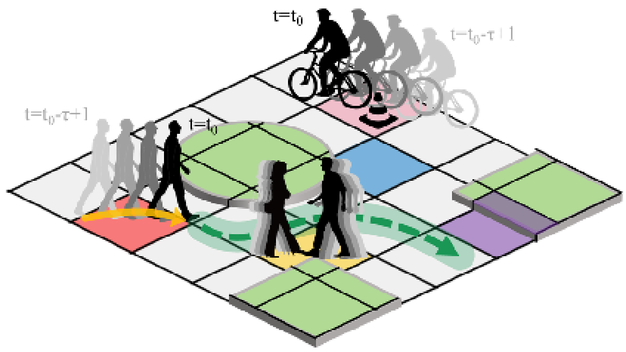
In this domain, they are the first to use those relational behaviors for future trajectory forecast, but the real novelty is the specific relation gate module in their work.

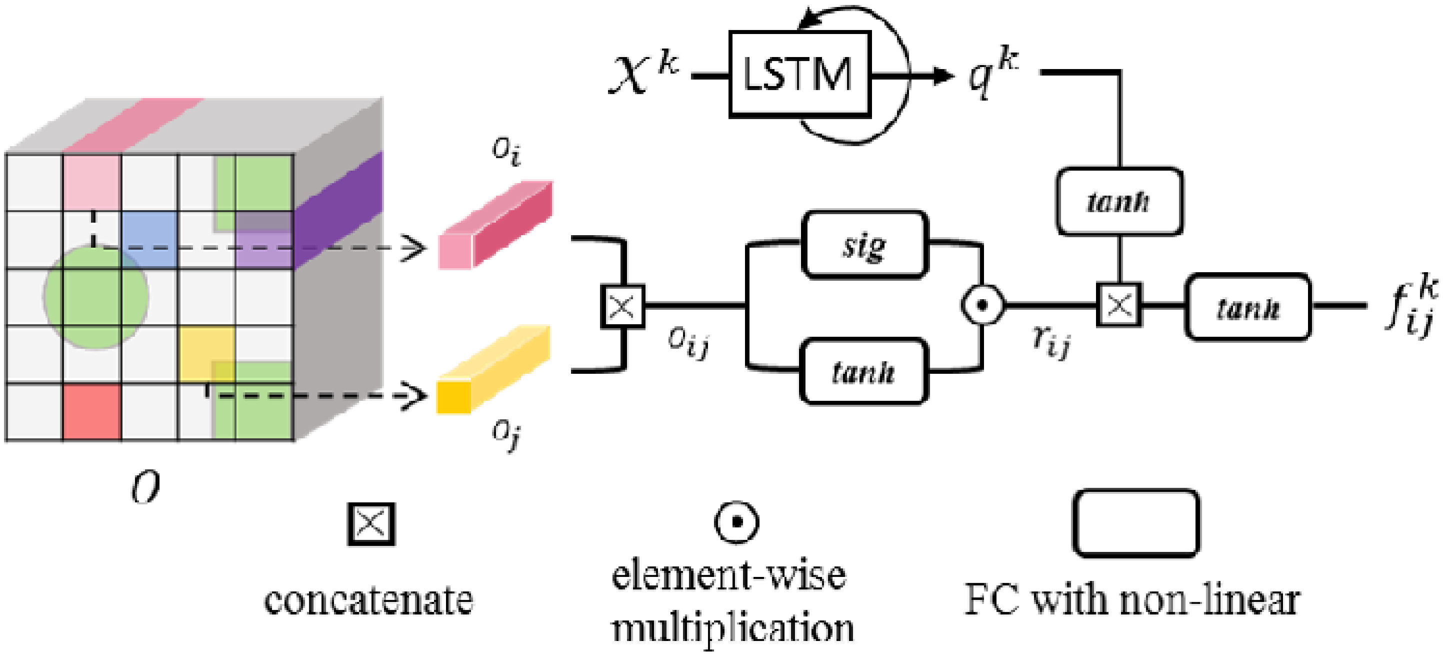
They are predicting the future, which is usually unpredictable, so there are lots of challenges to consider, including solve this by applying an uncertainty modeling method.

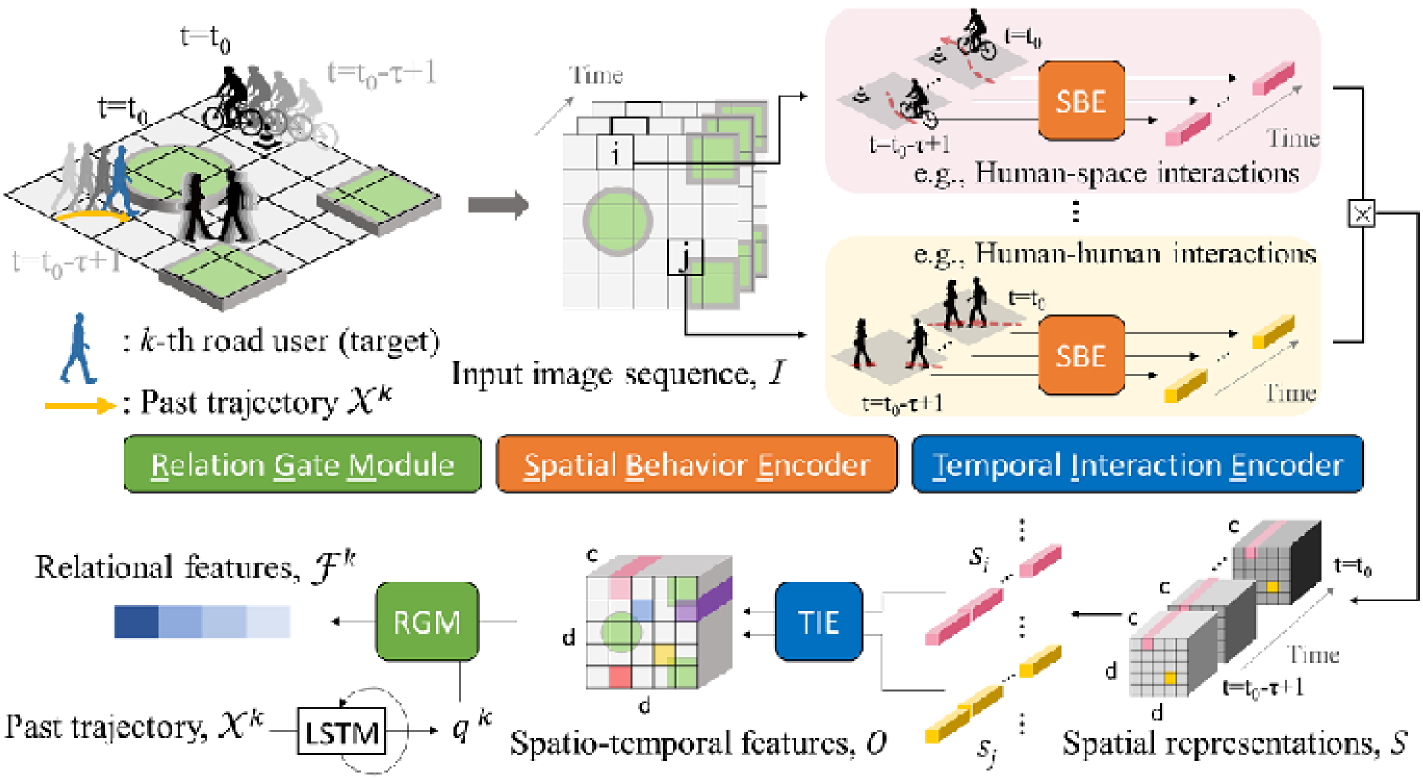
Chiho explains: *"In the literature there exists a Monte Carlo dropout method to model uncertainty of the neural network which is applied for semantic segmentation, image classification, and those kinds of things. We extend the use of Monte Carlo dropout for the prediction problem. However, it is very hard to say that our work achieves multimodality over the future prediction problem because even using Monte Carlo dropout, the variation of the predictions is very low. We would extend this work to incorporate actual multimodality to get rid of the uncertainty estimate."*

Continuing to think about what's next,

Chiho says the dataset they are currently vehicle interactions and human-vehicle using focuses on human agents such as **pedestrians and cyclists**, but in the future, Honda want to apply this work to **actual vehicles** to understand interactions together.







If there is no trust, nothing will come”

Georgia Gkioxari is a research scientist at Facebook AI Research (FAIR). She received a PhD in computer science and electrical engineering from the University of California at Berkeley under the supervision of Jitendra Malik in 2016. Her research lies in computer vision with a focus on object and person recognition from static images and videos. Georgia received the Marr Prize at ICCV for "Mask R-CNN" in 2017.

"I love dogs! They don't always love me back!"

Georgia, tell us about your work.

I can start with the 3D stuff that we’ve been working on. We actually had a tutorial on Monday. We presented our work on Mesh R-CNN, which is basically inspired by Mask R-CNN, but trying to move beyond the 2D grid, a grid we are very comfortable in. We want to predict in the 3D, get the shapes of objects, not just the silhouettes of objects. Then we also presented torch3d, a library of 3D operations used in Mesh R-CNN and beyond, including a differentiable renderer. It’s particularly designed to enable and facilitate research in deep learning and 3D understanding. This will all be published very soon. We hope that the community will appreciate the work and build on top of these tools.

You work in a very celebrated team with very special people. That must be challenging and rewarding at the same time.

Definitely - My colleagues have set a very, very high bar, not just for me, but for everyone at this conference. I’m very proud to work with them. When I work on a project, I have to think, *“What will my colleagues think?”* If I know that they will be proud of this work, then I know it’s good work. I care more about their approval than I care about CVPR or ICCV being in my favor.

How do you deal with the pressure?

I don’t think I feel pressure every day because, if you meet my team, they’re super normal. I don’t feel a particular pressure every day, but I think, overall, when I’m finalizing a project or when I’m writing up a project, I think, *“How* *would Kaiming react? How would Kaiming present it? How would Kaiming feel afterwards?"* So I'm actually more inspired rather than feeling pressure.

Also, you have a very strong leadership at Facebook AI that must bring the best out of you.

Correct - The leaders at Facebook AI, not only the leaders, but a lot of the senior members are people in the academic community who are very popular: Jitendra Malik, Kristen Grauman, Abhinav Gupta, Yann LeCun, Rob Fergus, Leon Bottou. All of these people are hugely famous and hugely influential. Interacting with them is another level of inspiration and pressure. To actually have your work read by someone like Yann, it has to be solid.

Wow! Can you tell us a story about what it's like to work with such a special bunch of people?

Before publishing Mask R-CNN, that paper was ready well in advance.

Actually, it was named Best Paper of 2017 by our magazine.

Awesome, thank you! When we were publishing that paper, it was done like a month and a half in advance. I was working on another paper, another submission for that conference, and I remember my colleagues. I was on grad student mode. Before the deadline, I have to work very hard. I have to keep pushing, pushing, pushing. Mask R-CNN was so ready that we had to talk about how to align the figures, like how to make the papers perfectly aligned with the image and the text so that it's pleasing visually. I was like, *"Oh my god! This level of detail for me is like a week before the deadline! I don't sleep! I work super hard to get the paper ready!" They were talking until like 2 AM on how to align figures! [both laugh]*

How do brilliant ideas like this grow in peoples' minds?

I don't think there is a recipe. When people get together in a room, good ideas come out of it. It's something that I've seen, not just back then, but throughout all these years. Better ideas come when people collaborate more, rather than with single author projects.

So trust is a major component in success, trust in your team.

Yes - For me, that's a given. If anyone who is reading this is in a team, and there's no trust, they should take action. That can be either to try to address those problems or get out. If there is no trust, there are no good ideas. Nothing will come out of it.

That's very inspiring. Tell me about you. I know you probably don't like me saying, but how does a young Greek girl become a famous scientist?

Oh boy!

[laughs] **I knew you wouldn't like that! Forget the word *famous*.**

Me being where I am right now is not a step function. It was tiny, tiny steps from like ten years back to today.

"Me being where I am right now is not a step function."

So let's start from the ten years back!

Oh! *[laughs]* That was a long, long time ago!

This is the interesting part. We want to be inspired and hear how you made it.

If I were to make bullet points of what was important for me and what shaped me, it was definitely having a figure, like my advisor. I had two great advisors. I had my advisor back in undergrad, Petros Maragos, who is at NTUA. He was one of the people that gave me the confidence that "you know what, I can actually do this!" Then, of course, Jitendra Malik. Like what we said before, I trusted both Jitendra and Petros. I had full confidence in what what they were advising me to do. When you have great leaders, you also have a great team. For example, in Berkeley… in 2010, I got on a plane from Greece to the US. On my first day at UC Berkeley, I met Bharath Hariharan. He is now faculty at Cornell, one of the most brilliant, wonderful people I've ever met. I had the luck to work with this guy and be next to him for six years. Then, of course, more came after that with all of these labmates, this team around you that is wonderful, that you can trust, that you can ask for help, that you can talk to if you have technical difficulties for your paper, or personal, when you're mad, upset, or depressed because of what's

"From my trip to Kigali, Rwanda, where I taught at the African Institute of Mathematical Sciences."

going on in grad school. It's a wave. It's a roller coaster.

"This team around you that is wonderful, that you can trust, that you can ask for help, that you can talk to if you have technical difficulties for your paper, or personal, when you're mad, upset, or depressed because of what's going on in grad school."

When I interviewed Jessica Sieren, she told me: *"If you're not having a crisis during your PhD, then you are not doing it right."*

Exactly, you have many crises… low points, high points, low points, high points.

How do you get through it?

To me, I feel that it was the people. It was the fact that I had my labmates. When you're in a low point, your friends are going to be very helpful! Your labmates are also going to be helpful because they're the ones that can give you the confidence that you can actually do this technically. To me, that was very important. I had the moral support from my friends, and I had the technical support from my lab mates.

That's a very inspiring lesson: mentorship and friendships are the way to succeed.

Mentorship, not only between an advisor and a student, but also peers, people that are with you all the way.

Of all these brilliant people that you worked with, what is the main lesson that you learned from any of them?

I think I would say from my advisor, Jitendra, the biggest lesson that I learned from him is to work on important problems, how to distinguish what is important and what is not. That is something that Jitendra has taught me throughout the years, back then until today. Another really big impact was the advice from my other mentor and colleague, Ross Girshick: whatever you do, you have to do it right.

What if you don't?

Then take a step back. Reformulate. Rethink. Redo everything.

Did this happen to you?

Of course - If you're working on a project in which, at the time of submission, you're like, *"Oh, this is perfect! I've solved the problem. Let's move on."* Then you haven't solved the real problem.

"Me in Seoul drinking tea! I prefer coffee!"

Yoshua Bengio told me that his success in his work is not because he's smarter than the others, but because he's able to focus a lot.

Correct - This is something that I actually value a lot, being focused on a problem, making a dent, and solving it. This is something that the young generation doesn't quite get, to focus on a problem and contribute to this one problem.

After interviewing many brilliant scientists in this community, I find that very often, scientific greatness comes with exceptional personal kindness. Did you notice that?

I want to believe that's true. I think that, otherwise, if that's not the case, the community will put you in the fringe. If you're not a mentor, both in terms of your work but also in terms of your persona, in terms of what you inspire others to do, then I don't see how you can become a prominent figure in any scientific field.

"If you're not a mentor, both in terms of your work but also in terms of your persona, in terms of what you inspire others to do, then I don't see how you can become a prominent figure in any scientific field."

How long have you been away from Greece?

Oh! Too long!

What is the thing that you miss the most?

My parents and my family...and the food... and the Greek Islands.

Well, you can visit often!

I try to go there every summer.

Which Greek Island is the best?

Oh! That's like choosing your favorite child! I am very biased because my parents actually reside on an island, Paros, which I find very beautiful. One of my favorite islands, if I didn't pick this one, is Folegandros. It's one that's not on the radar. Tourists have not picked up on it, so I like that. It's mostly Greek tourists that go there in the summer. It's wonderful scenery, beautiful beaches, and the food is great.

Greece is underrated. How good it is to go and see the beauty of Greece…

Definitely, but the last couple of summers that I've been there, I've seen more and more tourists. A lot of people want to go. It's also very far.

Well, not for me. I live very close to there so every few months, I visit one of the islands.

I hate you! [*laughs*] Oh! You're so lucky!

"How would Kaiming react? How would Kaiming present it? How would Kaiming feel afterwards?"

Tell us about Greece. What does Greece need to do now to succeed?

You mean in science?

Yes.

What I would like for the Greek leadership and government to do is to try to attract a lot of the talent that has left the country. There is actually a term for this. It's called brain drain. It means that a lot of people who are successful in science leave the country because there is no opportunity for them there.

What is the nicest thing that you found in the US that you didn't have in Greece?

I think my friends!

"Judy Hoffman (brilliant, funny and very talented, now faculty at Georgia Tech) and I in one of the most fun CVPR conferences in Portland. You can't see the empty margarita glasses!"

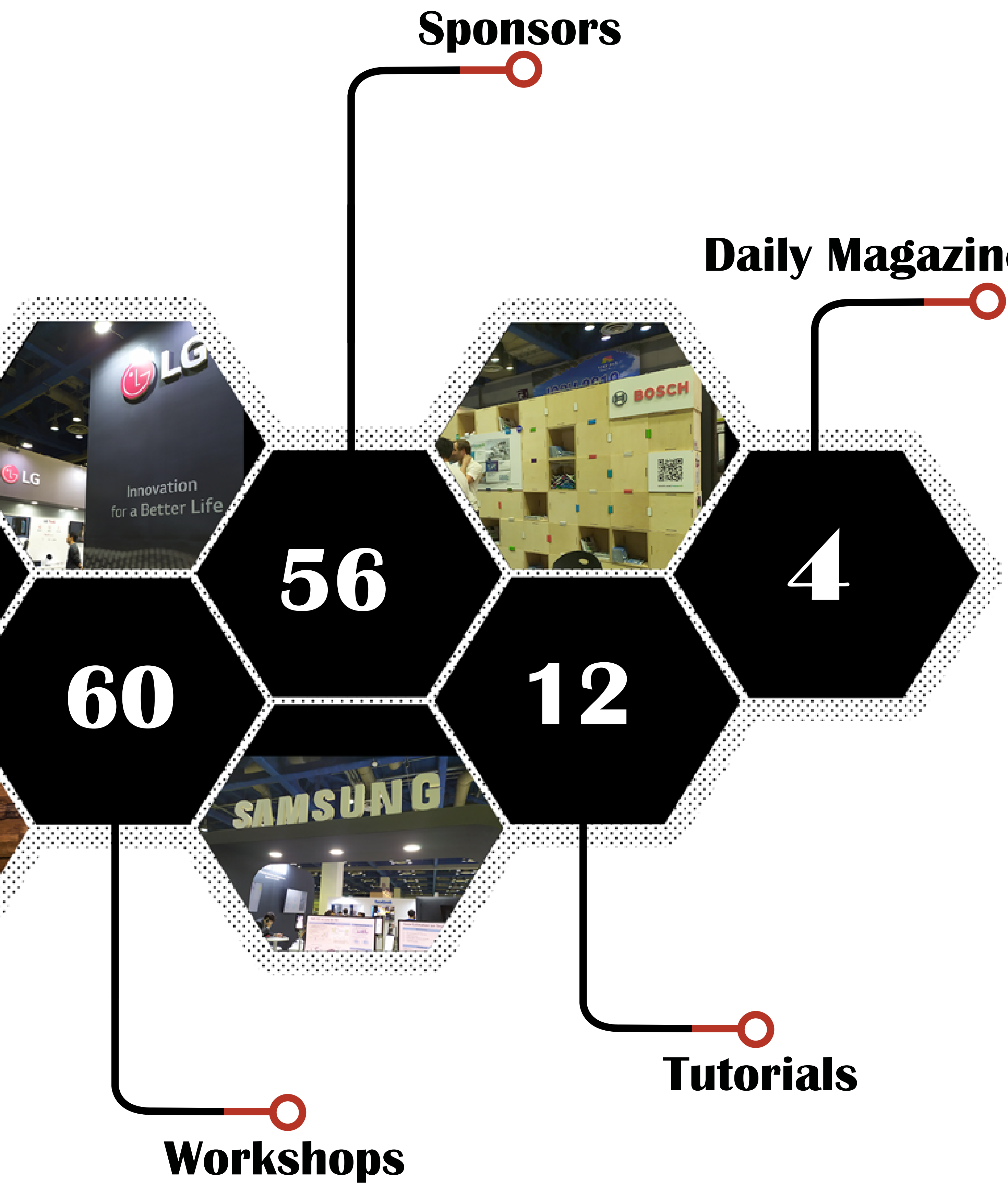
The top ICCV of all times

Submissions: 4,303

Exhibitors: 72

Attendees: 7,501

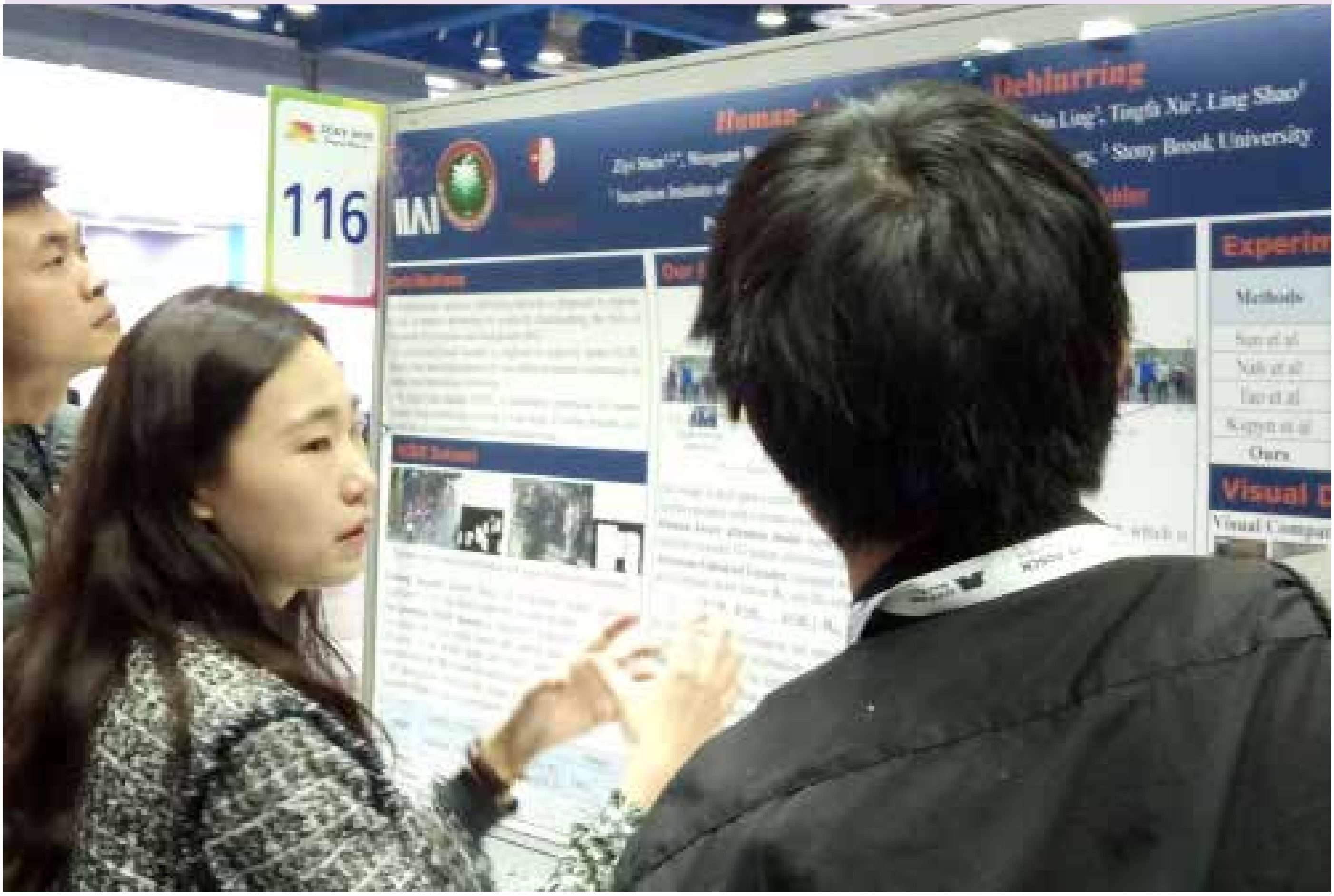
Accepted Papers: 1,075


Sponsors
56
Daily Magazines
4
60
Workshops
12
Tutorials
LG
Innovation
for a Better Life
BOSCH
SAMSUNG

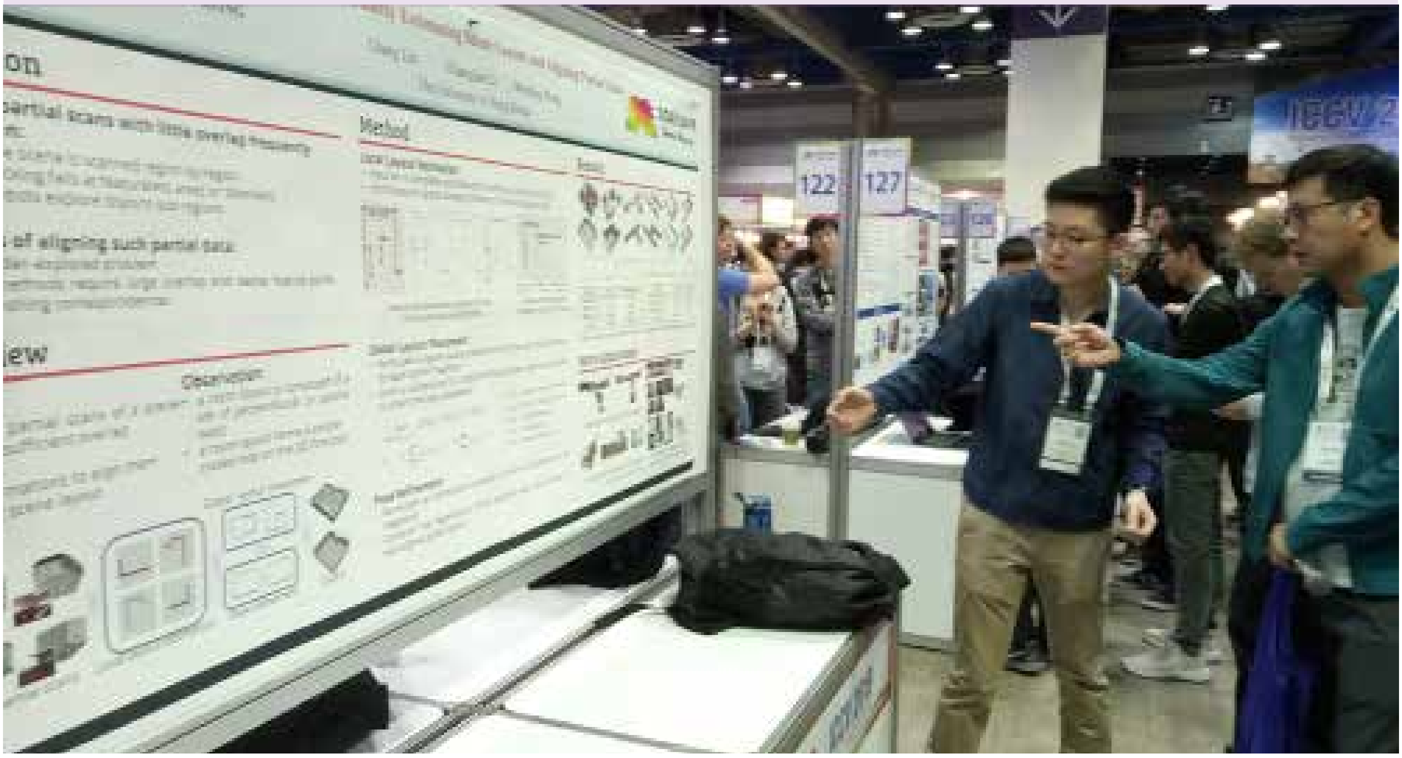
Yiming Wang presenting Active 3D Classification of Multiple Objects in Cluttered Scenes. Robust object recognition is an essential skill for robots to perform autonomous manipulation, while classifying objects in cluttered scenes can be challenging with single-shot methods due to severe occlusion. They therefore propose an active vision approach through 3D reconstruction following a next-best-view (NBV) paradigm.

Qing Liu presenting Localizing Occluders with Compositional Convolutional Networks: they introduce a compositional deep model that is regularized to be fully generative in terms of its high-level features. Their model learns a dictionary of part models, represents 3D objects as mixture of 2D templates, and localizes occluders with high precision. The proposed compositional convolutional network is robust to occlusion and outperforms baseline models on object classification tasks by a large margin.

Ziyi Shen presenting Human-Aware Motion Deblurring



Cheng Lin presenting Floorplan-Jigsaw: Jointly Estimating Scene Layout and Aligning Partial Scans




ICCV 2019

RSIP Vision boasts extensive experience across a range of **medical segmentation** projects: CT, MRI, fluorescent, ultrasound, OCT, laparoscopy images, endoscopy and more. All this work is based on **deep learning technology**, which offers much more robust solutions.

By Ilya Kolver

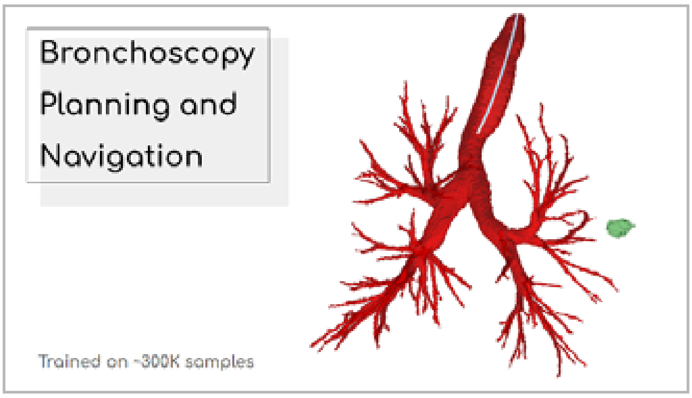
Let's have a look at three projects in particular. The first project deals with airway segmentation for the purpose of **bronchoscopy surgical intervention**. This segmentation helps during pre-operation planning. In addition, it assists the surgeon in planning the optimal path to reach the lesion, for instance, to extract a tissue for biopsy or some other form of treatment.

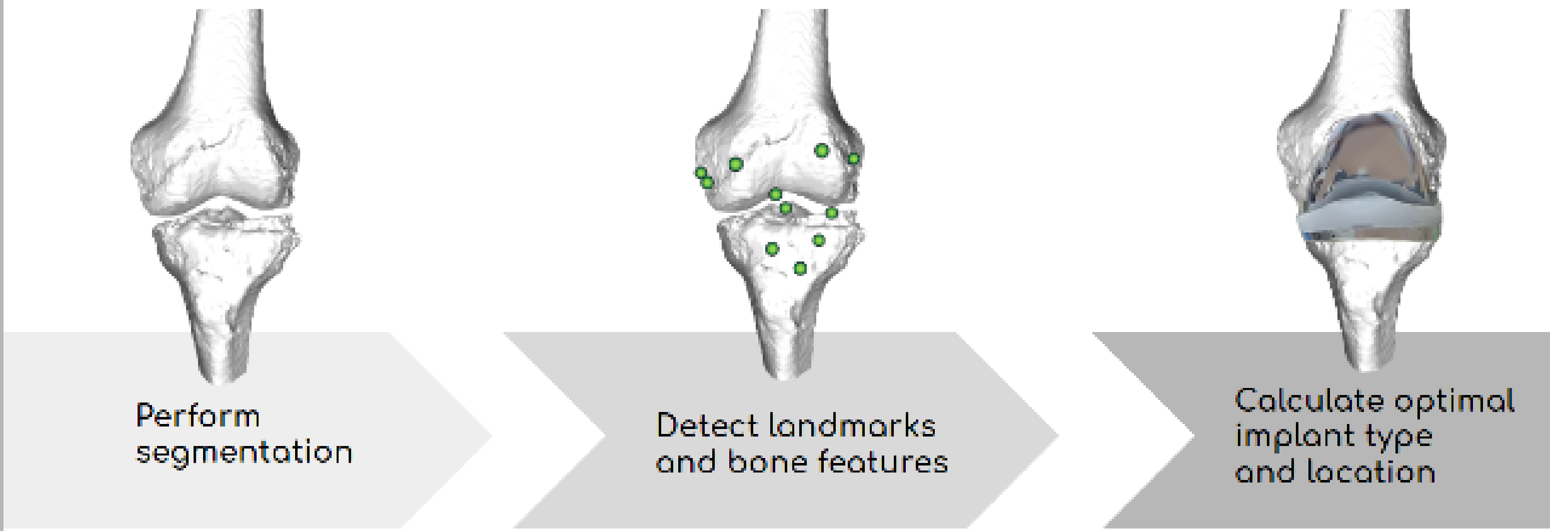
Since the catheter does not carry a camera, the surgeon must find another method to guide the navigation. This is generally performed with the help of an electro-magnetic sensor placed at the tip of the catheter. In such cases, the surgeon only knows information gathered through fluoroscopic imaging or by the position of the catheter in the electromagnetic field. The previously designed segmentation places a strong constraint on the position of the catheter in the CT used in pre-op planning.

The second project involves **joints segmentation**, a crucial process in developing a solution specific to the patient's body. This segmentation produces a patient specific implant based on the most accurate size and angles. In some cases, a 3D print of the specific implant integrates with the precise anatomy of the patient.

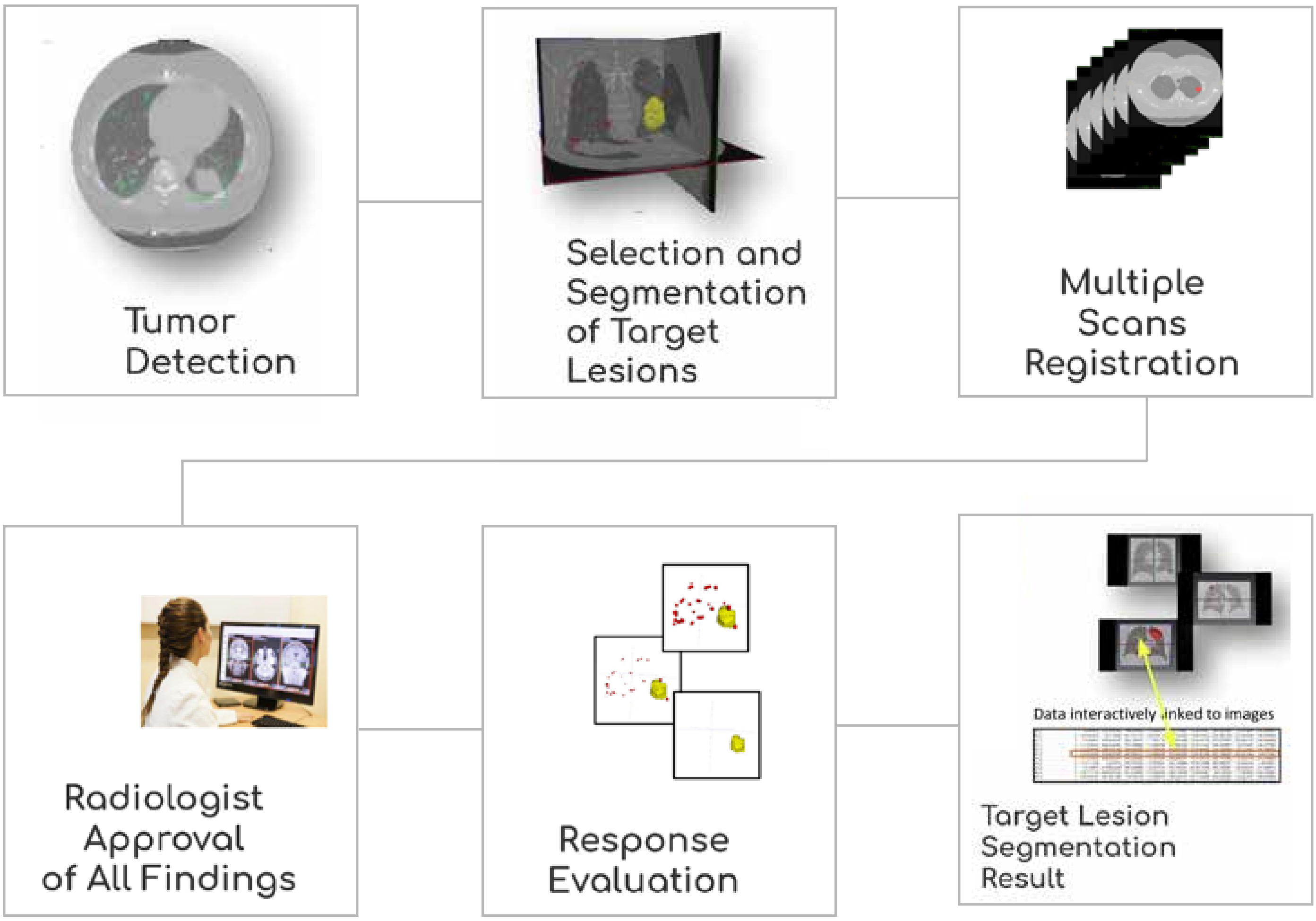
The third project focuses on **lung nodules**. RSIP Vision developed a system called **AUTORECIST**, an automated system capable of calculating the volume of nodules and their evolution during and after treatment. The radiologist measures the volume of nodules at the time of the diagnosis. This measure is then compared over time to assess the progress of the nodule in reaction tothe treatment and to monitor whether new nodules have sprouted. Segmentation provides very precious information: the shape and the size of the nodule, which might reveal whether specific treatment is successful or the need for an alternative treatment.

The **RSIP Vision engineering** team has successfully developed these systems. currently in use by our clients. If you also have a medical segmentation project, take us along and **put the power of artificial intelligence and deep learning in your hands**.





AUTORECIST



by Amnon Geifman

Open3D for 3D processing

A few weeks ago, we have seen together how to generate a 3-dimensional model from a collection of 2D images using Visual SfM. Although the final 3D model we generated was quite appealing, it was far from being an industry-level one. When dealing with a point cloud generated from sensors, there are several challenges such as outlier removal, registration, retrieval and more. Since 3D objects might be very complex to handle, a special tool is needed. To this end we tried for you the Open3D library in python. Open3D is an open source library that supports rapid development of software that deals with 3D data. The backend of the library is highly optimized and allows efficient usage of many 3D algorithms. Among the core feature of Open3D you can find: basic 3D data structures and algorithms, scene reconstruction, surface alignment, visualizations and more. In this article, we will familiarize you with a few useful features of this library.

Open3D settings and I/O ✓

Open3D can be used in either C++ or python. If you are using the C++ version, you can check out which compilers are supported and compile it on your machine. Luckily enough, on python we can get Open3D using pip install so we just write in the command line:

```
pip install open3d
```

Most of the algorithms in Open3D are self-contained, so that we don't need to install any further libraries and we are ready to go.

In this article we build upon the work we did in the VisualSfM article. There, we end up with 3D reconstruction, meaning we had camera locations and 3D points. We also explained how to parse the 3D model so you only need to copy the parser of the 3D points and save the result into a .ply file. You can also use as input any point cloud file you desire from any other Structure from Motion software.

Open3D was meant to be simple and easy to use. In order to read the file "pointcloud.ply", print the result, and write it down, we simply run the following lines of code:

```
import numpy as np
import open3d as o3d
import copy
print("Reading and writing file...")
pcd = o3d.io.read_point_cloud("pointcloud.ply")
print(pcd)
o3d.io.write_point_cloud("copy_of_file.pcd", pcd)
```

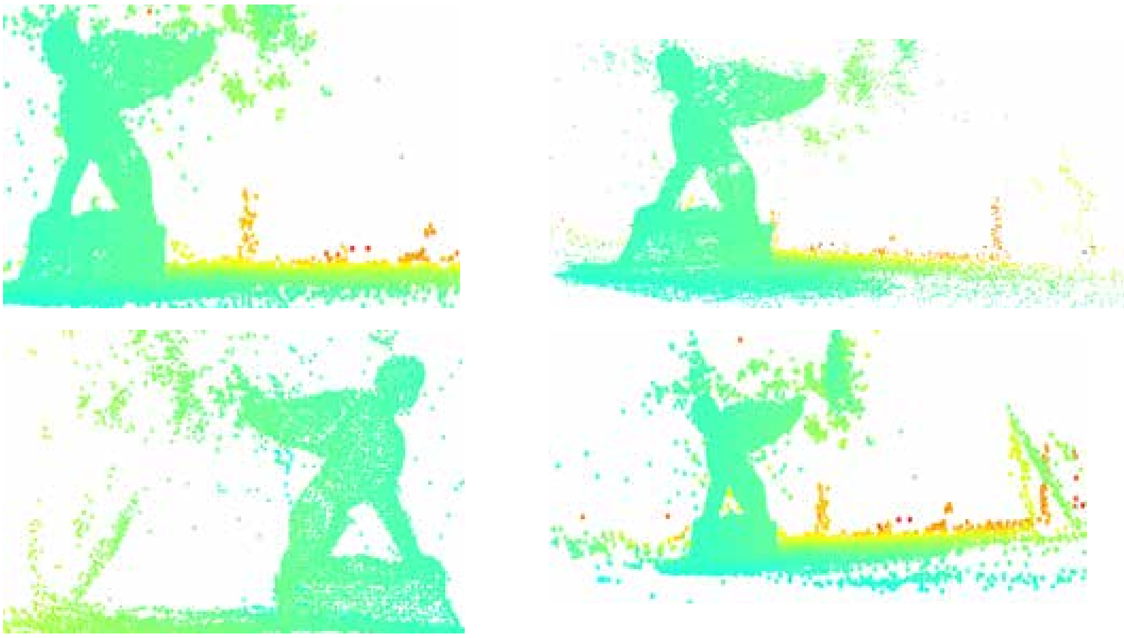
This is fairly simple, and now we are ready to do the cooler stuff.

Outliers Removal

When collecting data from scanning devices, usually the point cloud contains many outliers and artifacts. Since, in our case, the point cloud was generated by a 2D images, some outlier matches were triangulated to an erroneous 3D point. To cope with these phenomena, Open3D offers several outliers removal algorithms. We next describe each one of them.

The first method we use is downsampling. In order to represent the object more accurately in a point cloud, it is sometimes beneficial to downsample the data using some heuristics. We show below two kinds of downsampling; the first is downsample by voxels, meaning that we divide the space into voxels and represent all the points at each voxel by their centroid. The second down sample method we demonstrate is uniform, meaning that we collect every n-th points uniformly. The code and the results can be seen below:

```
print("Downsample the point cloud with a voxel of 0.02")
voxel_down_pcd = pcd.voxel_down_sample(voxel_size=0.02)
o3d.visualization.draw_geometries([voxel_down_pcd])

print("Every 5th points are selected")
uni_down_pcd = pcd.uniform_down_sample(every_k_points=5)
o3d.visualization.draw_geometries([uni_down_pcd])
```



The results of the voxel downsamples are the top figures, while the bottom figures are the results of the uniform. It seems that the downsample is nice, to begin with, but alone it is not enough. To improve it, we apply the next methods on our downsample point clouds.

The second method we use is statistical outlier removal. In this method, we opt to remove points that are far from their neighbors compared to the average of the point cloud. We need to choose the parameters of how many neighbors to take and a threshold level based on the standard deviation of the points. We use here an aggressive removal by setting low std value. The code and results are below:

```
print("Statistical oulier removal")
cl, ind =
voxel_down_pcd.remove_statistical_outlier(nb_neighbors=20,std_ratio=0.005)
```

Here the red points are the points that have been removed. We can see that this method starts to look better and our main object in the image starts to be clearer. The last method we present is Radius outliers removal. This method removes points that have a few neighbors in a given sphere around them. Here again, we need to tune the number of neighbors and the radius we are looking at.

```
print("Radius outlier removal")
cl, ind = voxel_down_pcd.remove_radius_outlier(nb_points=16, radius=0.05)
```

As before, the red dots are points that have been removed. It seems that we have a winner and that the radius method is the best for our purpose. Now that we have a clean point cloud, we can save it and move on to some other useful applications of Open3D.

ICP Registration

There are many applications where we would like to register two point clouds of the same object. To this end, Open3d offers a few registration algorithms: we will use the iterative closest point. ICP algorithm is applied to minimize the difference between two clouds of points, i.e. to register them. It uses an initial rigid body transformation and refines the solution at each iteration by minimizing the distance between the clouds. There are a few variants of this algorithm where Open3d supplies point to point registration and point to plane registration. In the lines that follow, we will use the former one.

To apply the algorithm, one can load his two point clouds directly. What we will do is to apply a random rigid transformation on our object with the open3d.geometry. PointCloud.transform method; then, we save it into a separate file. The initial pair of point clouds (in the same coordinates system) looks like this:

Next, we align the two point clouds to one another. Usually, a global registration algorithm has to be applied in order to give a good initial guess for the rough transformation. Since we know what is the true transformation, we start with a perturbed version of it. Note that the ICP algorithm only refines the transformation, hence a good initial guess is required. The code below performs the point to point registration. The first function is for visualization while the rest is the ICP algorithm.

```
def draw_registration_result(source, target, transformation):
    source_temp = copy.deepcopy(source)
    target_temp = copy.deepcopy(target)
    source_temp.paint_uniform_color([1, 0.706, 0])
    target_temp.paint_uniform_color([0, 0.651, 0.929])
    source_temp.transform(transformation)
    o3d.visualization.draw_geometries([source_temp, target_temp])


source = o3d.io.read_point_cloud("D:/CVN dir/november19/icppc.pcd")
target = o3d.io.read_point_cloud("D:/CVN dir/november19/icppc2.pcd")
threshold = 0.002
trans_init = np.asarray([[0.9, 0.01, -0.50, 0.55],
                             [-0.1, 0.9, -0.2, 0.7],
                             [0.45, 0.25, 0.83, -1.4], [0.0, 0.0,0.0,
                              1.0]])

draw_registration_result(source, target, trans_init)
print("Initial alignment")
evaluation = o3d.registration.evaluate_registration(source, target,
threshold,trans_init)
print(evaluation)
print("Apply point-to-point ICP")
reg_p2p = o3d.registration.registration_icp(
          source, target, threshold, trans_init,
          o3d.registration.TransformationEstimationPointToPoint())
print(reg_p2p)
print("Transformation is:")
print(reg_p2p.transformation)
print("")
draw_registration_result(source, target, reg_p2p.transformation)
```

Below you can see the final alignment we were able to achieve:

KD-tree

The last tool that we present in this article is the KD-tree (K dimensional tree). This is a data structure for point cloud that is used to perform fast retrieval of points and nearest neighbors from the point cloud. It partitions the space into a tree-like data structure, in which a search can be done efficiently. We will usually use it when we would like to inspect some specific areas in the cloud. Open3d uses FLANN to build the KD tree.

In our example, we paint a point in red and we show how to retrieve its neighbors. We search for the point's 200 neighbors and assign them the green color. The code and the results can be seen below:

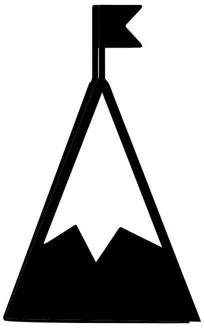
```
print("Testing kdtree in open3d ...")
print("Load a point cloud and paint it gray.")
pcd = o3d.io.read_point_cloud("D:/CVN dir/november19/icppc.pcd")
pcd.paint_uniform_color([0.5, 0.5, 0.5])
pcd_tree = o3d.geometry.KDTreeFlann(pcd)

print("Paint the 1500th point red.")
pcd.colors[1500] = [1, 0, 0]

print("Find its 200 nearest neighbors, paint blue.")
    [k, idx, _] = pcd_tree.search_knn_vector_3d(pcd.points[1500], 200)
np.asarray(pcd.colors)[idx[1:], :] = [0, 0, 1]

print("Find its neighbors with distance less than 0.2, paint green.")
    [k, idx, _] = pcd_tree.search_radius_vector_3d(pcd.points[1500], 0.2)
np.asarray(pcd.colors)[idx[1:], :] = [0, 1, 0]

print("Visualize the point cloud.")
o3d.visualization.draw_geometries([pcd])
```

Conclusion

Open3d is an extensive open source library for 3D processing. It supplies many useful tools to handle 3D data structure algorithms and visualizations. It can handle many types of 3D data such as meshes, point clouds, RGBD images and more. In this article, we showed its usage on point clouds generated from a collection of 2D images. We demonstrated how to clean outliers, how to register point clouds with point to point ICP and how to use KD trees. From our experience, open3D is quite robust allowing for industry-level applications in both C++ and Python. We recommend you try it yourself with the code above. Enjoy!

Task-driven vs challenge-driven project management

By Nissim Avitan



Task-driven project management is the standard method used to manage software programming projects. The customer provides specifications and requirements for the software component, and the project manager derives a well-defined set of tasks designed to meet the specifications. These tasks, and the inter-relations between them, form the basis for management of the project

This approach works best when the there is a clear path leading from the customer specifications to the finished product, in other words, when the uncertainties are minimal, and the project mainly involves completing each of the well-defined tasks.

On the other hand, when there are inherent challenges in the project, and the solution to these **challenges are non-trivial**, then the task-based approach breaks down. In such cases it is meaningless to try and derive specific tasks, since the path forward is not clear. Here, challenge-driven project management, which is used by RSIP Vision in many of our clients' projects, is much more effective.

Challenge-driven project management is all about identifying the main hurdles, or challenges, in the development path, and finding solutions to these challenges. Once the challenges are well understood, and the outlines of a solution for each challenge exists, then the project becomes a standard development project, and the task-driven management approach can be applied.

Examples of challenges may include finding an algorithm to achieve the most consistent and accurate results or finding a modification to an existing algorithm, such that it runs in an allotted time on a given hardware platform.

Challenge-driven project management usually includes a **proof of concept (PoC) stage**, where a solution to the challenge is devised, and preliminary tests are performed to verify that the solution is valid. In some cases, different solutions may be devised and tested in parallel by different

members of the development team to save time, with the best solution being selected based on the results. In still other cases, different members of the team may work on different challenges, so that by the end of the PoC stage all the major challenges have a solution outline.

Challenge-driven project management also requires a high level of customer interaction. While in task-driven project management the customer usually reviews progress and makes some decisions (such as GUI design, input and output format, etc.), in challenge-driven project management the customer is much more actively involved, especially during the PoC stage. Often, while developing a solution, new scenarios may arise which were not initially foreseen, and the customer needs to refine the specifications to deal with these scenarios. In other cases, the initial specifications and requirements may turn out to be unfeasible, and the customer may need to relax some of the requirements.

Don't miss the latest talk by Moshe Safran on Applied Deep Learning for Digital Pathology!

Computer Vision Project Management

Computer Vision Project Management is a series of lectures and articles conducted by RSIP Vision's CEO Ron Soferman, many of which are published as a regular column on magazine Computer Vision News, in the project managements tips section.

Everything a project manager in computer vision should know... **at the click of a button**

"Even the biggest hammer cannot replace a screwdriver!"

- How to solve all kind of challenges
- Team leadership and management
- How to implement deep learning
- Validation and test techniques
- What are the best practices

Did you miss an article? No worries, you can find them all in the **Project Management section of RSIP Vision's website!**

COMPUTER VISION EVENTS

RSNA 2019
Chicago, IL ---- Meet Us ---- Dec 1-6

Hamlyn Winter School on Surgical Imaging...
London, UK ---- Dec 2-6

BIOMEDevice
S.Jose, CA ---- Dec 4-5

NeurIPS: Neural Information Processing Systems
Vancouver, Canada ---- Dec 9-14

The AI Summit 2019
New York, NY ---- Meet Us ---- Dec 11-12

CES 2020
Las Vegas, NV ---- Jan 7-10

PrecisionAg VISION Conference
Seattle, WA ---- Jan 14-16

EI - Electronic Imaging
Burlingame, CA ---- Jan 26-30

RE-WORK Deep Learning Summit
S.Francisco, CA ---- Jan 30-31

SPIE Medical Imaging Computer-Aided Diagnosis
Houston, TX ---- Feb 15-20

ICPRAM - Pattern Recognition Applications and...
Valletta, Malta ---- Feb 22-24

VISIGRAPP - Computer Vision, Imaging and...
Valletta, Malta ---- Feb 27-29

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

FREE SUBSCRIPTION

(click here, its free)

Dear reader,
Do you enjoy reading Computer Vision News? Would you like to receive it for free in your mailbox?

You will fill the Subscription Form in less than **1 minute.** Join thousands of AI professionals and receive all issues of Computer Vision News as soon as we publish them. You can also read Computer Vision News in PDF format (though this online view is way better) and visit our archive to find new and old issues as well.

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