

Computer Vision and Machine Learning

Visual Computing

Dynamic Vision and Learning

At the increasingly open border between research enabling computers to see and advances in teaching them to learn, things are hopping. Interdisciplinary research is progressing rapidly, with wide-ranging implications for technology, business, and society. And a lot of the action is taking place right here.

As Rudolf Mößbauer Tenure Track Professors, Laura Leal-Taixé (Dynamic Vision and Learning Group) and Matthias Nießner (Visual Computing Lab) were both recruited through a TUM-IAS program designed to establish novel and often interdisciplinary fields at TUM. Daniel Cremers heads the Chair of Computer Vision and Artificial Intelligence in the TUM Informatics Department and currently is a TUM-IAS Carl von Linde Senior Fellow. Nießner is the host of the TUM-IAS Focus Group on Visual Computing, with Hans Fischer Senior Fellow Leonidas Guibas of Stanford University and Hans Fischer Fellow Angel X. Chang of Simon Fraser University. Cremers hosts the Focus Group on Computer Vision and Machine Learning, in collaboration with Rudolf Diesel Industry Fellow Michael Bronstein, whose current affiliations are with Imperial College London and Twitter, which acquired his startup Fabula AI in 2019.

In March 2020, the TUM-IAS conducted interviews with these six researchers about their trailblazing work and the framework for collaborative exploration that the Institute offers them. Their answers have been edited for clarity and length.



Daniel Cremers



Michael Bronstein

» ... the core question revolves around how to expand the success of machine learning and so-called deep networks to other application domains beyond standard images.«

Q: Clearly, all six of you have a lot in common in terms of research interests. What are the most important common threads, and what kinds of differences in topics or emphases distinguish one group from the other?

Cremers: For me, the core question revolves around how to expand the success of machine learning and so-called deep networks to other application domains beyond standard images. One of the core ingredients that Laura and I and Michael in particular are working on is so-called graph neural networks. One of the most important enabling advances in the last couple of years, which has enabled many things not just in vision but well beyond computer vision, is the advent of so-called deep neural networks. They were actually developed for vision, back in 2012, to address one of the biggest quests in computer vision, a quest that one of our colleagues set out to solve – it's a quest called the ImageNet Challenge. One of my friends and colleagues at Stanford, Fei Fei Li, compiled with her students a data set of millions of images, everyday-type photographs, where every photograph is classified as being an image of an airplane, an image of a car, or a cat, or some other object. The challenge was to devise a machine that tells us what's in each image. And to summarize a long story, with the advent of deep networks, we were able not only to reach human-level performance, but even to outperform the average human on this challenge.

This was a big breakthrough, and I don't think it's really been acknowledged much in the public. When people hear about artificial intelligence, they generally think about chess-playing computers. And the truth is, for a machine to play chess, that's not a challenge. Humans on the other hand did not evolve to play chess, but they did evolve to recognize objects and images. To beat the human on a task that humans are arguably designed to solve was a much more significant breakthrough. And from computer vision, these deep networks have since swept into virtually all areas of science and data analysis. The question is how to go beyond standard image analysis. That means for example how to deploy it for dynamic data, as Laura is doing with videos.



Laura Leal-Taixé



Matthias Nießner



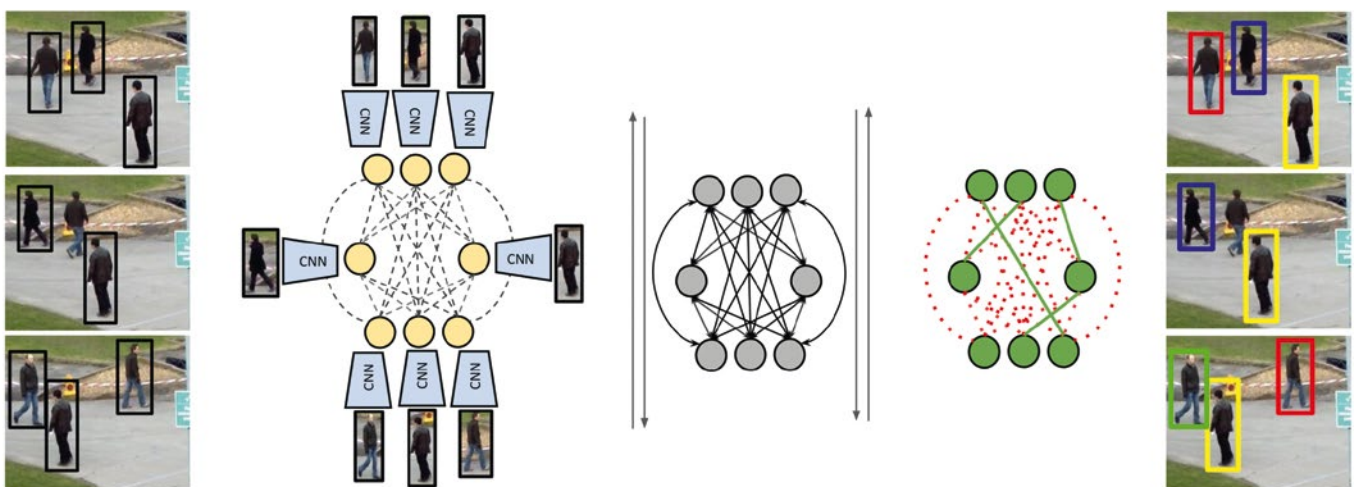
Leonidas Guibas



Angel X. Chang

Leal-Taixé: The interest of our group is dynamic vision and learning. The key word there is dynamic. We are interested not only in analyzing images, but in actually analyzing videos. A very strong component is analyzing motion in a scene. In particular, we look a lot at human motion. For example, for an autonomous car, you want to know where are the pedestrians around the car, where are they going to go in, let's say, the next ten seconds. So there's a lot of trajectory prediction and tracking. We want to allow the robots or autonomous cars to have a sense of what is around them and a sense of what is moving around them and how it is moving. We focus on humans because their motion is very interesting. It's hard to predict. The motion of a car is constrained by the lanes, for example,

and the rigidity of the car itself, while the motion of a pedestrian is quite random sometimes. It's hard to predict where pedestrians want to go, and they might suddenly change where they want to go and turn around. All of these things make the problem very interesting from a scientific point of view, because we're interested in exploiting not only the motion of each person individually, but also the relationship of motion between the pedestrians, for example, how they avoid each other when they cross paths, as well as the motion of pedestrians with respect to cars. So when you reach a crossing, you typically stop a little bit before crossing, you look around, and these are interactions that can be modeled with such a neural network.



a) Input

b) Graph Construction +
Feature Encoding

c) Neural Message Passing

d) Edge Classification

e) Output



Michael Bronstein (left) with the team of Fabula AI at Twitter office in London, Photo: Stacey Conti.

Cremers: The so-called graph neural networks operate not just on Euclidean data – images are typically Euclidean, that is, they have planar structures – but also on graph structures that are non-Euclidean.

Q: Does this help to explain how it could be that your research touches on applications like protein interactions, food design, fake news, and particle physics as well as enabling computers to better understand, and also better synthesize, moving pictures?

Bronstein: What they have in common is this new class of machine learning methods that we call geometric deep learning. Graph neural networks, or graphic representation learning, is a particular example. Basically these methods try to extend deep neural networks to non-Euclidean structured data: graphs, manifolds, point clouds. We want to exploit the geometric structure of the data in a mathematically principled way. We try to build deep neural networks that allow us to learn on data that has these underlying structures, of a graph or a manifold or a point cloud.

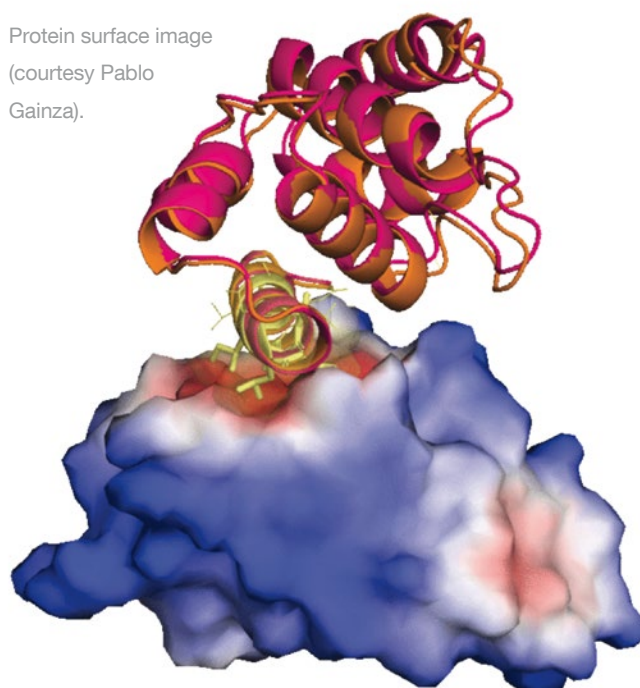
In all the examples you mention, we have data that lives on a graph, and the graph is part of the data. In the case of fake news, we have a social network and a propagation graph representing how news spreads in time. In the case of foods, we have a graph that represents protein-protein interactions. There are about 20,000 proteins in our body; their binding to each other is responsible for multiple biochemical processes.

Simplistically speaking, sometimes these processes break down and we get sick. We then take a drug, which is designed to bind to one or multiple proteins and fix these broken processes. We can represent the effect of a drug as a signal on the protein-protein interaction graph and use graph neural networks for “drug repositioning” – finding molecules that can act, for example, as oncological drugs.

Let’s say, we take examples of drugs that are approved for use against certain types of cancer, and train a classifier that predicts oncological “drug-likeness” from their protein binding signals. We can then use this classifier to screen other molecules. Take for example food: Some fruits and vegetables contain compounds that belong to the same chemical classes as some chemotherapeutic drugs. By using our drug-likeness classifier, we can identify foods rich with molecules that are similar to oncological drugs, though in much lower concentrations. We call them “hyperfoods.” Unsurprisingly, all the boring foods like cabbage or celery or green tea are hyperfoods.

The thing about these methods is that it’s completely data-driven. So if tomorrow you want to find, for example, potential drug-like candidates that allow you to beat a particular kind of cancer, or potential antiviral drugs to fight the new coronavirus, we could apply the same process.

Protein surface image
(courtesy Pablo
Gainza).





TEDx Talk by Michael Bronstein in Lugano, 2019.

Q: To what extent are the neural networks you're talking about implemented in software, and to what extent in hardware?

Cremers: Initially they're in software. Typically the hardware you need to run them is so-called GPUs. Laura and I spend a lot of time discussing how much more GPU time we need to buy for our team to continue working. We're spending a significant amount of funding on that kind of hardware. There are efforts to develop hardware implementations of deep networks, different ones that are, say, more power-efficient. But this is not really our expertise. We are more on the algorithmic and software side.

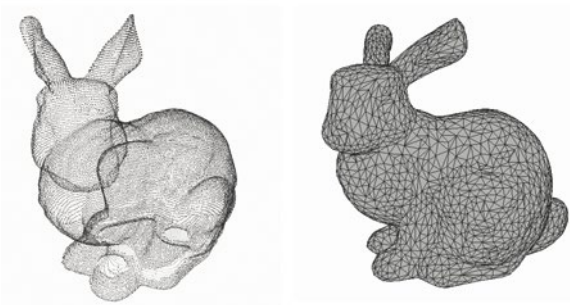
Q: Matthias, Leo, and Angel, how would you describe the research topics, approaches, and potential applications that best define your Focus Group?

Nießner: We've been working on bilateral projects for a very long time actually, at least five or six years now. This is how various topics are emerging. One commonality is the underlying 3D understanding of things. From a geometric perspective, or from a language perspective with Angel. My specific angle to it is more

the visualization part. I want to recreate photorealistic images. Eventually I would like to create holograms of 3D worlds and make sure we use all the content available in virtual environments. To synthesize photorealistic humans, faces, and environments, that's already being done, and by better understanding the environment you can get better synthesis. With Leo we're asking: Can we first figure out the object itself, do we know where objects are, and can we get the geometries of these objects? And then, can we use this information to synthesize better results? All of these areas come together.

From a methodology standpoint, all of our efforts are anchored in 3D learning techniques and 3D computer vision, with a lot of current progress in neural networks focused specifically on 3D data. We all argue that the world is not 2D, right? If you want to understand what things do in a 3D environment, you want to directly learn in 3D. There are good examples for that. A human has two eyes to see things in stereo. We have depth perception, we learn spatial correlations, we learn how to interact in 3D. We know how to describe and to touch things in 3D. And then we also know how to visualize things and imagine things in 3D.

56 **Guibas:** Three-dimensional understanding poses many challenges when it comes to machine learning, because 3D data tends to be different. Two-dimensional data is regular pixel grids, all with the same format. But in 3D there are many different representations that have existed for decades and serve different communities – point clouds or meshes, for example – and all of them tend to be irregular. A lot of the traditional machine learning techniques, especially convolutional networks, require regularity to be able to share coefficients and other optimizations. So if you have this irregular data, you have to do things differently. Part of our work has been to address this problem, how to represent 3D and how to process it even though it is not regular.



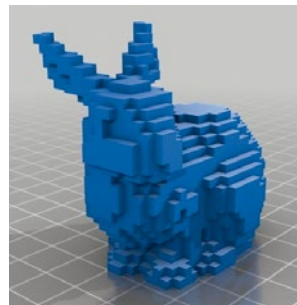
Point-Cloud

Surface-Mesh

Chang: I come from a natural language processing background, so what I'm interested in is how people talk about 3D things, using language. And so what can be very, very interesting is the connection between language and 3D representations. Recently there have been lots of methods where you can basically, using neural networks, take language and then take images and put them into a shared representation. And then through that we can connect language and other modalities. So with Dave Chen, a doctoral candidate in Matthias's team, we are working on being able to localize objects in 3D – where someone might say, for instance, “the chair that is in the corner of the room,” and then be able to actually identify, in a 3D bounding box, what the person is referring to.

Guibas: With Matthias, I share many interests in computer vision, and in fact inspired by Angel I have also started dabbling a bit in language in relation to geometry. My own background is much more in geometric algorithms and geometry processing, the more classical field before the advent of deep learning.

So I'm dealing with both the design of geometric algorithms and also suitable representations for 3D geometry, including how to take noisy data and try to improve it in various ways. We do research in classic areas like image processing, where one has noisy images and tries to get rid of the noise, and also in geometry processing where one has noise and tries to improve the geometry. But together we can try to do something that didn't exist before, such as in texture processing, being able to take images that are living on a manifold, images living on meshes, and capture them better and understand them better. That's one interesting direction we're pursuing in collaboration with Matthias.



Volumetric

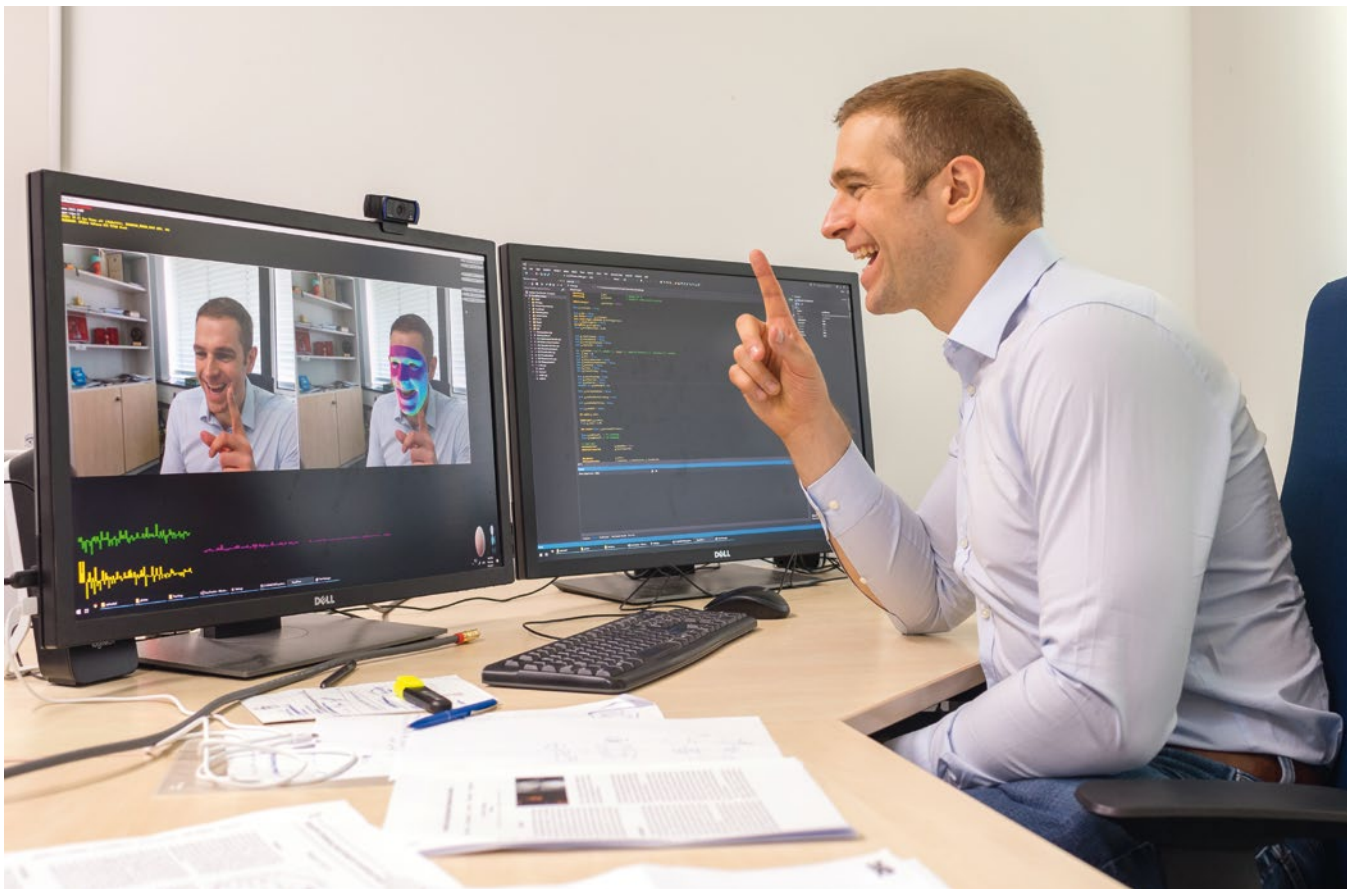


Multi-View Images

On my side, I have a bit more emphasis on the content creation side, that is, being able to not just capture some object that's out there in the world, but to have interesting tools for creating new geometry, new objects essentially.

Q: Would you call that 3D synthesis?

Guibas: Synthesis, indeed. And in fact these come together in another collaboration with Matthias where the goal is to replace extant scanned geometry with a CAD model that has been adapted to the scan, so that you can have a clean model that represents an object in the world. We've started a project that focuses on understanding scenes, with the purpose of acting on them. Not just what is there, but what could be there, how things could be different, how can I change them, so that we make it possible for an effect to happen in space. How do I close the door? How do I open the drawer? How do I move my laptop from this table to that counter? The focus on understanding not just what is but what could be is a central part of this effort.



So there are two specific themes of research we could highlight. One would be trying to replace dirty, noisy scanned objects with clean-cut models. Another is to be able to acquire high-quality textures of objects and to be able to use that also to understand the scene.

Q: What are the hurdles to achieving those two things?

Nießner: You're never going to get perfect information. A camera has noise, and you only see things from the current perspective. Humans are pretty good at recognizing things. Computers are pretty bad at recognizing things if they're not the same. Just comparing two numbers is a very difficult problem. Two integer numbers, it's very easy, 5 equals 5, right? But if you have two floating point numbers, like 5.00001 and 5.00002, they're two different numbers, and the computer makes only a binary decision there. And this expands to the whole machine learning field essentially. You have to learn features that make these comparisons easier. This counts for the recognition task,

but it also counts for the task of making things look good. If I do a reconstruction, I want to have textures on top of it, and I want to make it appealing, and these things are very difficult for computers to do.

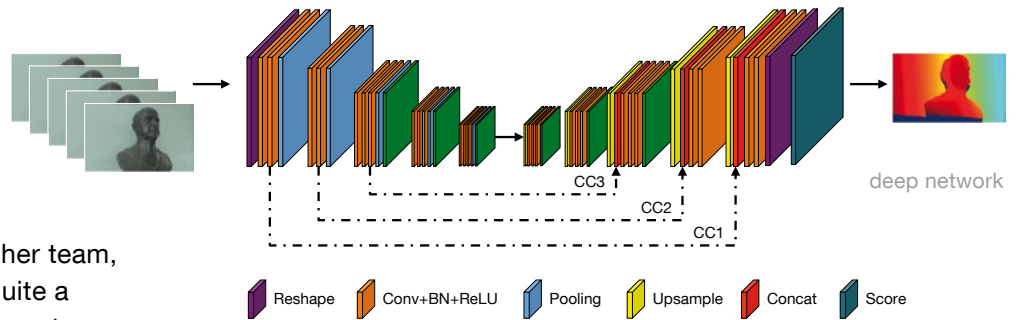
Guibas: Think about this problem of trying to replace a noisy scanned object with a CAD model. There are many, many, many objects in the real world. We don't have a CAD model of everything. You will never find the perfect CAD model. So then the question becomes, how do you adapt the CAD model to the actual data? This is quite tricky, because this adaptation has to be aware of the semantics of the object. Maybe there's a sedan parked on the side of the street. I have a similar sedan, but it's a little too short. I can't simply stretch it, because then the wheels will become ovals. I want to make the body longer, but the wheels should stay round. I have to understand the semantics of what's there. And I can only learn that when I understand many, many models together in a joint structure. There are many subtle problems in extracting this "wisdom of the collection."

Nießner: This brings up another important theme: How do we teach machines these things that Leo just described? With Angel and her team, we have been working for quite a while on how to use the human input to teach the machine to think. How do we know that the sedan is this way, and another way, it's the same car. We do it indoors, we do it for furniture, tables and chairs and things like that. But eventually we have to annotate data, label data, and devise a user interface to do this very efficiently. Do we do this with images and 3D space and so on? How do we get the information from the humans to the machines, and the other way around? There are also projects between Angel and Leo, where they are for example adding the natural language descriptions.

Q: What are some of the strongest links between the groups?

Nießner: Laura and I are pretty much the two people organizing the curriculum in deep learning at TUM right now. Also, we started at more or less the same time, as professors, and our groups are relatively closely connected. Of course there are different topics. Laura is doing a lot of research on localization, video, tracking, and things like that. Daniel has a lot of shared interests with Leo on the geometry processing side. Everybody has a few specialties, but I would say locally we are all very connected. We are all part of an artificial intelligence and vision cluster, as well as a computer vision cluster.

Cremers: To give you another perspective on how important deep networks are becoming, when Laura started teaching the first classes on deep networks here at TUM, the class was a hundred or so people. At this point, after just a few semesters, she has more than a thousand students in her classes on deep networks, master's level classes. We don't actually know where this is heading, but it shows you, even the students sense this need for deep networks that the world has, and Laura is at the center of it all.



Q: This kind of success can be a bit of a burden.

Leal-Taixé: It's not easy. It's an issue, how to balance research and teaching.

Q: I don't suppose you could employ an AI program to take over some of your duties.

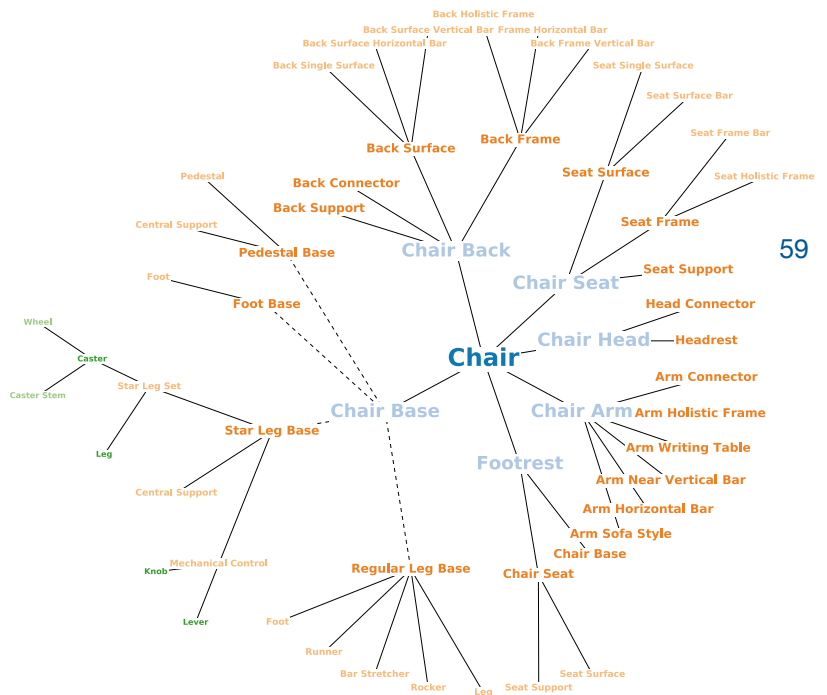
Leal-Taixé: That's not a bad idea.

Guibas: I have had many connections with Michael over the years, even before our association with the TUM-IAS. For example, in 2011 we published a paper called "Shape Google" about shape search, how to find similar shapes. And last year we had a collaboration on how to build robust nets against adversarial attacks, using several kinds of graph convolutions. So our associations and interactions form a connected graph, in multiple ways and with many edges.

Bronstein: I spent some time in Leo's group when I was at Stanford after my PhD, more than ten years ago now, and have collaborated with him and his students, many of whom are faculty members in their turn now. Our most recent collaboration was on using graph neural networks to make convolutional neural networks – the type of deep learning that is used in self-driving cars – more robust to adversarial attacks. It is known, for example, that making a few changes to the traffic signs can confuse the computer vision system that recognizes these signs. I think the actual threat this poses may be a bit exaggerated, but it does showcase potential vulnerabilities of convolutional neural networks, a particularly popular type of deep learning used in computer vision. Last year we published a finding that we can regularize convolutional neural networks with graph-based approaches and make them significantly more robust to these kinds of attacks.



Chair neighborhood



Q: Michael and Matthias, you both have received a lot of attention – from high-profile news coverage to Twitter’s acquisition of the startup Fabula AI – regarding the detection of “fakes” of one kind or another in social media. But your targets and approaches seem completely different. What’s the best way to clarify this?

Nießner: Creating synthetic imagery has been a topic for decades. In computer graphics, creating realistic – photorealistic – imagery from synthetic content is a thing that people have been doing for a long time. Nothing new. It’s just gotten easier. And a lot of what people refer to as “deep fakes” is just face-swapping, where you can take some face and copy that to a different face. What people don’t realize is that most of the time in fact you’re actually getting a hybrid between two people; for example you’re getting a mixture between Trump and Putin out of it. So I ask why do they think this is a problem. And they’re telling me, well, you’re making Trump and Putin “be” each other. And I say no no, you’re creating a person, it’s somebody who doesn’t exist. That person cannot jeopardize democracy, because he doesn’t exist. It’s not a person who can do any harm. The reality is that the deep fake is not really a problem at this point. Most of it is actually pornography or other sketchy areas, not fake news.

Still, you want to provide tools on the detection side, to reliably identify whether this is a real person or a fake. We have a major project, FaceForensics – currently the leading project in the field – that not only covers deep fakes but also a larger variety of facial

manipulations, that for the most part can reliably detect these changes and edits. At the moment detection is much easier than generation. If you know the method, you only need to know that a few pixels are wrong, and then you can detect it. You need to know what you’re looking for. You need to know the methods, but then it’s easy.

Q: And Michael, do I understand correctly that you are detecting fake news entirely without regard to the content? You train neural networks to look at its propagation characteristics?

Bronstein: That’s a special thing about the technology we have developed at Fabula. We have shown that graph propagation features contain important information that allows us to classify whether a piece of news spreading on the social network is fake or not fake. In many cases you really cannot use content, especially when it is language-dependent. We build deep neural networks that allow us to learn on data that has these underlying geometric structures. In the case of graphs, we can apply deep learning to social networks. On Twitter or Facebook, you post something and people interact with this content; they like it, they view it, they repeat or repost it, and then you get a kind of cascade. And by looking at the way in which this information spreads, we have been able to train a classifier – a graph neural network – that allows us to predict with high accuracy, with just a few hours of propagation of a piece of content, whether it was true or fake. That was the technology that we developed into the startup Fabula AI, which was acquired by Twitter last year.

Q: Of all the various ways your research might have an impact on society, what other areas do you think it's most important to highlight?

Leal-Taixé: From my side, I think that if you want to have a society where robots are interacting with people, whether it's robots running around in your home or autonomous cars, these robots need to have a really strong perception. And this is essentially what we are working for. So in the end, giving such a human type of perception to robots is, I think, super challenging, but at the same time at the core of robotic intelligence, and very much a need we can address.

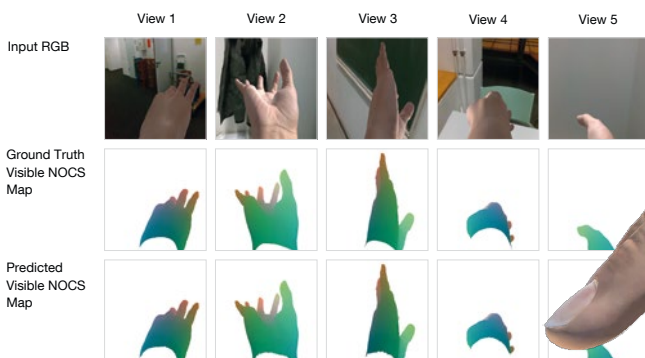
Guibas: There are several different directions in which the work can have an impact. Understanding the 3D environment can be useful not just for robots, like self-driving cars or home robots, but it can be useful to humans, to offer assistance to humans. Maybe there's an elderly person who has difficulty carrying out some task. If a system can understand the environment, understand what the person is doing, and infer the intent, it can offer assistance. And that can mean either providing information or creating visual content that fits that person's environment and makes it clear how they should proceed to complete the task. That's one direction. Providing instruction, education, and assistance, and creating these virtual actions that help people. Another direction is entertainment. Once you can start to freely pull content from the real world into the virtual and from the virtual back to the real, then you can create new experiences for people on top of the real world, or you can create new virtual experiences that use their own objects, their own world. I think both are interesting.

Chang: It's very important to understand that you need to come to it from the point of view of 3D, in the sense that there are these spatial relations that we use all the time: like top and bottom. Even if we're saying the stock market now is "tanking" or "cratering," this also has a geometric interpretation. A lot of these metaphors that we use are related, so I feel that fundamentally for us to have a deeper understanding from the natural language processing side, it's necessary for us to understand how it relates to the real world and to the geometry of things.

» I think that if you want to have a society where robots are interacting with people, whether it's robots running around in your home or autonomous cars, these robots need to have a really strong perception. «

And if we can give a machine – whether it's a robot or just something in the cloud – a better understanding of the physical world that we exist in and how we talk about it, then it is better equipped to meet some real-world need. We may not be so far away from being able to tell a robot, "Bring me the chair from the living room," or "Get me my coffee." Or, once we have this space where we are virtually interacting with each other, if I think of something I need that's not currently in the room, I can ask for it. Or if you want, maybe a virtual assistant or agent can predict what we're going to need before we ever ask for it, just by listening to our conversation.

Nießner: I agree that robotics is very much a consumer of the research that we're doing, self-driving cars that can get you from A to B. But I think the ultimate goal will be virtual environments, so you don't have to go from A to B any more. Right now everything is shut down because of the coronavirus, and there are limitations on how we can communicate. A longer-term question is how do we work in the future, how do we socially interact in the future, how do we communicate in the future. And this goes from entertainment to workspaces alike. There will be language barriers. How can we translate languages automatically? How can we integrate natural language processing into it? And the video can be adapted toward the specific target, whether that means telemedicine or repairing a machine. You may not have all the expertise where you need it, but you can remotely communicate. Having this combination of the real world and a virtual world requires, first, a fundamental understanding of



the real world; and the second thing you want is to be able to connect people in different parts of the world in virtual environments – a combined, mixed reality.

Cremers: One way what we develop becomes really of use to society, and gets used, is through technology transfer, creating startups and bringing things into the market. I've been involved personally in a number of startups, most recently one called Artisense, where we are developing technologies for self-driving cars, autonomous cars, and driver assistance. These are technologies where we leverage cameras to do 3D perception. For me, Michael Bronstein is a great inspiration. He just keeps creating one startup after the other, and he addresses lots of open challenges in society. It seems like he's almost driven to solve important problems for humanity, from fake news detection to predicting protein structure and function, or even trying to identify cures for Covid-19 by using intelligent algorithms. So it's not only that he has these ideas, but he actually brings them to life and makes them happen. This is one way to make sure that what we do actually affects and helps humanity.

You can solve lots of societally relevant problems with this power of the deep networks. In this context, we are in the process of setting up a big institute called the Munich Data Science Institute, and one of the ambitions we have is to bring the success of machine learning, and in particular deep networks, to all areas of data analysis: physics, chemistry, materials sciences, earth observation. There are so many areas that will profit enormously from this kind of transfer of knowledge.

Q: How would you describe the role the TUM-IAS plays in facilitating and enhancing your collaborative research?

Nießner: First, you basically want to get people together who have some shared interests but also add different expertise. And in our case that's exactly what's happening. Angel is a very good example. At TUM right now, we don't have any natural language processing expertise, so without her, we couldn't do research in that area. Same thing goes for Leo's expertise. He's probably the top expert in 3D geometry around the whole globe. Without the TUM-IAS, we would not be able to do these kinds of collaborations. The second part is help with the funding. It enables us to co-advise doctoral candidates.





Guibas: There are a lot of really excellent people at TUM, and it's always a pleasure to visit and spend time and interact with them. And as Matthias said, I think a lot of the real action happens with the PhD candidates we co-advise and engage with in joint projects. Ultimately this builds more permanent bonds, at a more fundamental level.

Chang: I just want to agree that it's a wonderful opportunity. I've also worked with Matthias's and Leo's PhD candidates quite a bit. And the funding for travel, which we hope will eventually be possible again, gives students as well as professors a chance to spend significant time at their collaborators' home institution.

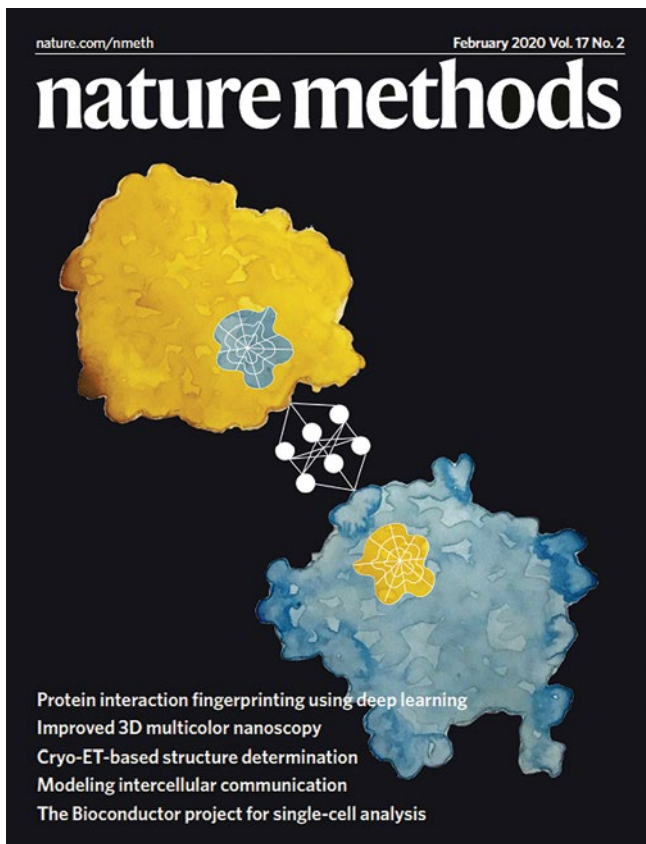
Leal-Taixé: I would add that, as a scientific hub, the TUM-IAS provides an interesting network. I've met many tenure track professors like myself who work in other fields and actually get super excited when I tell them what we can do with neural networks, because they are not really aware of these techniques.

And now we have even started collaborations with their students to bring our knowledge to their problems. They can potentially be solved using deep learning, it's just that they don't know it. So it's exciting to work on these very different topics.

Q: For example?

Leal-Taixé: One is in astrophysics. It has to do with gravitational lensing. Another project is concerned with mapping brain signals to images – essentially trying to predict the behavior of neurons when you show a subject a very specific object or a very specific person. So apparently there are these conceptual neurons – for example, there could be a neuron that is dedicated to firing when a picture of Maradona appears on the screen. We want to help them find out how these neurons are formed.

Cremers: The great thing about the TUM-IAS is that it is a platform that brings together talented people who



Cover illustration by Laura Persat.

Reproduced with permission of Springer Nature.

For example, we invited three protein experts, one from Harvard doing protein folding and two from EPFL doing protein engineering. They publish their work in biological journals and normally do not intersect with our community. One direct result from this meeting was our paper on designing synthetic proteins.

The problems in protein science are very geometric. You can think of protein molecules as surfaces that have to fit together like pieces of a 3D puzzle. It is somewhat more complicated than simple geometric complementarity, because there are electrostatic forces and chemical phenomena involved. There are multiple classical problems in protein science that can be addressed, or can be improved, when you think of them in geometric terms. One is protein folding: how a long one-dimensional chain of amino acids can fold into a complex 3D structure. Another is protein binding, basically how these proteins stick together. Understanding how proteins interact is fundamental to a lot of biological processes in every form of life that we currently know, and also crucial for the design of future drugs.

So as a result of this collaboration, we had a paper that appeared in February 2020 on the cover of *Nature Methods*, on the use of geometric deep learning to gain insight into protein interactions. It is the first paper in such a high-profile journal that I am aware of to have “geometric deep learning” in its title.

Q: And that can be traced straight back to the conference you held at the TUM-IAS?

Bronstein: That is correct.

share common interests and offers a forum to exchange ideas, to discuss, to do workshops together, to collaborate together. The TUM-IAS provides support for organizing workshops, for bringing top people from all over the world to Munich. A perfect example was the TUM-IAS Workshop on Machine Learning for 3D Understanding in the summer of 2018.

Bronstein: Leo was one of the organizers of the workshop, together with me and Daniel, and Lourdes Agapito from University College London, who is also the co-founder of the startup company Synthesia together with Matthias. It’s a small community.

Everyone knows each other. We’re connected in many ways. We wanted to break from our usual circle, by bringing together not only people from geometry, machine learning, computer vision, and graphics, but also from other communities like genetics and protein science. This is not something you would usually see in a machine learning or computer vision conference.