Computer Vision News

The magazine of the algorithm community

A publication by

RSIP VISION

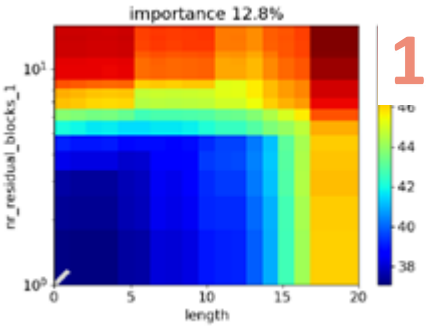
November 2018

Exclusive Interview with Yann LeCun

Contents

04

Guest
Yann LeCun - Facebook



14

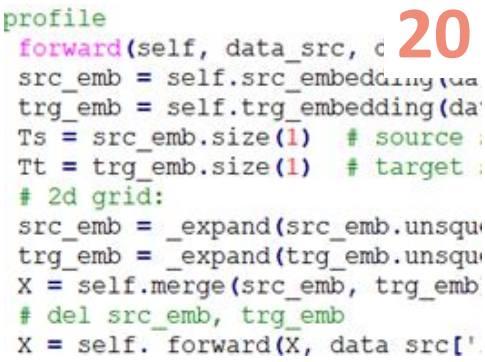
Research Paper Review

18

Challenge
IDG-DREAM

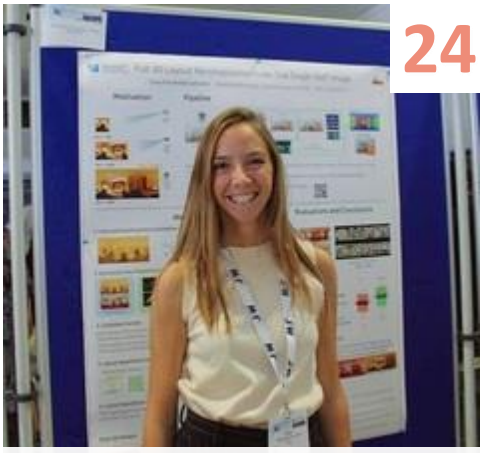
12

Project by RSIP Vision
Distracted Driver Detection



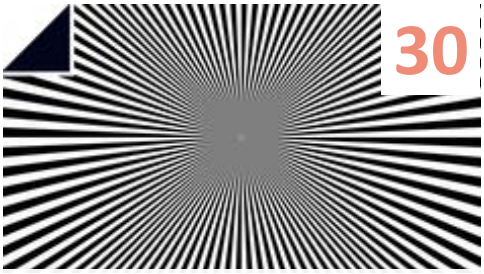
20

Focus on:
RNN and LSTM networks



24

Women in Computer Vision
Clara Fernández Labrador



30

Spotlight News

31

Upcoming Events

What's next for RNN? ➔

Did you subscribe to Computer Vision News? It's free, click here!

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editor:
Assaf Spanier

Publisher:
RSIP Vision

Contact us

Give us feedback

Free subscription

Read previous magazines

Follow us:

"One thing I like to do, and perhaps it's something that I do quite well, is try to really get to the core of what is the central problem behind a question!"

Dear reader,

The above quote is one of the many high points of the exclusive interview that **Computer Vision News** had the chance to conduct with **Yann LeCun**, **Facebook AI**'s charismatic leader. Besides learning about his past work, you will find out his own scientific ambitions for the next five years and also discover the vision behind current AI research at **FAIR (Facebook AI Research)**. As the interviewer, talking to Yann has been a fascinating experience; it will also be a fascinating read for you: please share it with your colleagues and friends!

Again, **RSIP Vision** shares with our readers its cutting-edge work behind one very recent **Deep Learning** project: this time you will learn about **distracted driving detection** obtained through computer vision and neural networks. It's one more project by RSIP Vision in the automotive arena and another life-saving success for our company. Read it on page 12.

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

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

Yann LeCun is Vice President and Chief AI Scientist at Facebook. He is also a Professor of Computer Science and Data Science at New York University.

"I was fascinated by the general concept of intelligence!"

Yann, you have had a major impact on the artificial intelligence field in the last couple of decades. Can you tell us how you got to where you are today?

I've been interested in AI since I was very little. I was fascinated by the general concept of intelligence. Not just machine intelligence, but intelligence in general. I always thought that learning was an essential part of intelligence. I studied electrical engineering and discovered during my studies that people had been working on learning machines back in the '60s and '50s. I learned that a little bit by accident and started reading all the literature about it when I was still an undergrad. It got me thinking that I want to do research on this, so I did a couple of projects when I was at school, and then decided that I should study it.

I found a bunch of people in France who were working on something called Automata Networks that were vaguely related. We're talking about 1983 or so. Nobody at that time in computer

"I read the paper by John Hopfield on Hopfield nets, and discovered the existence of Geoff Hinton and Terry Sejnowski"

science was working on neural nets or even simple machine learning. There was a little bit of machine learning in the context of AI, but a very, very small community. I met some people who had thought about the question of emerging properties of networks composed of lots of simple, connected elements, which is really what neural nets are. I got in touch with them and discovered that there was an international community that was starting to work on neural nets. I read the paper by John Hopfield on Hopfield nets, and discovered the existence of Geoff Hinton and Terry Sejnowski. They had just published a paper on Boltzmann machines, which I thought was wonderful and I really wanted to meet them.

Then I started my PhD. My advisor didn't know anything about neural nets and said: *"I can sign the papers, you seem smart enough, but I can't really help you from a technical point of view."* I had a scholarship through ESIEE, my engineering alma mater in Paris. Eventually I discovered a version of backprop on my own around 1984 and ended up meeting Terry Sejnowski and Geoff Hinton at conferences in France in 1985. At one of these conferences, I met Larry Jackel and John Denker from Bell Labs who eventually were going to hire me. I did a postdoc with Geoff Hinton in Toronto in 1987/88. Larry Jackel had started a group at Bell Labs on neural net hardware and hired me right after my postdoc.

Let's get to the talent part of how you got here. I interviewed Yoshua Bengio and he told me: *"I did not succeed because I am cleverer than the others, but because I know how to focus very well."* Would you agree with that sentence? If that is his secret, what is yours?

I surround myself by people who are smarter than me, so I certainly don't see myself as particularly talented in many areas. I'm much more impressed by other people. For example, I have a very long-term interaction and collaboration with Léon Bottou. He is a well-known figure in machine learning and is better than me in almost every way! [*he laughs*] One thing I like to do, and perhaps it's something that I do quite well, is try to really get to the core of what is the central problem behind a question. How do we get machines to learn? Things like that. Sort of general orientations and intuitions about what are the important problems. Simplifying questions down to their core.

Sometimes an idea or concept seems very complicated because of all the complicated maths you have to use, but at the core there is usually a very simple idea. I don't want to compare myself to Richard Feynman, but his way of thinking was very similar. In the sense that it's trying to ask the elementary questions and to reduce everything to the simplest questions you can imagine. I'm not at all putting myself in the same league.

As for Yoshua, Yoshua is very disciplined and organised. I'm not. I'm very messy.

Well, you're French!

Guest

With Yoshua Bengio (left), Prime Minister Justin Trudeau and Joëlle Pineau (center)

Guest

Well, Yoshua was born in France too! But I'm not a great theoretician. Contrary to my friend Léon Bottou, who is fantastic at maths and things like that. I'm okay at implementing things and making them work, but I'm not spectacular. People have different skills.

Well, you're French and I'm Italian. Some say that we are not too ordered!

[*both laugh*] I don't know, the French tend to be very Cartesian, but I'm not Cartesian at all. I'm more intuitive.

"I learnt English by reading papers written by Japanese people, so my English at the time was horrible!"

Have there been mentors or teachers who have influenced you? Or were you mostly self-taught?

There have been a bunch of people who have had a big influence on me. The maths professors at my engineering school were very supportive. I did a number of projects with them on topics that they were not experts in, but they let me explore for myself, so that gave me the taste for research really. They were very nice. That was when I started working on neural nets. I didn't know how to do research. I had to figure out for myself how it was done. I learnt English by reading papers written by Japanese people, so my English at the time was horrible! [*he laughs*] Most of the research on neural nets in the late '70s, early '80s was from Japan, because people in the West had completely abandoned the field. That's the papers that we were reading from. Then there are, of course, figures of science that I was interested by. Certainly, early on I read a lot about Einstein. I was very much into physics

also. Later, my reading matter was people like Geoff Hinton. I worked with Geoff as a postdoc and discovered we had a lot of common interest.

"We don't have a monopoly on good ideas, despite the fact that we're hiring a lot of the top people!"

Let's talk about Facebook. You are hiring a lot of the best software talent that is available today in the artificial intelligence world. I've interviewed many of them and there are many more who are working for Facebook now and for some of the other major corporations. What is funny is that most of them don't work in the core business of Facebook, but in something that is important for you at the mid to long-term. Without asking you anything confidential, what can you tell us about this?

We don't have anything that's confidential. Well, not many things are confidential, because Facebook AI Research is the fundamental research lab in AI at Facebook and it's outward facing. It's very much connected with the research community. We publish everything we do. We distribute a lot of our code in open source. We collaborate with universities. We have interns and resident PhD students in France and the US. It's very open. This is good in general, and of course, it will be good for Facebook in the long-term, because the main limitation of AI today is not whether Facebook is ahead of Google or Microsoft or IBM, it's rather that the field itself is not where we want it to be. If you want to build intelligent virtual assistants, for example, then the science or the technology to build intelligent machines that have a bit of common sense to interact with humans does not exist. So our objective is to develop the techniques that will make that product a reality. We don't have a monopoly on good ideas, despite the fact that we're hiring a lot of the top people. So, we have to interact with the broader research community. That's why we are open.

There's also another set of organisations within Facebook, part of the broader "Facebook AI" organisation, which are much more focused on problems related to Facebook – computer vision, natural language processing, search, things like that. A lot of those groups use technologies that were originally developed at FAIR, maybe for a different purpose, and so there is quite a lot of influence there, but these groups have a different modus operandi. They are much more focused on the needs of the company. They publish papers also, but not as much. They are focused on improving the services that Facebook provides or creating new ones. FAIR is working on developing new technology and advancing the field. Sometimes we think the horizon is 3 years or 5 years or 10 years, but sometimes, the things we come up with turn out to be useful right away. It surprises us sometimes.

In your opinion, what is the best computer vision paper of 2018? I'm sure you won't be surprised to hear that our magazine named Mask R-CNN as the best paper of 2017.

I would not disagree either! [*both laugh*] So many things are happening

Guest

Guest

in computer vision. So many things that I can't even follow! I will not pinpoint a particular paper, but I think

"... there is really interesting work going on in the whole area of self-supervised learning!"

there is really interesting work going on in the whole area of self-supervised learning. Either using generative adversarial neural networks or using other techniques, where people are trying to discover high-level concepts in vision – like objects, motion, depth, things like this – without actually exclusively supervising the system. I think this is going somewhere. This is the seed. There is no particular application of this yet, but I think it's the seed of something very big. The next revolution in computer vision. Probably the next revolution in AI, actually. There's something I have been talking about in all my talks for the last three years, which is that the future of AI is in self-supervised learning.

"The next revolution in computer vision. Probably the next revolution in AI ..."

This way you train a machine to learn about how the world works, without training it for a specific task, but then you can train it for a specific task with very, very little data. This is how humans and animals operate. You started seeing a bunch of papers on this from Facebook, Intel, Google, DeepMind and NVIDIA. There's this paper on colorization where you train

a machine to colorize objects in a video and basically it discovers the motion of objects. There's a bunch of things like this which I think are impressive.

I have a question that one of my engineers asked me to ask you. It's a funny question, but I cannot understand it because I am not an engineer myself. He asks, which one is your favorite between ReLU and batch normalization?

Oh okay… [*he laughs*]

So, it is funny?

It's funny! It's a pretty easy answer, but it asks many other questions. I would say ReLU, because it's a simple idea that everybody uses and it's the idea that basically allows us to train relatively deep networks. There's another idea, the ResNet residual connection idea from Kaiming He, that allows us to train even deeper networks. Batch normalization in the mind of many people, including me, is a necessary evil. In the sense that nobody likes it, but it kind of works, so everybody uses it, but everybody is trying to replace it with something else because everybody hates it. There's something about it that is not entirely satisfying. We're all under the impression that there's got to be something better than it. Also, people don't understand why it works and how it works. There are intuitions that we've had about how neural nets converge and learn, and the way batch normalization works doesn't fit in that picture, so there's a lot of work to do there to understand why it works and to try to replace it with something else. Kaiming He also came up with this thing called group normalization, which is meant to replace batch normalization and apparently works slightly better.

"This model of dual affiliation only works if - in the industry in which you work - the industry lab is a research lab, not a development lab, and if it practices open research and the company that runs it is not too possessive about intellectual property!"

Looking ahead, how do you think academia and industry might work better together in our computer vision and artificial intelligence community?

I have spent half my time in academia and half my time in industry in my career. I was initially at Bell Labs – that became AT&T Labs – then I spent 18 months at the NEC Research Institute. I became a professor and now I share my time between industry and academia. I think this idea that you can share your time between industry and academia is good. I actually wrote a piece about it.

I have read it. About the double affiliation.

That's right. There is a very important point that I think some people have missed in this piece, which is that this model of dual affiliation only works if - in the industry in which you work - the industry lab is a research lab, not a development lab, and if it practices open research and the company that runs it is not too possessive about

Guest

"That's probably my big contribution to industry research within the last five years..."

Guest

intellectual property. The whole point of this is the exchange of information between industry and academia so that you can take advantage of the fact that in industry you have engineering support and large computing facilities, and in universities there are students and young people with a lot of creativity whose incentives and modus operandi is different from industry.

It's good to have different motivations for people because they come up with different ideas when they're in different environments. That's good, but it only works if the industry lab you are in practices open research and actually does real research with publications. It doesn't work if the affiliation is with an industry where everything is secret, everything is very applied and engineering-oriented. There's been some responses to my piece and some of them basically blurred the difference between these two kinds of industry research. They said you can't have dual affiliation, because when you work in industry you have to basically work on whatever is useful to the company, and keep some things secrets, and that's incompatible with academia. I agree, it's incompatible, but I disagree that it's impossible to do this in industry. It depends how the industry research is organised.

One of the things I've done in Facebook is organise the research lab in such a way that it is compatible with academic collaborations. That's probably my big contribution to industry research within the last five years. It's something that basically did not exist to the same extent until now. There have been organizations in industry in the past that have been very influential in science, like Bell Labs where I used to work and IBM Research and Microsoft Research,

but they were a little too possessive about IP to really be open in the same way. There was no tradition of open source and things like that. Now, it's different, and this is a new way of doing industry research. I think some of the other companies are being influenced a little bit by us. For example, in the last five years, Google has become much more open about what they do in research than they were in the past. They're still a little secretive, but they're definitely much more open than they used to be.

I have a witness to what you said you are trying to implement at Facebook! Pauline Luc, whom I interviewed last month, told me: *"My lab at Facebook is the same as my lab at the school."*

That's right. I know Pauline's work very well of course, because I participate in her project and co-author the papers. I think the work she's doing is amazing.

"GANs are not very well understood"

Finally, what would you say is the biggest accomplishment you would like to achieve before you retire?

Finding good ways to do self-supervised learning in a generic way. I've been working on computer vision, but I'm not a computer vision person. I don't see myself as a computer vision person. At least not entirely. My interest is really in learning, so I like to find ways to get machines to learn how the world works, by observation. That means learning under the presence of uncertainty. If you give a machine a segment of video and you ask it to predict what's going to happen next, there's many things that can happen. Of all the possible things that can happen, one of them is going to be the thing that actually happens in the video, but there are many possible scenarios that may happen. When you train a machine to predict videos, if you're not careful, it produces a blurry prediction which is sort of an average of all the possible scenarios that can occur. That's a bad prediction.

"That's my goal for the next few years!"

One of the technical problems we're trying to solve is: how do you train a machine in a situation where what you're asking it to predict is not a single thing, but a set of potential things? You can formulate this mathematically as predicting a probability distribution instead of a single point. But we don't know how to represent probability distribution in high-dimensional continuous spaces. I think it's going to be a combination of, again, figuring out what are the essential concepts there, and coming up with simple architectures that are easily understandable and can deal with that problem of representing uncertainty. GAN is a promising approach. But GANs are not very well understood. They converge sometimes, but when they work they work amazingly well. They don't work every time, so we need to find ways to either understand how they work or find other techniques, then use those techniques to get machines to learn as much background knowledge as possible by just observing the world through video, images, etc. Then, once the machine has learned a good model of the world, it will only require a few samples or a few trials to learn any particular task. That's my goal for the next few years...

Every month, **Computer Vision News** reviews a successful project. Our main purpose is to show how diverse image processing techniques contribute to solving technical challenges and physical difficulties. This month we review **RSIP Vision**'s solution to track driver attention: **Distracted Driver Detection with Deep Learning**. RSIP Vision's engineers can assist you in countless application fields.

Project

… RSIP Vision's deep learning algorithms … provide a great solution to a real world problem!

More than **one million deaths** occur in the world every year as a consequence of car accidents. The number of injuries and severe harm is on the order of tens of million per year. It has been found that about one fifth of these accidents are due to **distracted driving**, defined as driving while performing other activities that take the driver's attention away from driving.

Traditional reasons for distraction are somnolence, looking at scenery, heated discussions with passengers of the car and operating various car/radio/CD commands. The **massive availability of smartphones** added new and deadly threats, under the form of talking on the phone, texting and reading social media - while driving. **Talking on a cell phone is said to quadruple the risk of crashing**.

Among the many projects conducted by **RSIP Vision** in the automotive field, today we will discuss the development of a system able to accurately detect when **a driver is entering a state of distraction**. A conveniently placed dashboard camera can detect several indicators of the driver's status and behavior, in particular **gaze** and **movements**. Patterns which differ from normal driving are identified in the images coming from the camera: typical patterns are labelled accordingly in view of training the system.

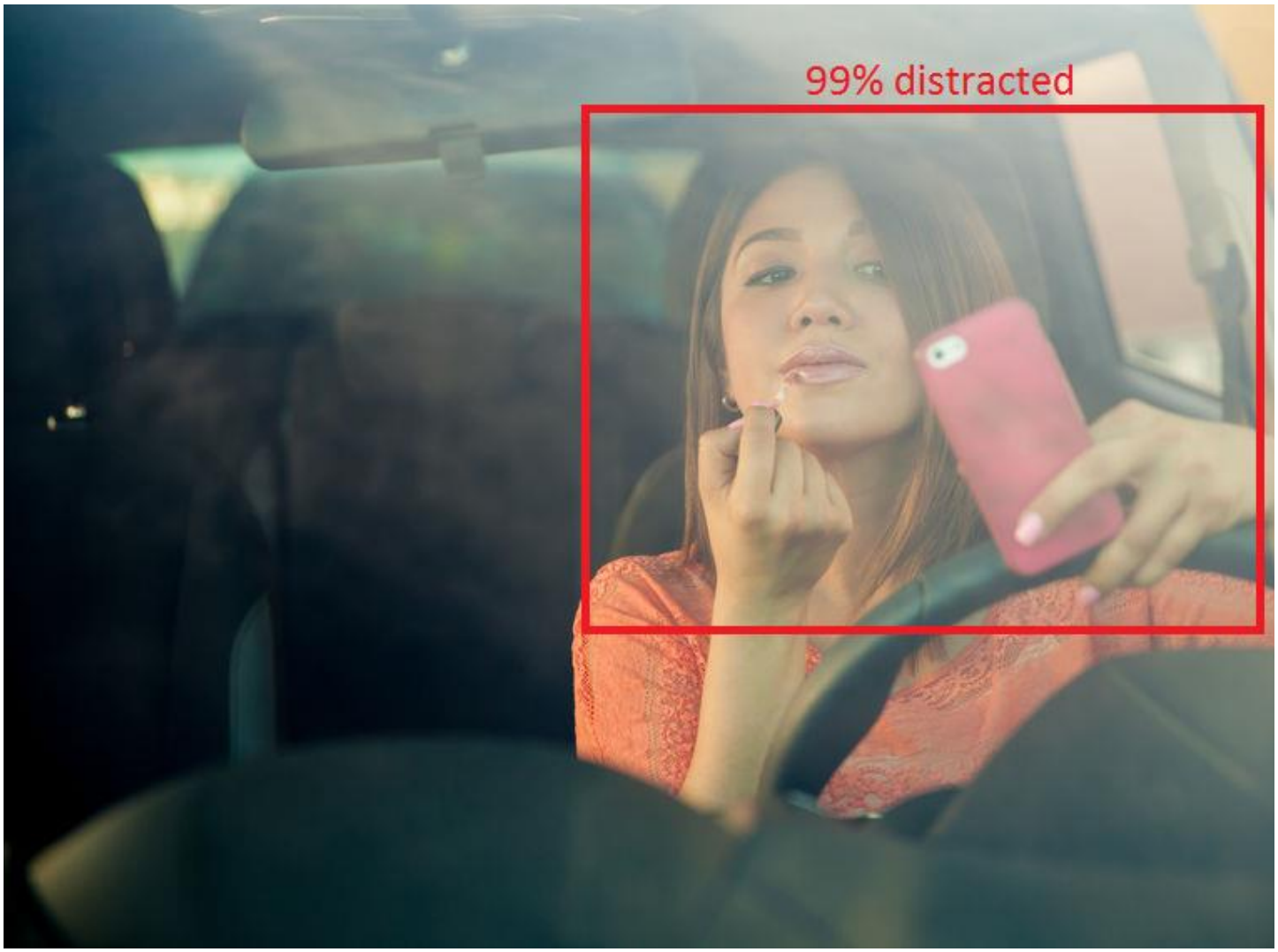
The main challenge of doing this is the **huge variability of situations**: a large set of movements can be made by the driver which are not dangerous; while only those signs which can be considered a clear indicator of fatigue or other distraction should be flagged. This means that **a very large amount of data** needs to be collected for the training.

Take us along for your next Deep Learning project!

Another challenge is due to the need for a very quick output: in order to provide a useful alert, the input collected by the camera must be processed by a local CPU via a relatively simple **deep learning** system; a more complex **neural network** would require a GPU, which would significantly increase the cost of the system. The network must be trained to know the **status of driver attention** at any time and to generate an alert if needed. The algorithmic work involves techniques of **pose detection**, **gaze detection**, **face detection**, **hand position** and **gesture detection** among other advanced computer vision techniques. Only a deep learning network can efficiently integrate the input information coming from the detection algorithms. The camera is placed conveniently on the dashboard and directed towards the driver. We are able to detect all possible scenarios with no need for an additional camera inside the car for validation.

RSIP Vision has already performed a large number of projects in the automotive fields. Do you have a project around **ADAS (Advanced driver-assistance systems)** or **autonomous driving**? Before you start any work on your system, talk to RSIP Vision's experts! This is another example of how RSIP Vision's deep learning algorithms are able to provide a great solution to a real world problem.

Project



by Assaf Spanier

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen **Towards Automated Deep Learning: Efficient Joint Neural Architecture and Hyperparameter Search**. We are indebted to the authors (**Arber Zela**, **Aaron Klein**, **Stefan Falkner** and **Frank Hutter**), for allowing us to use their images to illustrate our review. Their article is here.

… efficient joint neural architecture and hyperparameter search

One of the key questions in the deep neural network field is how to find the optimal architecture among the seemingly infinite possibilities? And especially: what is the optimal number of layers for the network? What is the optimal filter size and number of filters? **Neural Architecture Search (NAS)** is the emerging field of research attempting to deal with and offer some solutions for these issues.

Traditionally, this type of research is conducted in two stages: first, **we find the optimal architecture** for a small number of steps (or using a partially pre-trained network). Then, as a separate stage, given the selected architecture, **you search for the optimal hyperparameters** for training the selected architecture. In this paper the authors demonstrate that a one-stage approach that combines the architecture selection and hyperparameter optimization is preferable to the two-stage approach. Furthermore, the authors demonstrate that the common practice of using training on a small number of steps during the initial NAS stage and a much larger number of training steps for the second stage is inefficient due to a low correlation between the initial training stage and the second longer one.

Novelty:

The paper's innovation and contributions to the field:

1. Combination of Bayesian optimization and Hyperband to perform efficient joint neural architecture and hyperparameter search.
2. Demonstration that the approach of using short-step-number training on an architecture is uncorrelated with the results longer (realistic) training will produce. And demonstration of how the combined one-stage approach deals with this challenge.
3. The authors showed how, given an "*uncompromising*" limit of just 3 hours for training and using one-stage combined architecture and hyperparameter optimization, they achieved competitive results.

The Method and the Experiment:

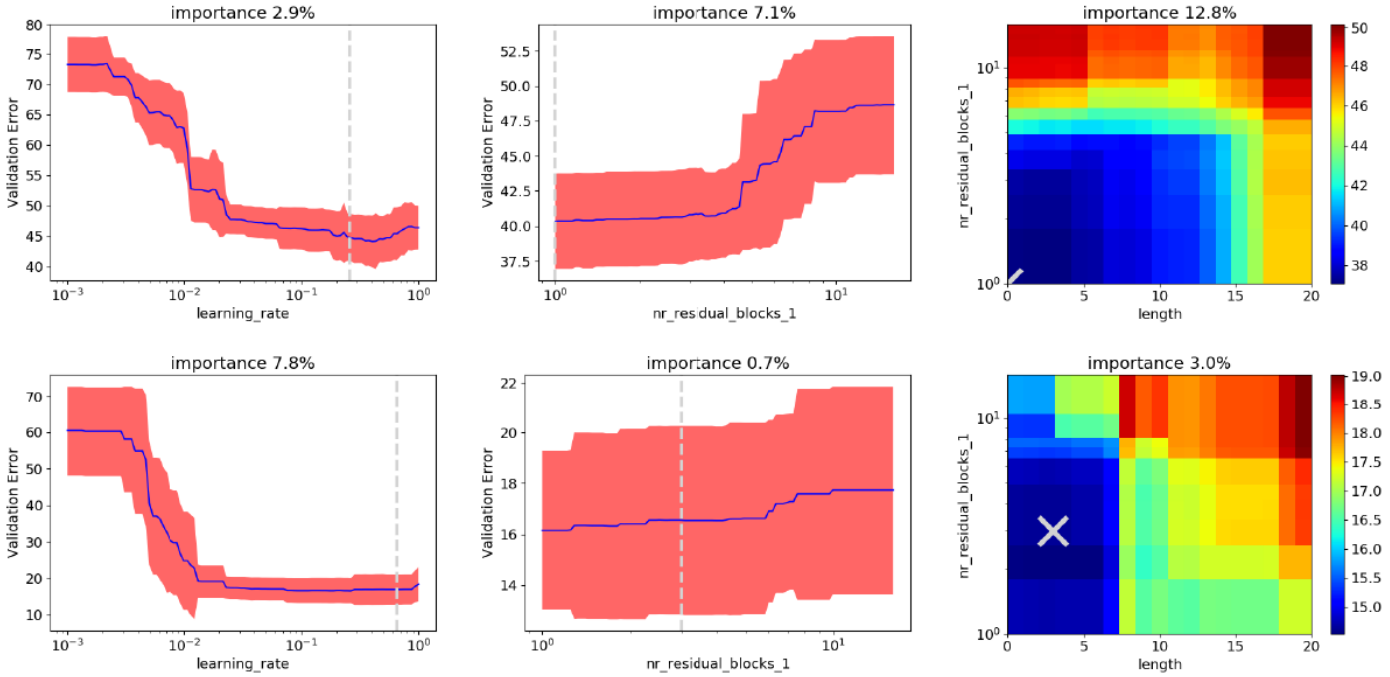
To efficiently optimize in the architectures and hyperparameters joint space, the authors chose to use **BOHB**, a recent combination of **Bayesian Optimization** (with its guarantees of convergence) and **Hyperband searches**. Like hyperband, BOHB uses evaluations with different training-time limits to accelerate optimization. BOHB allows you to run optimizations under different constraints, known as budgets. BOHB uses **kernel density estimators** to select promising candidates instead of sampling new configurations at random.

The baseline network for the experiment is the PreAct ResNet-18 (He et al., 2016) and WideResNet-28-10 (Zagoruyko and Komodakis, 2016). The dataset for the experiment was CIFAR-10. In the optimization phase the CIFAR-10 dataset was split into 3 subsets: 45k data points for training, 5k for validation and 10k for testing. The authors evaluated the following hyperparameters: The parameters of each model are initialized as described by He et al. (2016) and trained using SGD with an Initial Learning Rate of 0.1. Batch Size was 32. L2 regularization was applied with a factor of 10-4 and 5 * 10-4 to 3-branch and 2-branch networks, respectively. Momentum (Nesterov's) was 0.9. MixUp (Hongyi Zhang, 2018) with a value of 0:2. CutOut (DeVries and Taylor, 2017) was applied with a mask length of 16. Use ShakeDrop with alpha parameter set to 0 (Yamada et al., 2018). All networks were trained on one Nvidia GTX 1080Ti GPU. All the models in 2 are trained for 3h with the initial learning rate annealed using a cosine function with T0 = 720s, Tmult = 2 (Loshchilov and Hutter, 2017). So far we are dealing with “classic” hyperparameters, the ones usually handled in the second stage of NAS, given a specific selected architecture -- these are listed in the top part of the table below. The architecture parameters are found in the second part of the table and deal, as you might expect, with the number of blocks, number of filters and filter size:

Hyperparameter	Range	Log-transform	Value
Initial Learning Rate	$[0.001, 1.0]$	yes	0.648188
Batch Size	$[32, 128]$	yes	89
L_2 regularization	$[0.00001, 0.001]$	yes	0.000339
Momentum	$[0.001, 0.99]$	no	0.099601
MixUp α	$[0.0, 1.0]$	no	0.492042
CutOut *length*	$[0, 20]$	no	3
ShakeDrop *death rate*	$[0.0, 1.0]$	no	0.038439
$ResBlocks_1$	$[1, 16]$	yes	3
$ResBlocks_2$	$[1, 16]$	yes	4
$ResBlocks_3$	$[1, 16]$	yes	2
$ResBranches_1$	$[1, 5]$	no	1
$ResBranches_2$	$[1, 5]$	no	1
$ResBranches_3$	$[1, 5]$	no	4
$Filters_0$	$[8, 32]$	yes	16
$WidenFactor_1$	$[0.5, 8.0]$	yes	6.241141
$WidenFactor_2$	$[0.5, 8.0]$	yes	1.388867
$WidenFactor_3$	$[0.5, 8.0]$	yes	3.344766

Results:

Importance indicates the ratio of the variance the individual choice(s) explain. The dashed gray line indicates the value of the best found configuration:



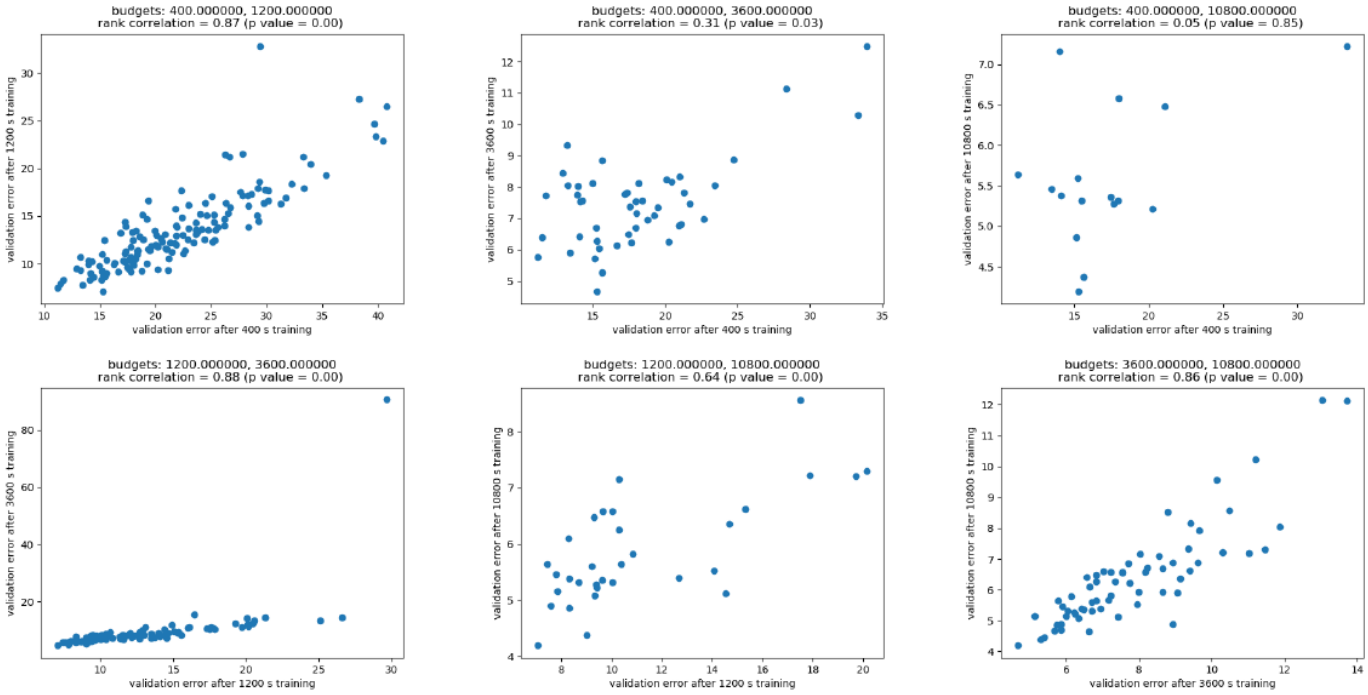
Parameter importance plots for three hyperparameters:

The top row are results after 400 sec of training and the bottom row results are after 3600 sec (1 hour). The columns are: left column -- evaluation of different learning rates; middle column -- different residual block numbers; right column -- number of residual blocks plotted against the **CutOut** length (a kind of regularizer that, as the name suggests, “cuts out” different portions of the image).

Correlations between the different training time limits (budgets).

You can clearly see that the correlation between similar lengths of training time is very high. Correlation between adjacent budgets is very high. In the top right-hand graph you can see the significant lack of correlation in error rate, with a 27-fold difference between 400 sec and 10,800 sec (3 hours) of training. In the bottom left-hand graph, on the other hand, you can see the high correlation between 1200 and 3600 sec of training -- only a 3-fold difference. Hence, there is no correlation in error rate between the shortest and longest training times.

To summarize, the correlation in error rate decreases as the difference in length of training time increases, making the short training time uninformative about the best configurations for the longest training time:



Spearman rank correlation coefficients of the validation errors between different training times:

	1200s	1h	3h
400s	0.87	0.31	0.05
1200s		0.88	0.64
1h			0.86

A performance comparison between manually constructed architectures and the network architecture found by BOHB with a training time limit of 3 hours. All networks used the same hyperparameter space for optimization. You can see that by using simultaneous architecture and hyperparameter optimization the authors achieved competitive results:

Network	**Params**	**Test error (%)**
ResNet-18	11.2M	3.34 ± 0.11
Shake-Shake 26 2x32d	2.9M	3.91 ± 0.09
Shake-Shake 26 2x64d	11.7M	3.38 ± 0.07
Shake-Shake 26 2x96d	26.2M	4.22 ± 0.06
Ours	27.6M	$\mathbf{3.18 \pm 0.16}$

Conclusion:

In this paper, the authors showed the advantages of simultaneously jointly optimizing architecture and hyperparameters. They demonstrated the powerful effect -- in term of training time -- of hyperparameter selection coupled with architecture selection. The authors demonstrated that this challenge can be solved with BOHB optimization method.

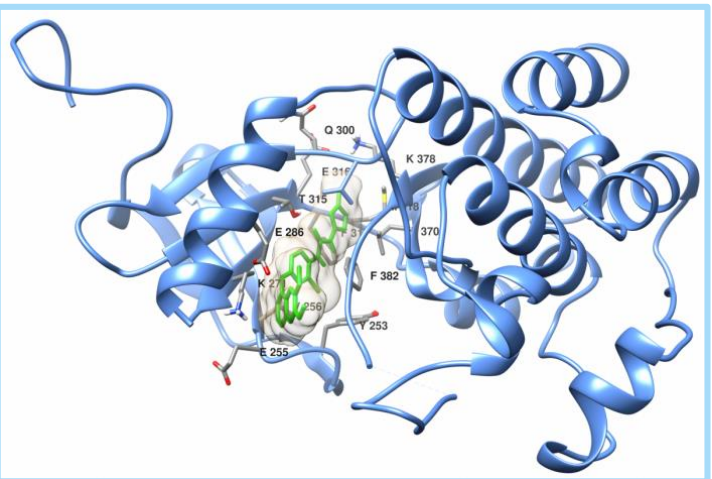
With Tero Aittokallio

The IDG-DREAM Drug Kinase Binding Prediction Challenge is launched to benchmark machine learning models in the effort to accelerate mapping of drug-target space by prioritizing most potent compound-target interactions for further experimental evaluation.

DREAM has launched the IDG-DREAM Drug Kinase Binding Prediction Challenge. The challenge is currently open and it will run until December 2018. Winners will present their winning model at the **DREAM 2019 Conference**.

The goal of mapping the bioactivity of the full space of compound-target interactions is probably unattainable, even using the most powerful hardware and software technologies: simply put, the size of the **chemical universe** is excessively large.

However, it is possible to prioritize most potent interactions for further experimental evaluation. Due to their clinical importance, the organizers of this **IDG-DREAM challenge** have decided to focus on **kinase inhibitors**. The objective is to further extend the druggability of the **human kinome space**.



Visualisation of the binding between tivozanib and ABL1

This is to be achieved by evaluating the practical power of statistical and **machine learning models** to predict drug-protein binding affinity and allow to prioritize the most powerful interactions.

This research requires a multi-disciplinary approach testing the boundaries and pushing the limits of both machine learning and drug discovery.

The project is consequent to a 2017 study by then PhD student **Anna Cichonska** et al.: **Computational-experimental approach to drug-target interaction mapping: A case study on kinase inhibitors** (Anna has successfully defended her dissertation only a few weeks ago).

Sage Bionetworks

"...a continuously-updated resource of predictive models and experimental data for the chemical biology community..."

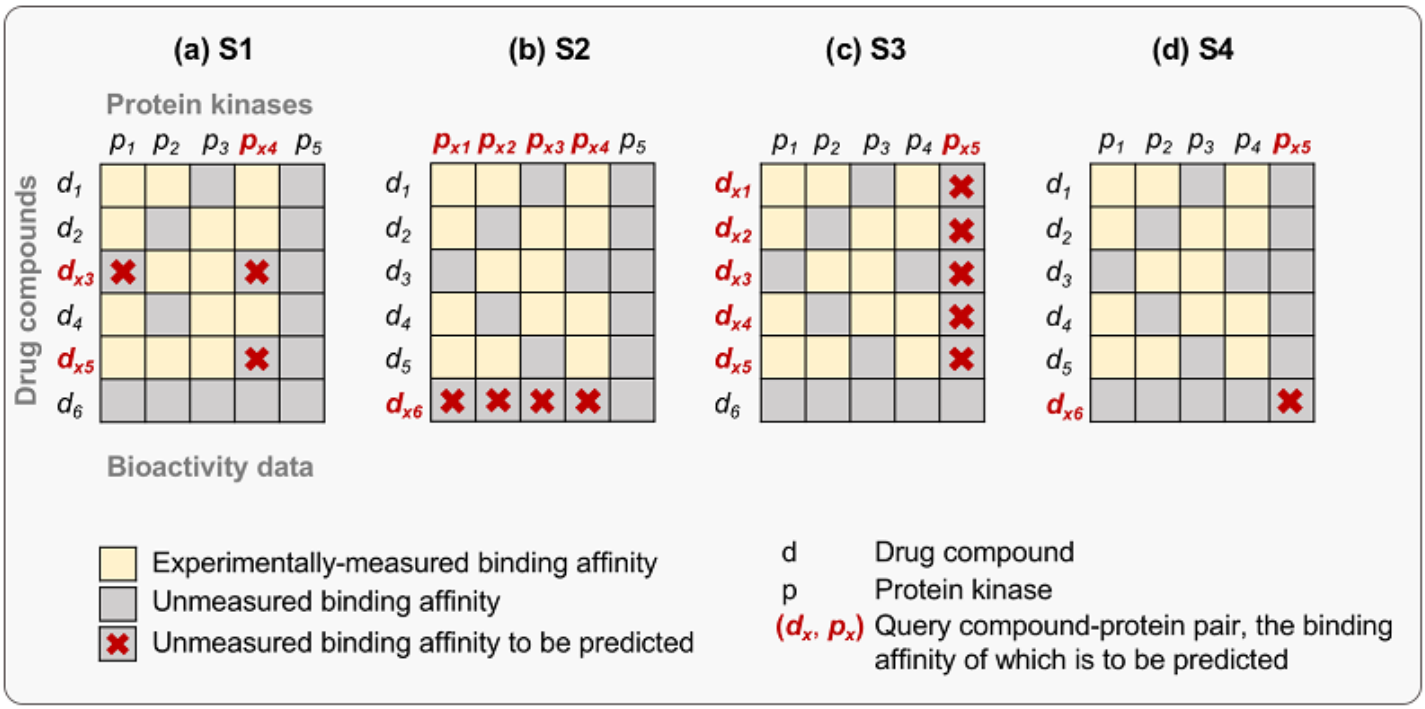
We asked **Tero Aittokallio**, Professor of Statistics and Applied Mathematics at the Department of Mathematics and Statistics of the **University of Turku (Finland)** and **FIMM-EMBL Group Leader** to tell us more about DREAM and about the challenge. Here is the answer that Tero, co-organizer of the challenge, gave us:

*"If your readers are not familiar with the **DREAM challenges**, here is a general introduction: DREAM is a non-profit organization that sponsors **scientific competitions**, so-called crowdsourced challenges, which pose fundamental questions about systems biology and translational medicine. The Challenges are designed and run by a community of researchers from a variety of organizations, and the solutions (e.g., machine learning models) are **tested by the organizers using independent data**. This allows individuals and teams to collaborate openly so that the "wisdom of the crowd" provides the greatest impact on science and human health."*

When asked about the vision behind the challenge, Tero replied:

"We envision that the IDG-DREAM Challenge will provide a continuously-updated resource of predictive models and experimental data for the chemical biology community to prioritize and experimentally test new target selectivities toward accelerating exciting drug discovery and repurposing applications."

The challenge page is here. Good luck!



Anna Cichonska

by Assaf Spanier

"What does the future hold for RNN and LSTM networks?"

Back in 2014, we all fell in love with **RNN** and **LSTM** networks and their various implementations: their potential seemed unbounded. What does their future look like? Is this the beginning of their decline? In 2014 RNN and LSTM came back from the dead, achieving impressive results on various tasks, thanks in large part to LSTM, as described in our overview of the field. For several years they were the go-to solution for difficult problems like seq2seq translation, and achieved impressive results in text-to-speech tasks.

However, two years later, technologies like **ResNet** and **attention** were proposed. The research community realized that LSTM is in fact a type of residual connection. The **attention mechanism** became one of the most crucial components of neural networks when dealing with sequences, whether they are sequences of video, images, audio, text or any other type of data.

Let's try and understand the attention mechanism in a nutshell (using a simple Keras code sample). The attention mechanism differentially weighs different parts of the input to the network in order to direct the network element, or nudge it, towards the most relevant parts of the input for classification, translation etc. (the attention weights will differ depending on the task).

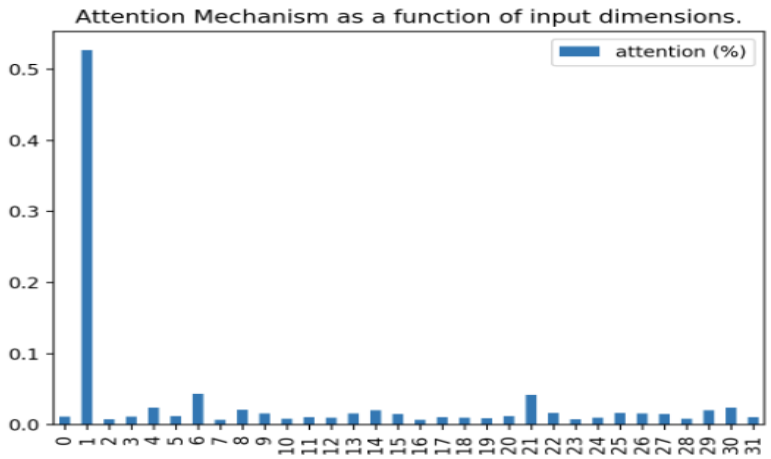
To get a better idea of what we're talking about, let's take a look at the following code snippet, which demonstrates an attention mechanism for a very simple network. The simple attention mechanism is implemented using a Dense layer (first line of code), the input is multiplied by the output of the Dense layer -- and the product of this multiplication is the attention mechanism.

```
def build_model():
    inputs = Input(shape=(input_dim,))

    # ATTENTION PART STARTS HERE
    attention_probs =
        Dense(input_dim, activation='softmax',
                         name='attention_vec')(inputs)
    attention_mul = merge([inputs, attention_probs],
                      output_shape=32, name='attention_mul',
                                       mode='mul')
    # ATTENTION PART FINISHES HERE

    attention_mul = Dense(64)(attention_mul)
    output = Dense(1, activation='sigmoid')(attention_mul)
    model = Model(input=[inputs], output=output)
    return model
```

Now, let's look at an example with data. For purposes of the demonstration a synthetic dataset of vectors of length 32 was prepared, where term 1 of each input vector is purposely determined to be equal to the label that vector should receive as the resulting output label:

```
input_dim = 32
attention_column=1
x = np.random.standard_normal(size=(n, input_dim))
y = np.random.randint(low=0, high=2, size=(n, 1))
x[:, attention_column] = y[:, 0]
```

After training the network, if we look at the weights vector output of the attention layer in this simplified case (see figure below), we shall see that the network indeed gives a much higher weight to term 1, as we would expect, given that we set this term to be identical to the label fitting the vector's data. That is, we see that the attention mechanism was successful -- focused the network's attention on the most relevant part of the input to help successfully classify the vector -- term 1.



That was an extremely simple case; a somewhat more useful case is integrating the attention mechanism into an LSTM network. Just like the following code demonstrates, implementing an LSTM network which includes a simple attention mechanism, effective for understanding how it works.

```
def attention_3d_block(inputs):
    # inputs.shape = (batch_size, time_steps, input_dim)
    input_dim = int(inputs.shape[2])
    a = Permute((2, 1))(inputs)
    a = Reshape((input_dim, TIME_STEPS))(a)
    a = Dense(TIME_STEPS, activation='softmax')(a)

    a_probs = Permute((2, 1), name='attention_vec')(a)
    output_attention_mul = merge([inputs, a_probs],
                                 name='attention_mul', mode='mul')
    return output_attention_mul


def model_attention():
    inputs = Input(shape=(TIME_STEPS, INPUT_DIM,))
    lstm_units = 32
    lstm_out = LSTM(lstm_units, return_sequences=True)(inputs)
    attention_mul = attention_3d_block(lstm_out)
    attention_mul = Flatten()(attention_mul)
    output = Dense(1, activation='sigmoid')(attention_mul)
    model = Model(input=[inputs], output=output)
    return model
```

We can see the LSTM layer included inside the `model_attention` function, immediately followed by the attention mechanism (`attention_3d_block` function), which, just as in the previous example, weighs the terms of the input by their contribution to successful prediction. The attention layer is again implemented using Dense, whose output is multiplied by the input received by the layer - giving a different degree of focus to different terms of the input. Note the need to appropriately permute the data, so the weights of the attention mechanism affect the terms of input vectors, not differently weighing the same channel for different samples.

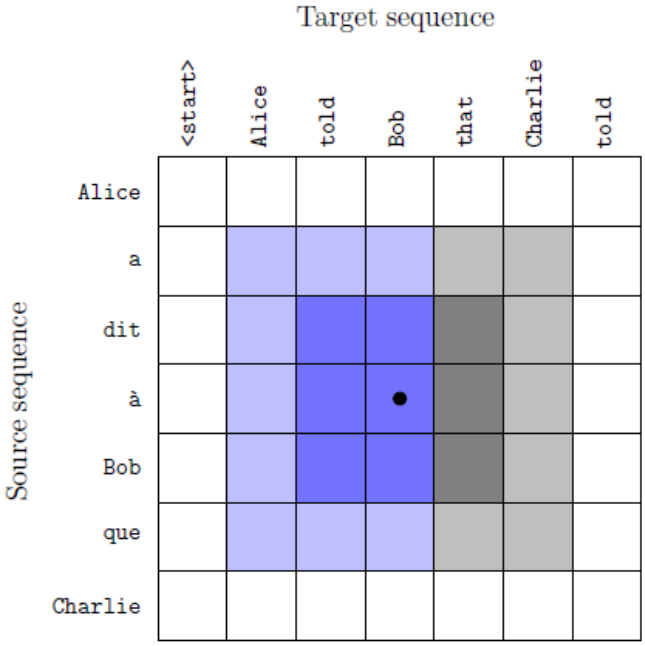
However, the LSTM with attention model is not without drawbacks and difficulties. First, there is still the vanish gradient problem for long sequences. Second, there is a computational difficulty.

LSTM requires 4 linear layers (MLP layers) per cell and for each sequence time-step. Linear layer computation is very memory bandwidth intensive, incapable of being run on many computation units because of insufficient memory. Memory-bandwidth-bound computation is one of the major challenges for hardware designers, ultimately limiting the usefulness of neural networks.

A somewhat more sophisticated attention mechanism is the Transformer, which Computer Vision News reviewed last year.

A brand new method, which might mark the end of LSTM networks is the new method described in the paper "**Pervasive Attention: 2D Convolutional Neural Networks for Sequence-to-Sequence Prediction**". In this method the authors use a CNN network traditionally used for image classification. However, in this paper the authors represent the data as "images" in a unique way: given a training pair of source and target sentence fragments (s, t) with lengths |s| and |t|, the model embeds them as {x1, . . . , x|s|} and {y1, . . . , y|t|} and concatenates them into an "image" where each "pixel" is a concatenation of the two embeddings --

the depth of the "pixel" is equal to the sum of the source and target embeddings. In this setting the label output for the image is the predicted next word of the target sequence. Below is the illustration of such an "image", trained for translating a sentence from French into English -- we can see the source on the left and the target at the top.



The code below is an implementation of the method described above: the first lines embed the target and source sequences and merge them to form the "image" (denoted as `X` below). Then, the function call the `_forward` function, which implements the CNN (DenseNet), with a special aggregation function, which projects the 2D data into a one-dimensional vector. (The authors evaluated several different aggregation methods, such as Max-pooling, Average-pooling and attention, with results detailed in the paper.) This one-dimensional vector is fed into softmax function to predict the next word.

This method outperformed the state of the art methods on the IWSLT German-English translation dataset. This initial article, with such promising results, seems to indicate a revolutionary direction for the sequence to sequence area. Detailed comparisons of different setting parameters, such as embedding size, network depth, and others, can be found in the original paper; the source code is here.

```
# @profile
def forward(self, data_src, data_trg):
    src_emb = self.src_embedding(data_src)
    trg_emb = self.trg_embedding(data_trg)
    Ts = src_emb.size(1)  # source sequence length
    Tt = trg_emb.size(1)  # target sequence length
    # 2d grid:
    src_emb = _expand(src_emb.unsqueeze(1), 1, Tt)
    trg_emb = _expand(trg_emb.unsqueeze(2), 2, Ts)
    X = self.merge(src_emb, trg_emb)
    # del src_emb, trg_emb
    X = self._forward(X, data_src['lengths'])
    logits = F.log_softmax(
        self.prediction(self.prediction_dropout(X)), dim=2)
    return logits

# @profile
def _forward(self, X, src_lengths=None, track=False):
    X = X.permute(0, 3, 1, 2)
    X = self.net(X)
    if track:
        X, attn = self.aggregator(X, src_lengths, track=True)
        return X, attn
    X = self.aggregator(X, src_lengths, track=track)
    return X
```

Clara Fernández

Clara Fernández is a second year PhD Student in Computer Vision and Robotics under the guidance of Prof. J.J. Guerrero at the Computer Science and Systems Engineering Department of the University of Zaragoza (Spain) and Prof. Cédric Demonceaux at the Lei2 of the University of Burgundy (France). More interviews here.

Clara, can you tell us about your work?

I am beginning my second year as a PhD student at the University of Zaragoza and at the University of Burgundy in France. It's like a double PhD, with a Spaniard supervisor and a French supervisor. I have to spend the first year-and-a-half in Spain and the second year-and-a-half in France.

You study in Spanish and in French?

Everything in English. I just started learning French. More or less, I can understand everything, but I feel more comfortable in English.

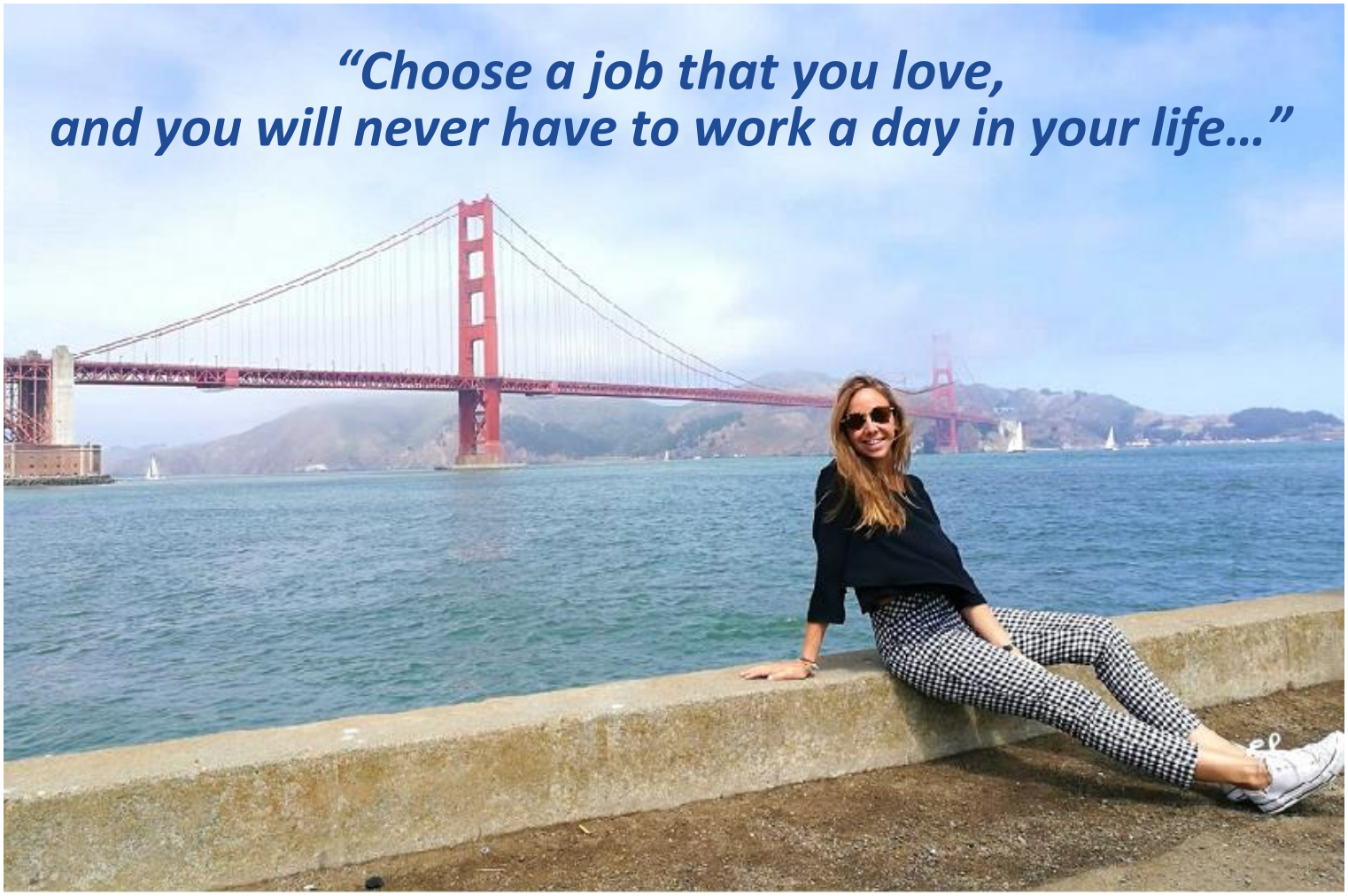
Lesson #1: Burgundy is Bourgogne.

Exactly!

[*both laugh*]

But this you knew already! Can you tell us more about your work?

My research interests lie in artificial intelligence with an emphasis on computer vision. My work focuses on 3D visual scene understanding with both geometry and deep learning. I'm using cameras with different fields of



view. My first project is using omnidirectional images - 360º images.

How did you get to this point?

Actually, it's funny because I did industrial engineering both in my master's and bachelor's. It's not really, really related to computer science. My interests grew out of when I had to decide on my master's thesis. I saw this proposal from a professor who is now my supervisor, José Guerrero. I thought: "*Oh, it's cool! But I don't know what it is!*" I sent him an email, and it said: "*Hello, I am really curious about this topic of computer vision, but I don't know anything about it.*" He told me: "*Okay, Clara, it's not a problem. If you are so motivated, for me, it's the most important thing.*" I was about to leave for Erasmus to Italy, to Polytechnic of Torino. He told me that we can see which subjects I could do there in order to start with some notions of computer vision. He proposed that I start with scene understanding. What we are doing is the recovery of indoor scenes. I started with this in my master's thesis. I really enjoyed it, and in the end, he proposed that I start a PhD I said: "*Yes, okay!*"

Where are you from in Spain?

Zaragoza. I did Erasmus in Italy.

In Trieste and Torino, right? I'm Italian. Tell me something nice about Italy.

It's a country where I feel very comfortable. To be in Spain, for me, it's home. When I arrived there, I didn't even know how to speak Italian, but in the end, I went out with Italians and tried to have an Italian life. I felt really, really comfortable. I love the language. I think that in the future, I could live in both countries. I love the food. I love the people.

Italians are fun.

Yes, I agree.

Thank you for the advertising! [*both laugh*] **How long did you stay there?**

The first year, when I was in Trieste, I stayed for the whole academic year. Then, in Torino, I stayed from September until March.

You probably know that the most popular Erasmus destination for Italian students is Spain.

Yeah, yeah, I know!

[*both laugh*]

I think that what you said about Italy is exactly what they say about Spain.

Yes, I guess we have very similar countries,

so it’s like to be at home.

It sounds like your plans for the next two or three years are already fixed. What happens after?

In the beginning, I had two things in mind. On the one hand, I was loving research. I wanted to do more of it. On the other hand, I wanted to start working in a company, maybe in a startup, for example. In the end, I decided to start doing my PhD Probably when I finish my PhD, I would love to find a place where I can do research at a startup, for example. There is a sentence that I love which is: “*Choose a job that you love, and you will never have to work a day in your life.*” For me, this sentence is really, really important.

Who said it first?

Confucius.

My father uses this sentence all the time. “*Confucius says…*” I wanted to know if your sources are the same.

[*both laugh*]

To be sure that you do not have to work a day in your life, does that also include teaching?

Actually, I have no experience teaching. Maybe, it’s not the thing that I really like. I love the part of research more. I think it would be nice, also. I prefer the research part, and for this reason, I would love to go to a company to continue doing research. If

you love also what you are explaining to the others, I think it's nice.

You were at ECCV a few weeks ago. It was probably your first big computer vision conference. Is that right?

Exactly!

Can you share with us what you learned there? Not all of our readers had the chance to attend ECCV.

I went to ECCV with two different workshops. One was the Women in Computer Vision workshop. The other was 3D Reconstruction Meets Semantics. I was really excited because it was my first conference. It was also my first time explaining my work to all these people. I really enjoyed it because I love explaining my work and discussing this with people. It's important to hear the feedback they can give to you. I was talking to important researchers. I follow their work.

I want names here!

In this case, for example, it was Thomas Funkhouser from Princeton University. My first work was based on a work of his students. I really liked this group. I also met with a lot of people whom I had met this summer, at a summer school in Sicily, the ICVSS in Punta Sampieri. We really enjoyed all the days.

School is not always boring!

"...my first time explaining my work to all these people!"

No! [*both laugh*] Daniel Cremers was also at ECCV. It was really nice to see everyone again. Also, I enjoyed the conference in general, going to this lecture, going after to this poster because I like it, to ask questions, to interact.

You also had a poster at ECCV. Can you tell me how it was to defend your work in front of people coming to ask you questions?

In one of the workshops, first, I had to do a presentation. I had like two minutes to sell my work. Actually, I felt very relaxed and comfortable at the poster. It's your work. You know how to defend it. People usually come because they are interested, and they say, "*Oh! It's very cool!*" My work is really visual. There are lots of 3D rooms moving. People come and ask really interesting questions and give good feedback.

What is the toughest question they asked you at the poster?

I think that the most difficult question comes from people who are not super into this topic. Maybe they ask you something that is not really related. You have to answer in a way like: "*No, because it's not really what I'm doing.*" When people who come and know about this topic, it's easy.

So you found a polite way to answer.

You have to be careful because it can also happen to me. If I go to a poster, I understand something different, and I might ask. They could say: "*No, it's not like this.*" You have to find a polite way.

What was the highest number of people you had at your poster at the same time?

I don't know. Sometimes you are really focused on the people closest to you. Suddenly, you look, and there are more people!

They all circle around.

Yes! For example, I was also in IROS in Madrid two weeks ago. I was with a second author. In the poster section, it's very comfortable when there are two people. Some more people can come and ask about your work.

So next conference, you will already be a veteran, and you will not fear presenting.

Well, it depends: because at ECCV you have the main conference. You have to speak in single track in front of a lot of people.

Thousands!

Yeah, at IROS, I was presenting at the main conference, but IROS was not in single track. It's fewer people.

It's not everybody. With single track, it's everybody.

[*laughs*] Yes, exactly! I think it's a challenge.

What does it mean to be a scientist?

On one hand, doing research on a topic that you are really interested in, to discover new ways of doing things. On the other hand, it's very important to show your work, to share your goal, and make it available for the others. It's a way to improve and advance this field. I think that computer vision, and artificial intelligence in general, is growing nowadays. A lot of people are interested in this topic. Society can advance in these kinds of ways. It's important to share your work.

So you have a double goal. One is to do a lot of research and the other is to share it.

Exactly!

What about if you do all this research, and nothing of it ends up in the real world?

I want my research to be useful for something also.

So it has to have some application in the real world.

Yeah, to be used in the real world or for another person to continue the work.

If I gave you the chance to be famous for one thing, what would it be?

Research. For me, it would be nice to do good research and have the opportunity to go to other places, to be older and have enough experience to also give good advice to younger people starting in a PhD.

What is the most precious teaching that you received from your two supervisors?

Both are really important. They not only help me in research, but also I can talk with them about myself if I'm worried about something. They are very close, and this is very important. In the future, if I'm the supervisor of a PhD, I would love to have this balance. They always tell me that the most important thing is to be happy.

And you are happy?

Yeah! For example, they put pressure on me when I must finish something, but when it increases, they know how to calm me. They keep this balance.

Jessica Sieren told me once: ***"If you don't have a crisis during your PhD, it means that you are doing it wrong."***

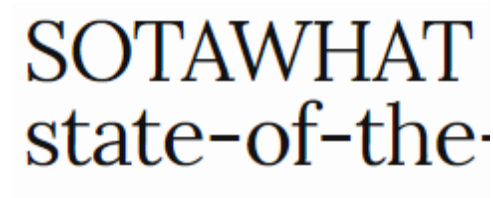
[*laughs*] I agree! If you don't have a crisis during your PhD, you are not doing a PhD!

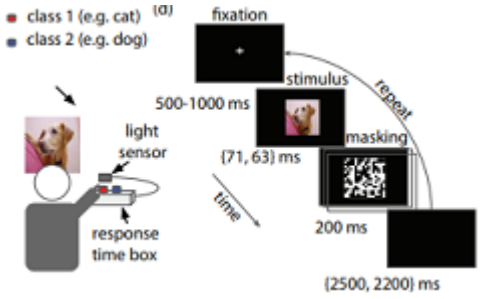
What do you love, Clara?

Research, travel and Italian food!

A Script to Keep Track of State-Of-The-Art AI research!

Let's start with a great tool developed by **Chip Huyen**: for lack of better search options, she wrote a script to **query arxiv** for abstracts that contain a specific keyword and return summaries of those abstracts. The script is simple (283 lines of code) and you can find it on Github. **Read More on Chip's Great Blog...**



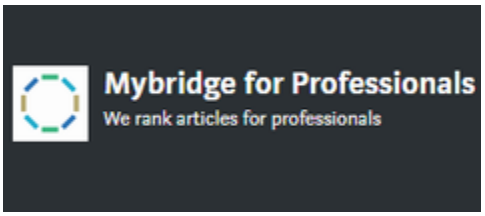


Adversarial Examples Fool both CV and Time-Limited Human:

Google AI has published a great paper that transfers **adversarial examples** from computer vision models with known parameters and architecture to other models with unknown parameters and architecture, finding on the way that those adversarial examples **fool humans as well**. Needless to say, **Google Brain**'s **Ian Goodfellow** is one of the authors. **Read More...**

How AI is Helping Amazon Become a Trillion-$ Company:

A bit long but very complete panoramic view about how **AI** shapes every aspect of **Amazon**'s business, from its warehouses full of products to **Amazon Go** and to your **Echo smart speaker**. Do you think you know everything about Amazon? **Well, Think Again!**

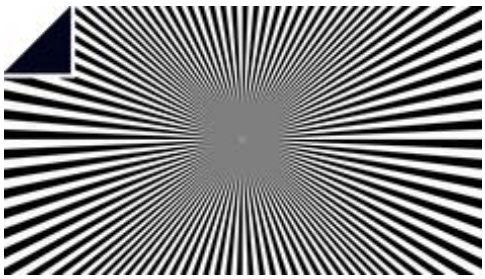


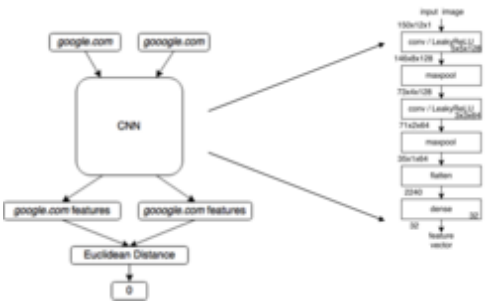
30 Amazing Machine Learning Projects for the Past Year:

Mybridge people have compared almost 8,800 open source **Machine Learning projects** from 2017 to pick their Top 30. This includes libraries, datasets and apps - graded for **popularity, engagement and recency**. We think their list is awesome. Read It...

Neural Networks Do Not Understand Optical Illusions:

I don't know how the image on the right looks on your screen. On mine it looks terrible! So why do I show it here? Because I recommend you look at it in its original size on an article explaining why **Machine Vision** systems cannot recognize **optical illusions**, which means they also can't create new ones. **Read (and look) More!**





Detecting Phishing With Computer Vision - Blazar:

Endgame Research has developed a new **computer-vision based phishing detection tool** called **Blazar**. Phishing attacks are very common confidence scams, mainly designed for **financial gain** but also for **espionage**. Apparently Blazar detects malicious **URLs** that masquerade as legitimate ones. **Read How!**

FREE SUBSCRIPTION

Dear reader,

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

Subscription Form
(click here, it's free)

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

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

BENAIC / BENELEARN 2018 - Benelux Conf. AI / Machine Learning
's-Hertogenbosch, NL Nov 8-9 Website and Registration

AI & Big Data Expo North America
S.ta Clara, CA Nov 28-29 Website and Registration

ACCV 2018 - Asian Conference on Computer Vision
Perth, Australia Dec 2-6 Website and Registration

AI World - Accelerating Innovation in the Enterprise
Boston, MA Dec 3-5 Website and Registration

NIPS 2018 - Neural Information Processing Systems
Montreal, Canada Dec 3-8 Website and Registration

AI Summit (practical implications of AI for the enterprise)
New-York, NY Dec 5-6 Website and Registration

CloudCom 2018 - Cloud Computing Technology & Science
Nicosia, Cyprus Dec 10-13 Website and Registration

IEEE Big Data 2018 - IEEE International Conference on Big Data
Seattle, WA Dec 10-13 Website and Registration

WACV: IEEE Winter Conference on Applications of Computer Vision
Waikoloa Village, Hawaii Jan 7-11 Website and Registration

RE•WORK Deep Learning Summit
S.Francisco, CA Jan 24-25 Website and Registration

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

Did you read our exclusive interview with Yann LeCun at page 4? Don't miss it!

FEEDBACK

Dear reader,

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

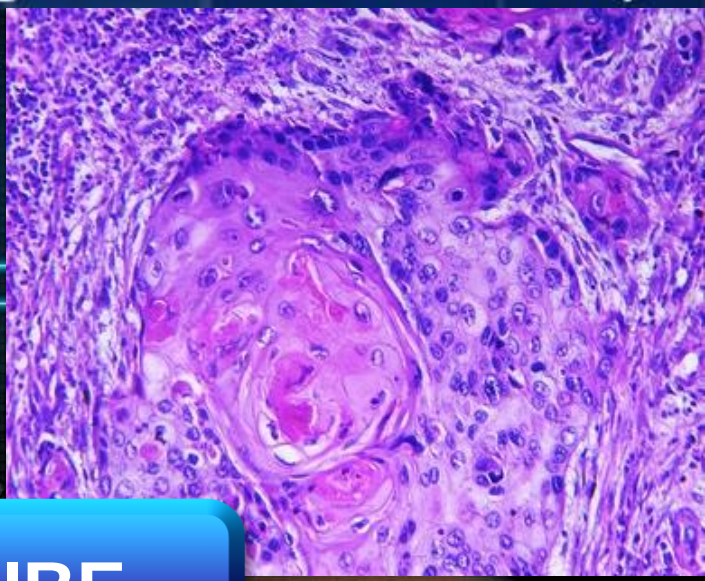
Give us feedback, please (click here)

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

IMPROVE YOUR VISION WITH

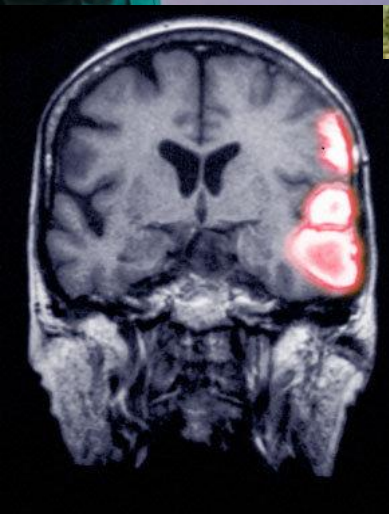
Computer Vision News

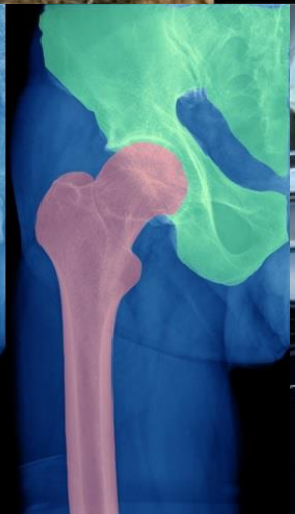
The Magazine Of The Algorithm Community



SUBSCRIBE

CLICK HERE, IT'S FREE





A PUBLICATION BY

RSIP VISION

Global Leader in Computer Vision and Deep Learning