Computer Vision News

The magazine of the algorithm community

December 2017

(a) Anger

(b) Happiness

(c) Disgust

(d) Sadness

(e) Surprise

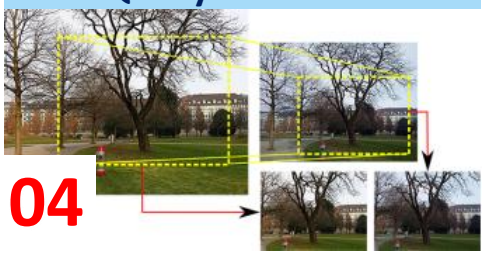
(f) Contempt

Challenge DCER & HPE - Emotions Recognition and Head-Pose Estimation

A publication by

RSIP VISION

Research by ETH
DSLR-Quality Photos on Mobile



04

Project
There is No Bad OCR

18

Woman in Computer Vision
Sarah Ostadabbas

24

Applications
Eyes of Things - Bonseyes with Oscar Deniz

12

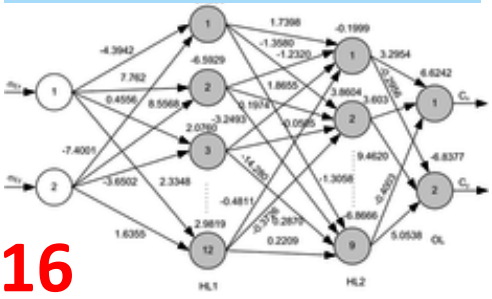
Challenge: DCER & HPE
Emotions Recognition and Head-Pose Estimation

20

Upcoming Events
Meet us at...

29

Spotlight News



16

Project Management
by Ron Soferman

23

Workshop
CoBCoM Winter School

30

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editor:
Assaf Spanier

Publisher:
RSIP Vision

Contact us

Give us feedback

Free subscription

Read previous magazines

Dear reader,

December is the right time to reflect on the past year and sum up everything that we did objectively.

We are more than satisfied with how things went in 2017: **Computer Vision News** and the conferences' Daily magazines are getting just short of **one million page views** during the year, an indisputable sign that many readers enjoy their content; the number of subscribers to the magazine has more than doubled; we have strengthened our partnership with some of the major conferences (CVPR, MICCAI) and events (RE•WORK, AI Expo); we have sealed new partnerships with ICCV and CARS; as a result, we keep hearing that RSIP Vision's publications are fulfilling a palpable need in our community.

Maybe some of our new readers don't know yet that the publisher of this magazine, RSIP Vision, is not a publishing house. We are a **software company**, providing global clients with custom algorithms, consulting and R&D services. Our position as a leader in technology enables us to offer these publications as **a gift to the community**. We hope that you will continue to consider **RSIP Vision** as the most suitable **outsourcing company** to develop great software for your key projects and products.

In this regard, we are preparing **great news for 2018**! You will hear about that very soon…

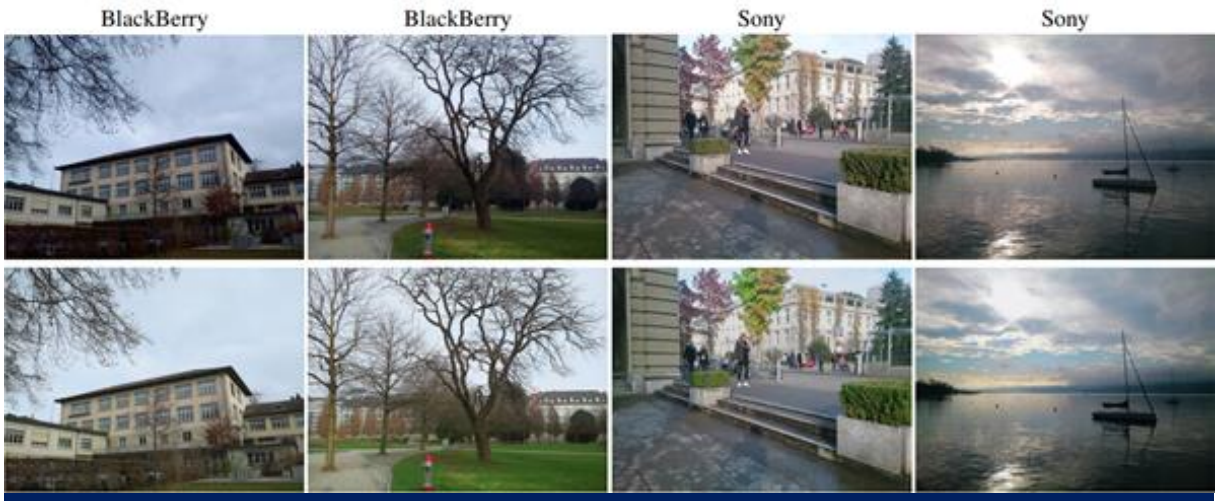
Enjoy the reading!

Ron Soferman
CEO, RSIP Vision

DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks

by Assaf Spanier

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen to review **DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks**. We are indebted to the authors from **ETH Zürich** (**Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey and Luc Van Gool**) for allowing us to use their images to illustrate this review. Their work is here.



Four examples of original (top) vs. enhanced (bottom) images captured by BlackBerry and Sony cameras

Aim and Motivation:

Built-in smartphone cameras, despite continuing fast-paced advances in technology and results, are prevented from producing photos equivalent in quality to DSLR cameras, by their physical conditions: small sensor size, compact lenses and the lack of specific hardware.

Novelty:

There are a number of innovative aspects to this paper: first the authors developed an end-to-end deep learning approach that translates ordinary photos into DSLR-quality images. The network was trained without any additional supervision or handcrafted features.

Additional innovations include:

- A loss-function custom-designed to handle mobile-camera photo quality enhancement that combines content, color and texture losses.

- Construction of a large and publicly available dataset of corresponding DSLR high-quality images and lower-quality smartphone-camera photos from three different devices.
- Quantitative and qualitative evaluation showed the method achieved results superior to the given low quality input images and comparable to the corresponding DSLR photos.

Background:

The last few years have seen smartphone cameras' photo quality improve by leaps and bounds. Today, even phones with relatively simple, affordable hardware produce reasonable quality pictures, thanks to a combination of hardware and advanced image processing software. But when it comes to professional quality photos, the images produced by smartphone cameras still cannot match those from DSLR cameras. DSLRs use larger sensors with higher capabilities, more advanced optics and higher quality aperture, which taken together produce a better, higher-resolution image. There are a number of utilities for automatic image enhancement on the market, but they mostly focus on global parameters, such as contrast or brightness, without improving texture quality or taking the image semantics into account. Moreover, these utilities mostly use a set of predefined rules, which aren't always relevant to the details of a particular mobile device. Thus, today, the leading approach to smartphone-produced image enhancement continues to be manual manipulation using dedicated software.

Degradation in mobile device image quality includes: loss of image sharpness, texture details and precise color variations; once these are lost by the camera sensors, a direct approach at recovering them is sure to fail. The challenge the authors set to themselves was to deal all at once with all these losses by learning a cross-distribution translation, from an input distribution of a given mobile camera sensor image pool to a target distribution of DSLR-produced images.

Method:

The main difficulty of the image enhancement task is that input and target photos cannot be matched densely (i.e., pixel-to-pixel): different optics and sensors cause specific local nonlinear distortions and aberrations, leading to a non-constant shift of pixels between each image pair even after precise alignment. Hence, the standard per-pixel losses, besides being doubtful as a perceptual quality metric, are not applicable in this case. We build our loss function under the assumption that the overall perceptual image quality can be decomposed into three independent parts: i) color quality, ii) texture quality and iii) content quality. We now define loss functions for each component, and ensure invariance to local shifts by design.

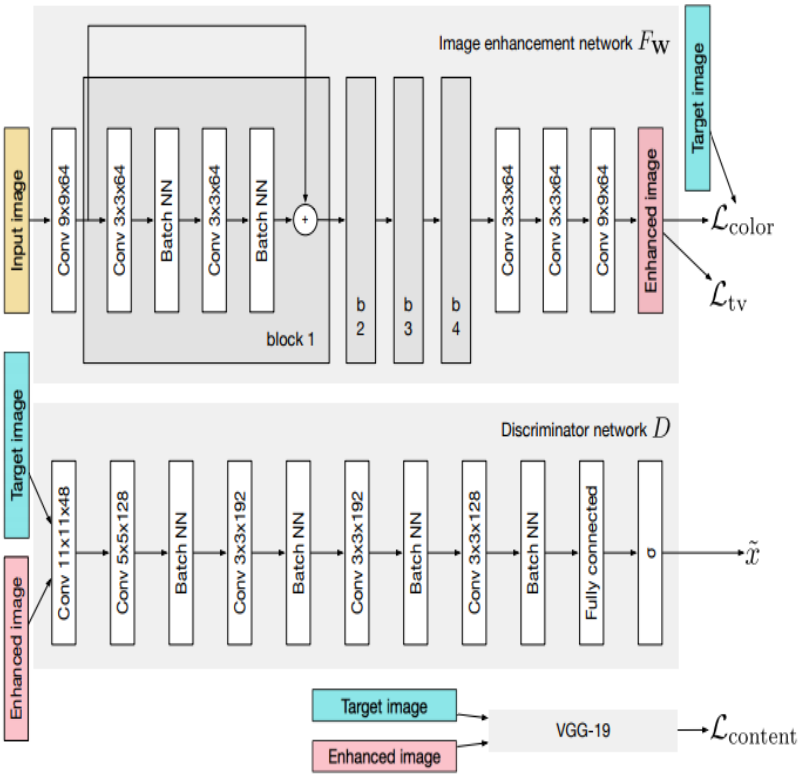
Input: training set$\{I_s^j, I_t^j\}_s^j$ consisting of N image pairs, where I_s is the low-quality mobile source image and I_t is the high-quality DSLR target image.

Output: $F_W$is the deep residual CNN, parameterized by weights W representing the learned translation function from I_s to I_t. The optimization function of W is given by:

$$W^* = argmin \sum_{j=1}^{N} \zeta(F_w(I_s^j), I_t^j)$$

ζ denotes a multi-term loss function, about which more details will follow.

Overall architecture of the transformation network:



A fully-convolutional network comprised of a 9×9 layer followed by four residual blocks. Residual blocks consist of two 3×3 layers alternated with batch-normalization layers. Following the residual blocks, two additional layers with 3×3 kernels and one with 9×9 kernels. All layers, except the last one, have 64 channels and are followed by a ReLU activation function. A scaled tanh is applied to the outputs of the last layer.

The discriminator network is comprised of five convolutional layers, each followed by a LeakyReLU with batch normalization. The last fully connected layer has a sigmoidal activation function and produces a probability that the input was produced by the DSLR camera.

The network was trained on an *NVidia Titan* X GPU with *Adam* optimizer.

Because the input and target images are produced by different optics and sensors -- causing specific local nonlinear distortions -- standard pixel-to-pixel loss functions, which are in any case not necessarily good stand-ins for perceptual quality, are inapplicable to the case at hand. The authors custom-designed loss-functions to handle mobile-camera photo quality enhancement, proposing four loss terms, representing: content, texture, color and total variation losses.

Total loss:

Defined as the weighted sum of the loss terms with the following coefficients:

$$\zeta = \zeta_{content} + 0.4 \cdot \zeta_{texture} + 0.1 \cdot \zeta_{color} + 400 \cdot \zeta_{tv}$$

Next, we'll look at each of the loss terms a little more closely.

Content loss:

$$\zeta_{content} = \frac{1}{C_i H_j W_j} \left|\left| \varphi_{j\,(F_w(I_s)) - \varphi_j\,(I_t))} \right|\right|$$

$C_i\ H_j\ W_j$ denote the number, height and width of the feature maps;
$F_w(I_s)$ the enhanced image;
$\varphi_j\,(I_t)$ the feature map obtained after the j-th convolutional layer.

To preserve image semantics, this loss term confines the network to produce images with similar feature representation of their content and perceptual quality.

"The authors developed an end-to-end deep learning approach that translates ordinary photos into DSLR-quality images"

Texture loss:

$$\zeta_{texture} = -\sum_{i} \quad log(D(F_W(I_s), I_t)$$

Comparisons of CycleGAN against other methods was evaluated on the Cityscapes Label↔Photo dataset of 2975 training images, with pixel-level annotations of the road, sidewalk, parking, sky, persons, etc., image size is 128x128. We used the Cityscapes validation set of 500 images for testing. In the table below you can clearly see CycleGAN outperforms 4 competing style transfer GAN methods, the results are presented using the following evaluation metrics: per-pixel accuracy, per-class accuracy and mean class Intersection-Over-Union (Class IOU). A sample of compared results can be seen in the image following the table.

Color loss:

$$\zeta_{color}(X, Y) = \left|\left|X_b - Y_b\right|\right|$$

X_b and Y_b are the blurred images of X and Y , respectively.

This term confines the network to minimize difference in brightness, contrast and major colors while eliminating texture and content comparison.

Total variation loss:

$$\zeta_{tv} = \frac{1}{C_i H_j W_j} \left|\left|\varphi_{j\,(F_w(I_s)) - \varphi_j\,(I_t))}\right|\right|$$

Being relatively low-weighted, this term doesn't harm high-frequency components, but quite effectively removes salt-and-pepper noise.

Dataset creation:

The authors produced a publicly-available large-scale "DSLR Photo Enhancement Dataset" (DPED). By mounting a Sony Xperia Z, iPhone 3GS, BlackBerry Passport and Canon 70D DSLR on a single tripod-platform (see image of the contraption below) - they produced corresponding sets of photos from all of these devices. They collected over 22K photos: 4549 photos from Sony Xperia Z, 5727 from iPhone 3GS and 6015 photos each from Canon 70D DSLR and BlackBerry Passport. The photos were taken during the daytime in a wide variety of places, illumination and weather conditions. In order to align and fit the images for translation, the authors compute and match SIFT keypoints across the images, which are used to estimate a homography using RANSAC.

Camera	Sensor	Image size	Photo quality
iPhone 3GS	3 MP	2048 × 1536	Poor
BlackBerry Passport	13 MP	4160 × 3120	Mediocre
Sony Xperia Z	13 MP	2592 × 1944	Average
Canon 70D DSLR	20 MP	3648 × 2432	Excellent

"Users could very clearly distinguish between the DSLR and original photos, but nearly as strongly preferred the* proposed *images to the originals."

Evaluation and Results:

Benchmark include comparison with the following 4 tools and methods:

1. APE (Apple Photo Enhancer) -- one of the most popular commercial software, for generating among the best visual results. Using its automatic setting for image enhancement.
2. Dong et al.'s is the baseline end-to-end image enhancement super-resolution-based method.
3. Johnson et al. is the latest state of the art in photo-realistic super-resolution and style transferring tasks.
4. Manual enhancement -- produced by a graphic artist using the classic Adobe Photoshop, given a single workday.

The evaluation metrics used were: 1) PSNR, which measures signal distortion and 2) SSIM, which measures structural similarity, a known cue for perceived quality.

Results can be seen in the following table:

RSIP VISION | Global Leader in Computer Vision

JOIN THE AI AGE IN OPHTHALMIC IMAGING

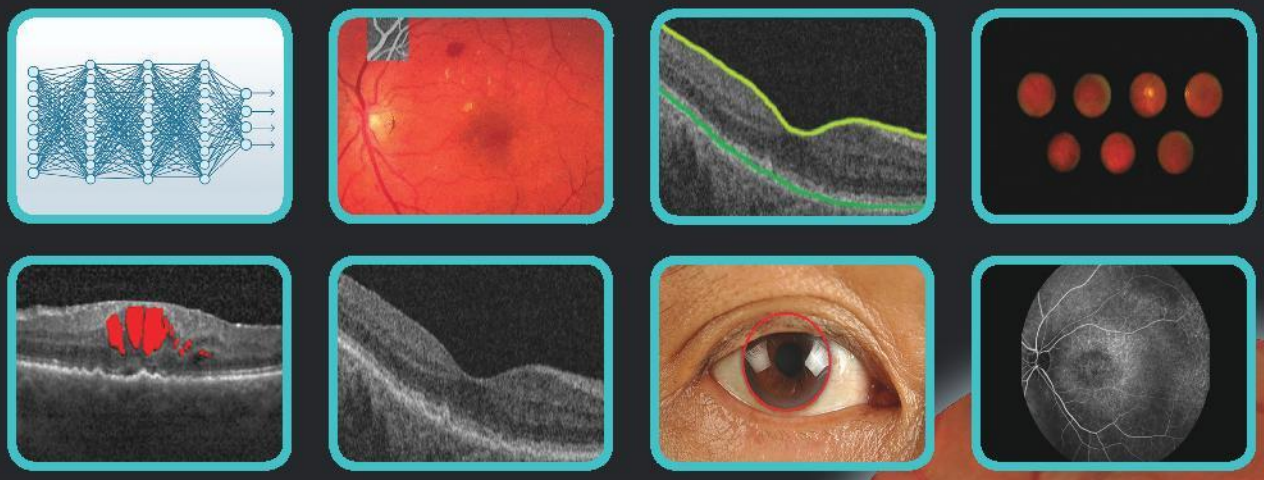
Be part of the AI revolution in ophthalmic imaging. RSIP Vision, a global leader in computer vision and image processing, has completed numerous ophthalmology projects involving development of advanced algorithmic software. In order to achieve state-of-the-art results, the company has invested in breakthrough technology like AI and deep learning.

Your applications can be first-class too! Talk with us about how to gain a competitive edge which will give your device a top-notch performance and a huge advantage in the marketplace. New technologies can benefit all eye care fields in all their components, from telemedicine to diagnosis, from data collection to cloud operations.

Consult our experts now!

Ophthalmology software Research & Development:

- Machine Learning
- Artificial Intelligence
- Image Processing
- Deep Learning
- Computer Vision
- Ophthalmic Imaging

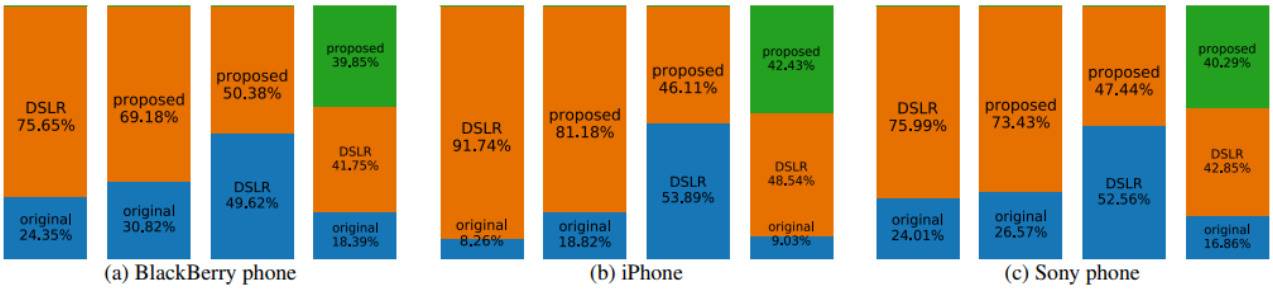


Proud sponsor of ARVO The Association for Research in Vision and Ophthalmology

www.rsipvision.com

Phone	APE		Dong et al.		Johnson et al.		**Ours**	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
iPhone	17.28	0.8631	19.27	0.8992	**20.32**	0.9161	20.08	**0.9201**
BlackBerry	18.91	0.8922	18.89	0.9134	**20.11**	0.9298	20.07	**0.9328**
Sony	19.45	0.9168	21.21	0.9382	21.33	0.9434	**21.81**	**0.9437**

For the purpose of qualitative evaluation the authors conducted a study where users were given pairs of images and asked which is of higher quality. Images were shown in full resolution and the users were allowed to zoom in and out at will. In this setting, all possible pairwise comparisons between the 3 classes of pictures (*original* phone photos, *DSLR* photos, photos enhanced by *proposed* method) were conducted.



(a) BlackBerry phone (b) iPhone (c) Sony phone

As can be seen from the above, users could very clearly distinguish between the DSLR and original photos, but nearly as strongly preferred the *proposed* images to the originals. In the most relevant comparison -- the direct one between the *proposed* images and DSLR photos -- the method achieved very good results, achieving nearly 50% in all three cases.

From left to right, top to bottom: original iPhone photo and the same image after applying, respectively: APE, Dong et al., Johnson et al., the generator network, and the corresponding DSLR image.

Oscar Deniz is a professor and researcher in computer vision at the University of Castilla-La Mancha in Spain. He holds a PhD from Universidad de Las Palmas de Gran Canaria.

You are working on two very interesting projects. The first one is Eyes of Things. Can you tell us more about it?

The name of Eyes of Things is indeed catchy. What we want to do in the project is design and develop an embedded platform for computer vision: on the one hand, it should be open and very easy to use; on the other, it is also optimized for low power consumption and very small. It has low cost and it is very flexible, with a full software stack that you can use to develop any computer vision application.

Is there anything this simple available anywhere else?

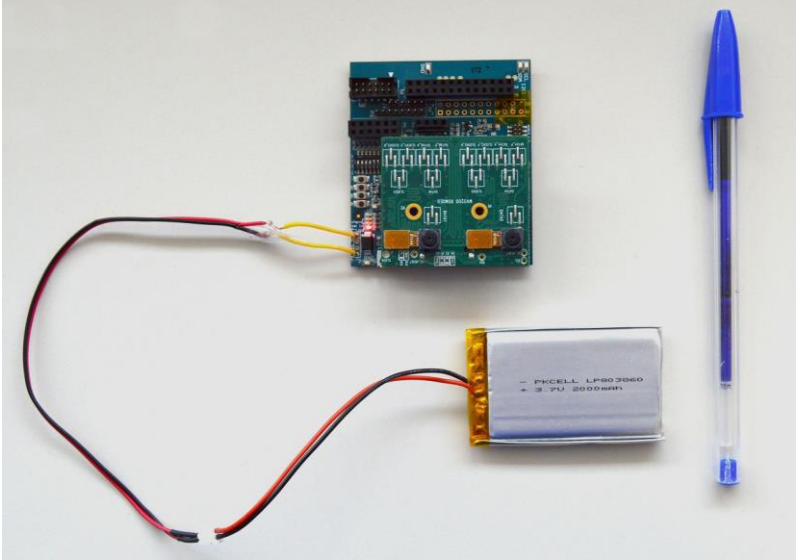
In the past, as computer vision practitioners we sought a flexible platform to develop some of our projects. In fact, we had a project to develop a system for the blind. At that time, we couldn't find something like this. The closest thing was Google Glass, but it was very expensive, and we didn't need the screen for that application, so it was not right. That was roughly four or five years ago. At that time, we had to use a smartphone. You had to wear a smartphone with the camera on, which is only valid for a very basic prototype. Thus we decided to embark on this project. After the beginning of the project in 2015, we have seen a few other projects, mostly from private companies, that aim to do roughly the same. I see that as a success.

Who will adopt your platform?

Since it is aimed to be flexible and open, we want anyone to use it, from hobbyists to SMEs and even large companies. We know, for example, that some large companies like Thales and Intel are now using it internally for prototyping. The software is already in GitHub. The hardware board schematics are also already available. You can do modifications and tailor it to your particular product or service. Hopefully, it can be used by anyone.

How did you manage to spread the word about the results of your work?

So far the main communication channel is through the project website. Then we have initiated what we call an early adopter program with some selected third-party partners that wanted to use and test it. For us, this is good in order to have some feedback from external parties and see what things we can do to improve the hardware or the software. Then we are also developing a number of demonstrators inside the



project that hopefully will entice people. Some of them, the coolest ones, may go on and eventually become commercial products.

What was the most difficult part of this project to achieve?

Yeah, the hardware [*laughs*], no doubt. Any design of new hardware is difficult. Time estimates are always difficult to achieve. On the other hand, that was the tricky part, but there have been some very interesting parts too. For example, the ethics associated with this kind of project is very interesting for us who, as engineers, are not typically dealing with it. It turns out that you can put this very small camera almost anywhere. It has internet connection. Any image, which is personal data, can go to the internet. This is a privacy issue, of course.

How did you solve the software challenges of this project?

We have a full array of software libraries that you can use for computer vision, machine learning, deep learning, and communication with the outside world. We have used lightweight protocols for communication. The device can easily interact with everyday devices like smartphones and tablets that you can use. For computer vision, we have both open libraries and proprietary libraries that are even more optimized for the hardware. For deep learning, we have implemented two frameworks. One is proprietary, while the other is open and is also available in other platforms as a PC, so you can very easily train your network and upload the weights to the device, and then run your application for object recognition or whatever.

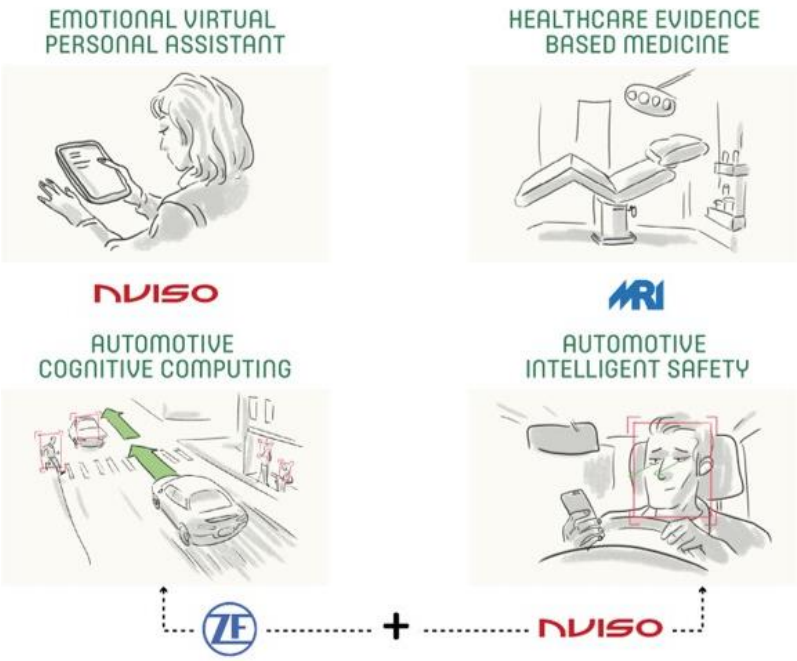
Which technique helped you the most?

Of course, deep learning has become the state-of-the-art tool. We have a cool demonstrator in which we run a CNN network in the device for emotion recognition. This is a network that gives one of seven facial expressions. Everything runs locally with a latency with roughly ¼ of a second. The application itself is for emotional dolls. We embed the board inside the doll's head along with a rechargeable battery. Then it provides some audio feedback to the child.

I understand that it's an H2020 project from the European Union. Who is working with you on this?

We are a consortium of eight partners. We are leading the project here from Spain. Then we have AWAIBA. They are working with a camera which is a tiny, low-power camera. We also have Movidius/Intel who designed the chip. Then we have DFKI from Germany, also working alongside us with the software. Then we have four demonstration partners: EVERCAM, Fluxguide, nVISO and Thales.

"We embed the board inside the doll's head along with a rechargeable battery. Then it provides some audio feedback to the child."



Tell us about the other project which is also very interesting and intriguing, Bonseyes.

The official name is Platform for Open Development of Systems of Artificial Intelligence. The name Bonseyes is just a catchy name. I don't think it is actually an acronym related to the title.

What is behind this?

This project is related to artificial intelligence. There are two key objectives. On the one hand, we want to develop a toolbox for deep learning, in which the main objective is to help you implement your deep learning-based application into a specific platform. This means that in real life you will have to do some approximations to make it run in a specific platform. This should enable you to do that by helping in approximating weights, reducing memory, enhancing sparseness... everything related to the network.

Why does the developer need that?

It is meant to help the developer. Of course, you could do these things on your own, but the idea is to set up a set of tools that you can use to help you do that faster. That is the main reason. Hopefully, we will develop new methods to make these things run in constrained platforms. That's more or less the research part of this project.

What is the current status?

It started in December of last year. We are more or less one year into the project. At the moment, we are mostly working on the second key objective, that is to develop a marketplace for artificial intelligence, a data marketplace. It will include also trained models and not only labelled data. The idea is that we can share trained models. Some of them will be very useful as a starting point. If you want to develop a specific application, you can use those as a starting point and then fine tune your network. Some of them will be free, but for others, you'll have to pay somehow. That is the whole idea. This is also supposed to contain a back-end with a pipeline for training deep learning models. We are currently working on the back-end. It's a full pipeline from training to benchmarking.

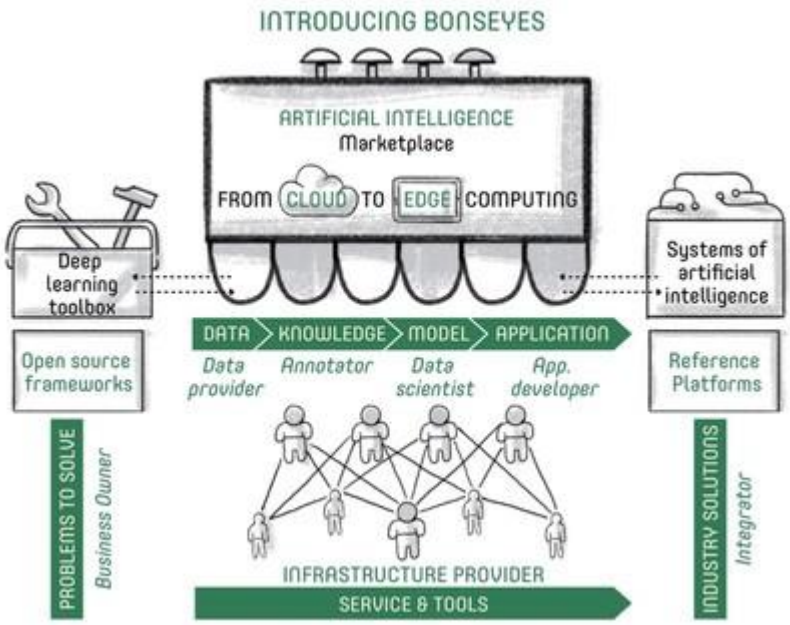
I understand that there is no working version of it. When do you think it will become available?

I think roughly by the middle of next year, we will start to have something for the public.

Do you know the costs for this platform?

No, that will also be studied. What I can guess is that there will be both free and paid options in this data marketplace.

What is the most challenging issue in developing this platform?



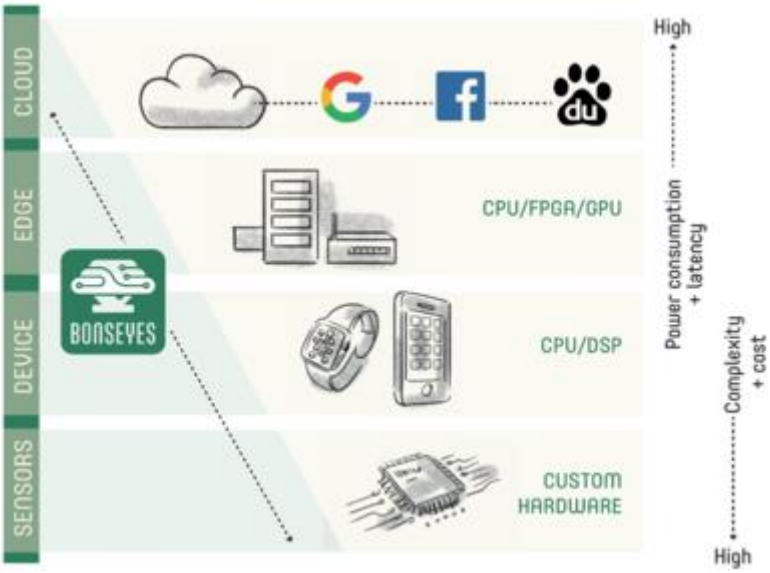
This data marketplace is supposed to be supported with that back-end. Yes, that is a very complex stack of software that we must not only assemble, but put somewhere that everyone can use somehow. We will also have issues with privacy and usage rights. Some partners in the consortium are working on that.

This is also a multi-national project. Can you tell us something interesting about everyone working together from different cultures?

Yes, first of all, I am not the coordinator of this project. Actually, it is being coordinated by nVISO, the Swiss company expert in facial analysis. We have partners in academia, industry, and SMEs. This European project is relatively large and very heterogeneous. Yes, it shows. When the project is large, it is more difficult to get people coordinated. Sometimes some partners are working on something, and you don't know anything about what they are doing. It may be related to what you are doing, but you don't know. We are being very well-coordinated, not just by nVISO, but also by another company that deals with the administrative stuff. We meet regularly, not only in consortium meetings, but also in workshops, which helps very much.

How do you find time to teach with all of that?

[*laughs*] That's a good question! Actually, during the Eyes of Things years, I'm basically free to coordinate the whole project, with an exemption from my teaching duties given by my university.



"We want to develop a toolbox for deep learning, in which the main objective is to help you implement your deep learning-based application into a specific platform"

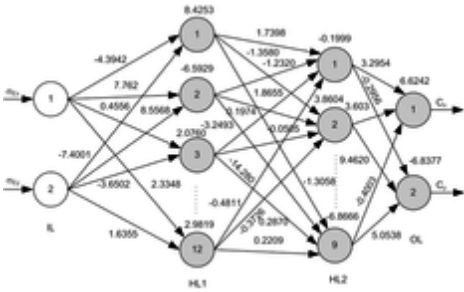
Project Eyes of Things has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643924. Project BONSEYES has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732204 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0159.

Computer Vision News lists some of the great stories that we have just found somewhere else. We share them with you, adding a short comment. Enjoy!

Visual and Intuitive Understanding of Deep Learning

Let's start with an exceptional video by **Otavio Good**. We call it exceptional for at least 2 reasons: one is that he perfectly explains **what deep learning is and how it works**; the other is that it does it with infectious **passion and enthusiasm**. The video is taken from a recent public presentation. Kudos, Otavio!
Click on the image to watch the video!

Software 2.0 - by Andrej Karpathy

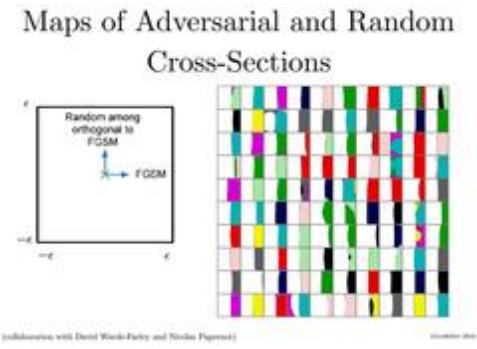


In this article, which we believe will be a milestone in our understanding of deep learning, **Andrej Karpathy** explains that **Neural Networks** are not just another classifier in our machine learning toolbox! *"They represent the beginning of a fundamental shift in how we write software. They are Software 2.0."*
Read Now...

Apple: On-device Deep Neural Network for Face Detection

It's not every day that **Apple** shares with us the **technical challenges** they face. The story says that with the release of the Vision framework, developers can now use the **deep learning algorithms for face detection in iOS 10**. Apple faced some issues when trying to run them efficiently on-device while at the same time preserving **user privacy**. Here are the challenges encountered by Apple when getting deep learning algorithms to run on iPhone.
Read More

How Adversarial Attacks Work?



Chances are that by now most of our readers know very well how **Generative Adversarial Networks (GANs)** work. For those who want to learn more or consolidate what they already know, here is a great article by **Roman Trusov**, full of **codes and examples** that you can follow step by step. We recommend it warmly.
Read it Now...

A Wired story about the woman making Uber's self-driving cars smarter, cheaper: Read
Though our readers knew about that already: Read our article on TorontoCity

Another Toronto story - David Fleet and others join the Vector Institute: Read

The Race to Train AI: Facebook Bested by Japanese

In the process of training a **Neural Network**, you basically jam as much data as you possibly can, as quickly as you can, into a computer so that it'll have some basic understanding of something. Last month (November) have seen unprecedented records in training a **Resnet50** model (a 50 layer network used as a benchmark) with accuracy comparable to **Facebook**'s model which was trained in 60 minutes. As of today, the top performer is Japanese AI company **Preferred Networks**, which used its own supercomputer, comprised of 1024 Nvidia Tesla GPUs (vs Facebook's 256!), to train the model in just 15 minutes. The race is open… Read More

Augmented Reality Glasses Could Soon Help Blind People

One more proof that **Augmented Reality (AR)** is good not only for retail apps and gaming. Scientists **Philip Torr** and **Stephen Hicks** from the **University of Oxford** have been developing smart **AR glasses**, which pick up on visual weaknesses in a person's eyesight, allowing individuals to navigate independently, avoid collisions, or see better in dark or low-light conditions. The trick is exaggerating contrast and features depending on the needs of the blind and partially sighted user. Before you ask, yes, it is already a startup (**Oxsight**) and **Google** is already on the deal.
Read More **Click on the top image to watch the video!**

A Few Reasons Why Using OpenCV Is Better than Matlab

Read it cautiously because it's a sponsored post. No, not here - we have no sponsors and no advertisers - it is a sponsored post on Techaeris: they claim that computer vision with **OpenCV** is better than **Matlab** for image processing tasks. They even offer a few concrete reasons, from speed to efficiency, as well as the specificity of OpenCV, which was created for image processing. We actually agree, with the tiny difference that we are not paid to say that. Read More

"The nightmare robot uprising scenario is wrong"

We hope that he is right when he says that. He is **Emmanuel Mogenet**, head of Google Research Europe. He and his 130-strong Zürich-based team specialise in AI and machine learning. Bottom-line, he says that robot are not going to take over all our jobs, because higher-level intelligence functions, like reasoning, inference and other abstract concepts are (at this point) unique to humans. Read More

There is no bad OCR, there is OCR which wasn't served well

Every month, Computer Vision News reviews a successful project. Our main purpose is to show how diverse image processing applications can be and how the different techniques contribute to solving technical challenges and physical difficulties. This month we review **RSIP Vision's image processing algorithms applied to OCR**. RSIP Vision's engineers can assist you in countless application fields. **Get effective R&D consulting from our experts now!**

"These algorithms do an awesome job in detecting challenging data with high accuracy"

Project

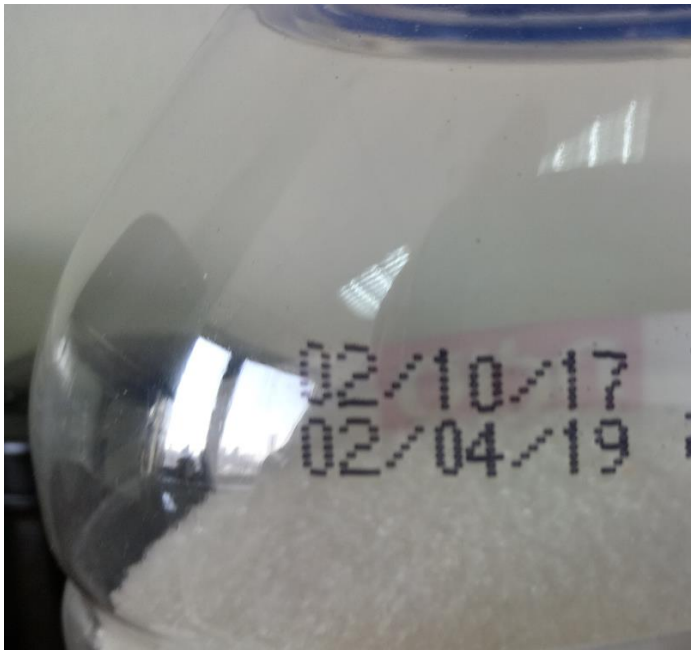
OCR (Optical Character Recognition) is one of the earliest subjects dealt (often successfully) by Artificial Intelligence.

The 80's and the 90's witnessed major progress in **OCR software**, which enabled satisfying results when reading regular quality texts printed on white paper. The leading companies in that area were Calera, Caere and Nuance, which commercialized successful OCR packages able to fulfil this task.

We still see today off-the shelf packages commercialized in the marketspace. However, these standard tools give much lower results when used on today's complex images: in fact, nowadays images are not anymore presented with the level of quality typical of scanners; source is rarely composed of black ink text printed on white paper and most often we need to deal with images taken from smartphone cameras with different lighting conditions.

This is the exact reason **why RSIP Vision's state-of-the-art technology** provides a better solution to the problem. During the 90's we decided not to develop one more general-purpose OCR package to compete in an already saturated market; instead we focused on preparation and post-processing for the regular OCR. A typical example of work in this area aims at reading production and expiration dates printed on bottles and packaging, which comes with several challenges:

- Nowadays, most data comes from photos taken by smartphone cameras; this entails high variability of distance and projections angles, which makes text processing very difficult.
- Different lighting conditions might negatively affect the image, including saturation problems due to reflections of the light source over the object.



- **The geometry of the object** might be challenging too: a label stuck on a bottle of wine or shampoo will have its own geometric distortions changing both size and appearance of the letters in the image.
- **Printing source** might be of different kinds: often, expiration date of products was printed by a dot matrix printer. Printed letters are composed of isolated points, which are more difficult to recognize than the standard input of regular OCR.
- **The simple background** that once was just a white paper to be inserted in a scanner is now more complex: all kinds of pictures, textures and colors make the segmentation of the letter a complicated task by itself.
- For the reasons stated above, **some letters merge** during the binarization of the document: text is digitized with less or no space between letters, which appear as fused, making the task of OCR much more difficult.

"RSIP Vision has focused for many years on understanding the structure of the text"

Computer vision algorithms are thus needed to pre-process and post-process the images, enabling thus the OCR engine to work properly. Nowadays, the breakthrough of deep learning technology allows better results. When the OCR is fed with images of good quality - or which have been pre-processed to reach the proper level of quality - the output is more accurate. **RSIP Vision** has focused for many years now on doing just that, as well as on understanding the structure of the text: we have worked on many OCR projects centered on text which was found on structured fields and tables, rather than free floating in the image. Our expertise enables our algorithms to analyze the image, isolate the field and identify the kind of information which we need to find in them (serial numbers, date, free text or product names that can be compared to a list of products). This model is very powerful! Engineers at RSIP Vision have applied it on many projects and can witness that these algorithms do an awesome job in detecting challenging data with high accuracy.

Get in touch with our experts now.

"Computer vision algorithms are needed to enable the OCR engine to work properly"

The subject of the **DCER&HPE 2017** joint challenge was **Dominant and Complementary Emotion Recognition Using Micro Emotion Features and Head-Pose Estimation.**

It was organized around FG 2017 (the IEEE International Conference on Automatic Face and Gesture Recognition). The goal of the organizers (**Gholamreza Anbarjafari**, **Jüri Allik**, **Cagri Ozcinar**, **Sylwia Hyniewska**, **Hasan Demirel**) was to discuss two important problems that can be very useful in applications like face recognition and 3D face modelling:

- recognition of detailed emotions (including micro-emotions);
- estimation of head-pose in relation to a person operating a computer, to facilitate facial alignment, to be used in applications such as face recognition and 3D face modelling.

To this purpose, the challenge was divided in two tracks: the first track estimates complementary and dominant emotions from a given dataset called **iCV-MEFED** (iCV Multi-Emotion Facial Expression Dataset) and consisting in 31250 facial faces with different emotions labelled under the supervision of psychologists. The goal of the other track is head-pose estimation from RGB-D data based on **SASE**, a 3D head pose database with varying poses and respective labels.

The main winner on all tracks was CBSR-CASIA from the **Center for Biometrics and Security Research** (CBSR), founded by the **Institute of Automation, Chinese Academy of Sciences** (CASIA) in Beijing, China. For the

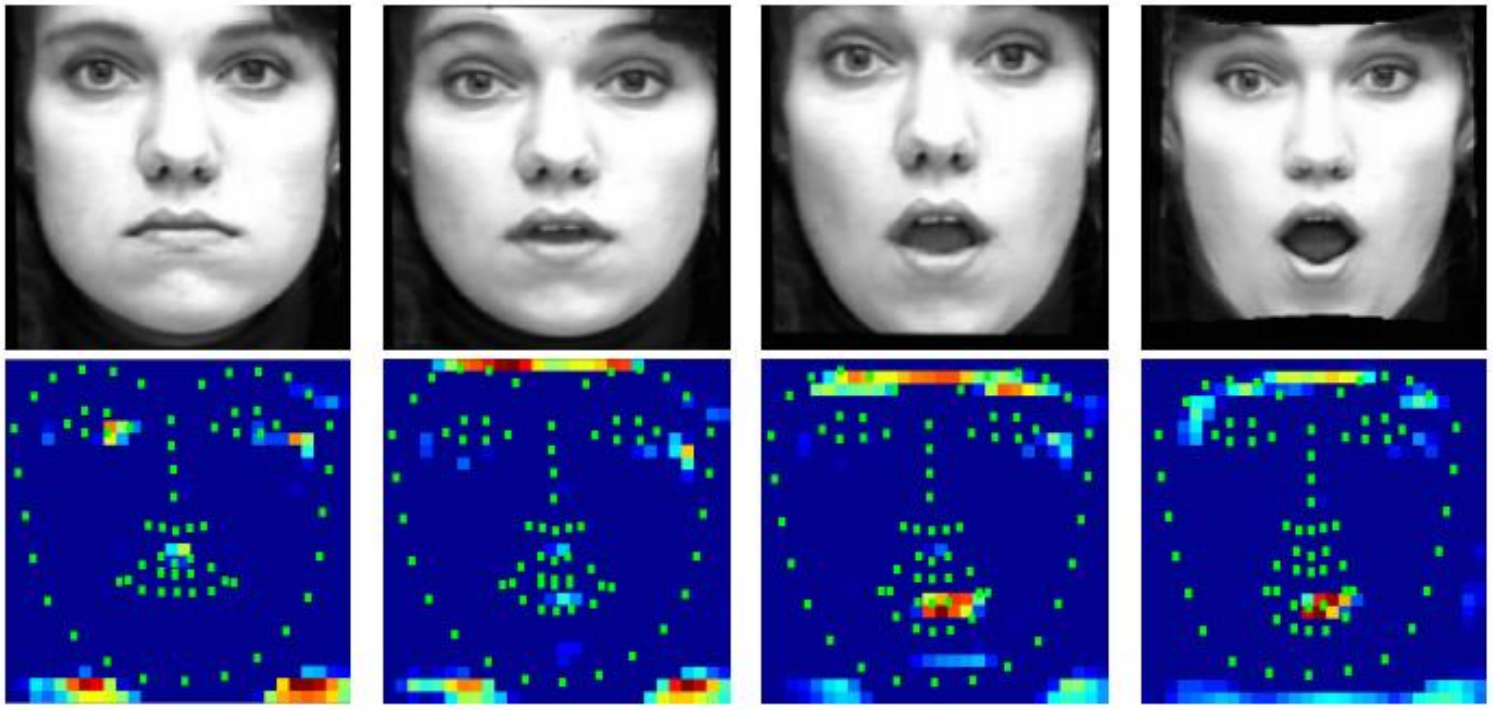


Fig. 8: Example of activation maps obtained from the *conv5* layer of the EMNet CNN for surprise facial expression of the CK+ dataset. As we can observe these feature maps activate around the lips as the emotion goes from neutral to the peak frame. The landmark points are superimposed on the feature maps. These temporal variations are captured with FGT-TPF descriptors. Best seen in colour.

Images courtesy of Ikechukwu Ofodile, Kaustubh Kulkarni, Ciprian Adrian Corneanu, Sergio Escalera, Xavier Baro, Sylwia Hyniewska, Jüri Allik and Gholamreza Anbarjafari from their paper Automatic Recognition of Deceptive Facial Expressions of Emotion

"This fast progressing computer vision community is advancing on all fronts..."

Fig. 1: People may have difficulties in displaying expressions that look genuine when lying. In the case of smiling, differences can be observed in the contraction of the *orbicularis oculi* muscle around the eyes. *Left*: no orbicularis oculi contraction, a marker of fake expression. *Right*: strong orbicularis oculi contraction, with very visible "crows feet" around the corners of the eyes, a marker of genuine expression.

"The beauty of this third method lies in the fact that it allows for partially retaining sequential information in the representation."

emotion track there are two first place winners: **HCILab** (Sejong University, Seoul, South Korea) and **NIT-OVGU** (Otto-von-Guericke-University Magdeburg, Germany). The third place was earned by **TUBITAK UZAY-METU** (Middle East Technical University, Ankara, Turkey).

We asked organizer **Sylwia Hyniewska** what are the main take-home learnings from the Emotion Recognition challenge track.

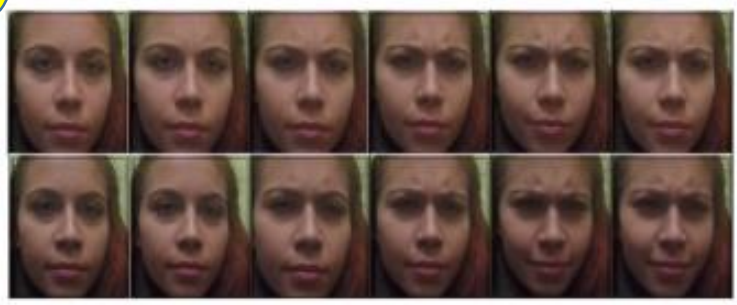
She replied:

*"The challenge attracted many ambitious researchers from around the world. The three best emotion challenge teams were able to differentiate expressions of **felt ("True") and unfelt ("Fake") emotions** by relying on very different approaches. The OVGU team focuses mostly on an estimation of the intensity of each particular facial action observed in videos, then applies a simple Rank-SVM, while the Sejong University team relies on a mirror neuron-like approach. In the latter case, researchers extracted facial features from videos using a face library (DLIB library) then trained a modified Recurrent Neural Network (RNN) which includes a Long short-term memory (LSTM). The team to earn the third place, on the other hand, relied on a sound theoretical stance and focused on brief emotional*

changes in eyes and mouth movements, a long short-term memory model, a Compact Bilinear Pooling and a simple SVM classifier. ***The beauty of this third method*** *lies in the fact that it allows for partially retaining sequential information in the representation.*

Our conclusion - Sylwia explains - *to this challenge track is that this fast progressing computer vision community is* ***advancing on all fronts****. While cooperation between labs and exchange of information is crucial as in all tech domains, it seems that so far there is no efficient golden approach to giving machine the ability to "look at people". More large scale initiatives are necessary in order to maintain the creativity of this vibrant community."*

Sylwia Hyniewska is a researcher in Affective Sciences at the Bioimaging Research Center in Warsaw, Poland



(a) Anger

(b) Happiness

(c) Disgust

(d) Sadness

(e) Surprise

(f) Contempt

Fig. 3: Expressed emotion pair sequence showing acted (top) and fake (below) for each emotion in theSASE-FE dataset

Navigation through the early stages of the project

RSIP Vision's CEO **Ron Soferman** has launched a series of lectures to provide a robust yet simple overview of how to ensure that computer vision projects respect goals, budget and deadlines. This month we learn about **Navigation through the early stages of a project**, another tip for **Project Management in Computer Vision**.

"Key steps in the R&D will lead to the right route..."

The early stages of any R&D project are very sensitive. They are a crucial time in the pursuit of the ideal path to the research area. Any mistake in these stages might lead to the wrong direction, wasting precious resources.

When the team is entrusted with a new challenge, it is advisable to start with key steps that will lead to the right route.

The first step is: getting acquainted with the data. Project manager needs to go through many images and use imagination to understand how the human brain deals with this task. Knowing the data will allow to prepare the following steps with additional ideas that will be tested later.

The second step involves scanning the literature to look for prior related work. The field of computer vision is enriched by much work done all over the world. The project manager must learn what has been done and which action items were successful in the project which is the closest to the task at hand. This may include finding shareware software enabling the initial testing.

The third step requires to focus on the core problem that needs to be answered during the early stages. The purpose is to concentrate all efforts on locating the technology that is needed to solve the key challenging issues of the project. This focus will also help all the stakeholders to understand where are the risks as well as the solutions.

This focus will require work (even manual work!) to get the exact data needed for these tasks. An example of manual work might be cutting the information needed for the core part, assuming that the remainder will be treated later on during the project.

The fourth step consists in locating the tools that are needed to estimate the difficulties. These tools might be shareware from prior art and (more importantly) transfer learning opportunities.

Using deep learning (often the best practice), requires to locate, very early, similar problems, the neural network of which may be used by transfer learning to obtain initial results. This will tell us whether this or other networks are suitable to solve the task at hand and provide the first meaningful results. This will lead us to the tools that we will adopt further on during the project.

The fifth step - conclusion and iteration: after summarizing the results of the previous steps you will be able to scope the next iteration in the R&D process.

These steps have a key importance even during the first week of the project: they are the inception of a long journey of research, during which the project manager wants to keep control over the resources.


FENWAY PARK
"The U.S. is my home now..."

Sarah Ostadabbas

Sarah Ostadabbas is an Assistant Professor at the Electrical and Computer Engineering Department of Northeastern University (NEU) in Boston, Massachusetts.

Sarah, I know that you have recently opened a lab. Can you describe what you do?

It's called Augmented Cognition Lab. It's about enhancing human information processing capabilities through the design of adaptive interfaces based on physical, physiological, and cognitive state estimation. This is the main goal of my lab. I would like to use machine intelligence to make humans perform at a higher performance. This is the ultimate goal, but actually the day to day job is at the intersection of computer vision and machine learning.

My lab looks like a regular computer science lab. We have supercomputers and use cloud services from Amazon when having huge models to train, but we do have some off-the-shelf equipment as well. We have different types of cameras, such as Microsoft Kinect. We have a range of infrared cameras since we want to collect data all the time, under different lighting conditions; many of these cameras are put together by us.

The other thing my lab has that not many labs in the Northeastern or even in the US have is the state-of-the-art Augmented Reality and Virtual Reality equipment. The reason is that I do a lot of simulation, which has some human factors involved. To make it realistic, with realistic human behavior, we place subjects in the virtual reality and augmented reality environments. In summary, the augmented cognition lab, which we call it ACLab, is composed of a bunch of very powerful computers because, most of the time, we do modeling and processing of images, videos, and sometimes signals. We also use different kinds of off-the-shelf cameras to collect data and when we fix them, they become state-of-the-art for VR and AR data collection. This is the whole lab.

How many people do you have on your team?

I have three full-time PhD students and two Master's students, doing their thesis with me. I would like to have a compact team at this point: we have weekly meetings, weekly journal clubs and book clubs, and also a lot of daily interaction. We talk about the projects, and sometimes I sit and program with them. The lab, at this point, is small, but I like the size because then the interaction is very seamless and frequent.

Did you choose the people in the lab?

Yes, people say I have a really rigorous interview process. It starts with an email, usually initiated by the student. Then I have an informal interview, in which I talk and make sure that they're interested, that they have a good math background and that they are passionate about the specific line of work that we are doing. This is not only about computer vision or machine learning. I am interested in understanding humans, estimating and predicting their behavior and then providing information in a way that is going to enhance their performance. This is specific, and I want the student to be passionate about this.

After this, when I see that the person has the interest, we have a two-hour technical interview which is mostly based on the linear algebra, programming background, and then I give the student a couple of our papers, and ask them to read them. I ask some questions about the papers to see how they pick up challenges, questions, problems or calculations. It is, in my opinion, very straightforward, but people think it is a bit rigorous, which I like. I have been fortunate to have very strong PhD and Master's students in my lab.

Who supports the lab, besides Northeastern University?

As a new assistant professor, I have a startup budget, which covers a couple of students for a couple of years. I have written grant proposals to government, to industry, and to foundations. I have been fortunate enough to get a couple of grants from MathWorks. The project led to several publications and also an internal grant at Northeastern. I have quite a few proposals under review that I'm hoping to get funded. The bottom line is that I do have my research goal. These fundings give you money for your equipment, for your salary and also for your students.

What is the most challenging for you? Finding the money? Getting results? Meeting deadlines?

That is a very good question. I believe faculty, before getting their tenure, are juggling several balls at the same time. You need to do a fantastic job in mentoring the students, which results in good publications, in journals and conferences. At the same time, I need to bring the money for the students so they are making enough results for me to write the proposals. You have a phase of getting money, and a phase of getting results. Sometimes you need to get money to produce the results. I need to work at the same time in order for students to get results and get enough funding. That funding can get the students to get the results. It is a little bit challenging. It's interesting. When I'm writing a proposal for funding, I have to be imaginative and think about what you want to get out of that. Another thing is that I need to do a very nice job at teaching and also provide services to the university, mentor undergrad students, go to open houses, be part of graduate meetings… These are all parts of the job, but this is the job that I always wanted. No complaints. But having all of these things and doing a very good job at all of them at the same time can be challenging.

What trained you to handle so many challenges?

First of all, I need to thank my teachers and mentors over the years. I went to a very good high school back in Iran, my native country. It was very science-oriented, but they gave us a lot of responsibility. We had our projects, and we had our teachers and mentors. Since that time, the ideal thing that attracted

ACLab's start-of-semester gathering

Senior undergrad end-of-year gathering on the coast of Caspian Sea (north of Iran)

me in becoming a professor was that I would have my own lab. I even picked out the name of my lab. Throughout my undergrad and master's in Iran I went to very good schools, AmirKabir Polytechnics University and Sharif University, the top in the technical fields back in Iran. I learned from different teachers, who made their own teaching models. Then when they made mistakes, I thought that these were data points that I shouldn't repeat.

Then I came to the United States for my PhD, at the University of Texas at Dallas. After that, I did my postdoc at Georgia Tech. Throughout, I had supervisors and advisors that were pre-tenure and tenured. This gave me a very good information about people's life and how much effort they should put into it. I can say that I'm thankful for every single teacher that I had because I could learn from them whether a pattern may lead to a success or not. I thank my parents who gave me confidence that I could become an engineer in the first place. At school, everyone was telling me to go into the medical domain. But I loved math, and I loved physics. I told my parents that I wanted to become one of the first engineers in our family. They believed in me. They said, "*Just go! Become an electrical engineer.*" They always accept my choices. They give me confidence, which I believe is very important.

Was there a teacher that had a big impact on you?

I can name Dr. Ali Fallah, who taught signal processing during my undergrad in Iran. At that time, there weren't many female students in the engineering department. I have seen that it is the same in the United States. At this point in Iran, the situation is better. Dr. Fallah gave me responsibilities. I was the only undergrad TA in this school. Let me tell you, it was nothing official nor even legal to be paid as a TA while being undergrad. He used the name of one of his grad students under the TA position. The student was paid, and the paycheck came to me. It gave me the confidence to go to class, to teach and talk to the students, have them listen to me and convince them that I am right. That's how I understood the reality of working with students.

That was a very, very interesting experience. Throughout my undergrad, he was very supportive, always behind me. That's why I wanted to become a professor. I had all of these crazy, big ideas like I wanted to publish a book or have my own journal. He was there, and I kept in touch with him. I am

Last day as a student:
goodbye party at Georgia Tech

thankful for all of the confidence that he had in me.

It sounds like what helped you the most was when someone gave you exceptional trust and confidence. Did you tell your students about this story?

I don't think I've shared this story with many people. Of course, my family knew because of how happy I was. I'm not sure if I shared this story with my students. That would be nice for them to hear. I have nothing to hide about that. I thought once that it was a privilege, but looking back, it wasn't a privilege. It was an opportunity!

Do you see yourself going back to live and work in Iran?

No, my husband is American and I have my career here. I haven't been going back and forth very often, but I have kept in touch. I have a couple of Iranian students in my lab. When I go to Iran to visit, the technical universities where I went for my master's and undergrad back in Iran often organize talks for me to give. But for living, I'll be here. The U.S. is my home now!

ACLab members win the Best Engineering Project award - RISE2017

"...enhancing human information processing capabilities through the design of adaptive interfaces based on physical, physiological, and cognitive state estimation..."

Sarah and husband Eric on top of the Alps

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.

NIPS 2017 Neural Information Processing Systems
Long Beach, CA Dec 4-9 Website and Registration

PReMI 2017 - Pattern Recognition and Machine Intelligence
Kolkata, India Dec. 5-8 Website and Registration

International Congress on OCT Angiography and Advances in OCT
Roma, Italy MEET US! Dec 15-16 Website and Registration

ICPRAM - Pattern Recognition Applications and Methods
Funchal, Portugal Jan 16-18 Website and Registration

RE•WORK Deep Learning Summit
S. Francisco, CA Jan 25-26 Website and Registration

AAAI conferences on Artificial Intelligence
New Orleans, LA Feb 2-7 Website and Registration

RE•WORK Intro to Machine Learning in Healthcare Workshop
London, UK Feb 14 Website and Registration

CardioFunXion Winter School 2018
Lyon, France Feb 19-22 Website and Registration

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

Did you read the Spotlight News at pages 16 and 17? Read the best stories we've found for you!

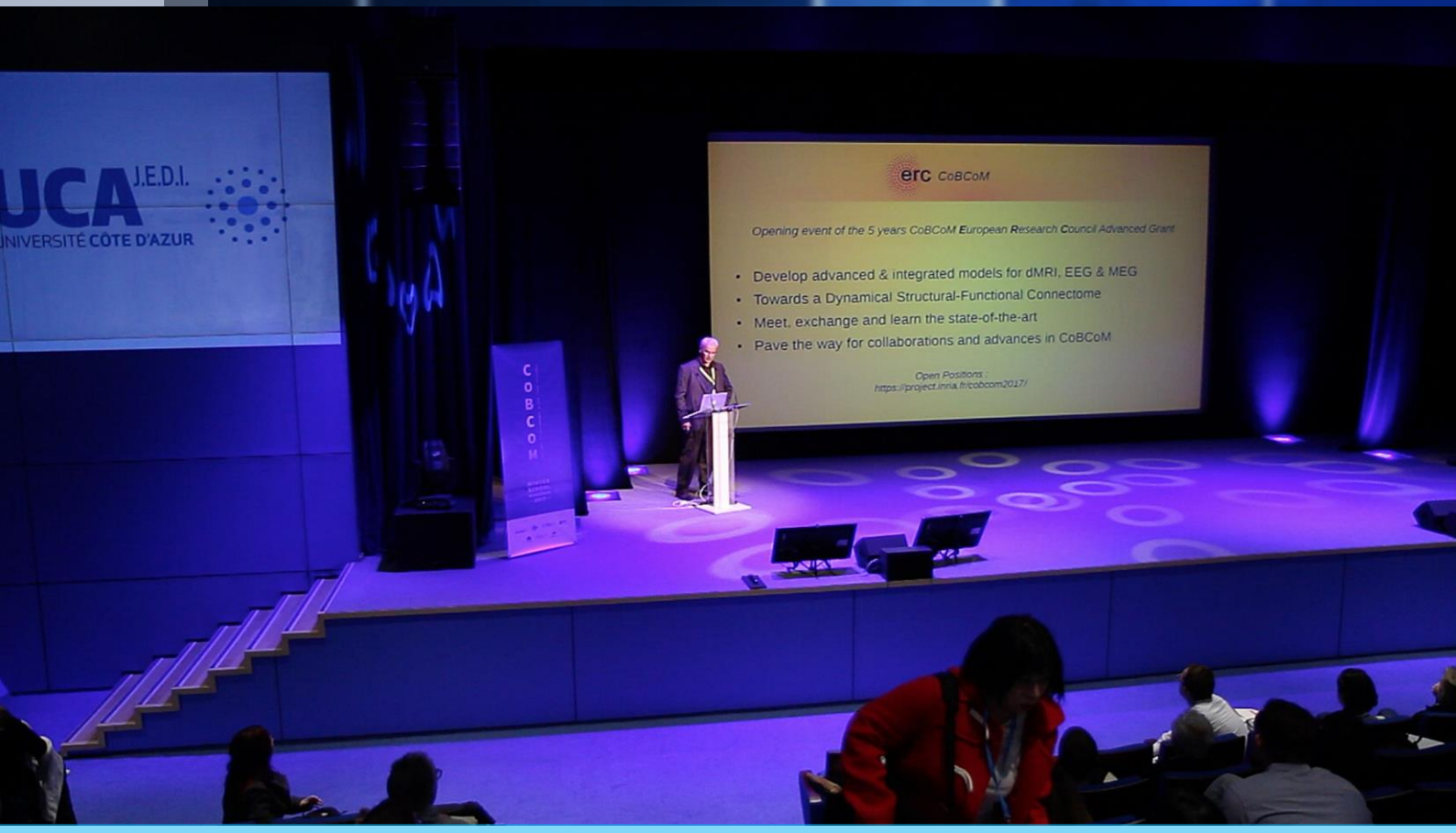
FEEDBACK

Dear reader,

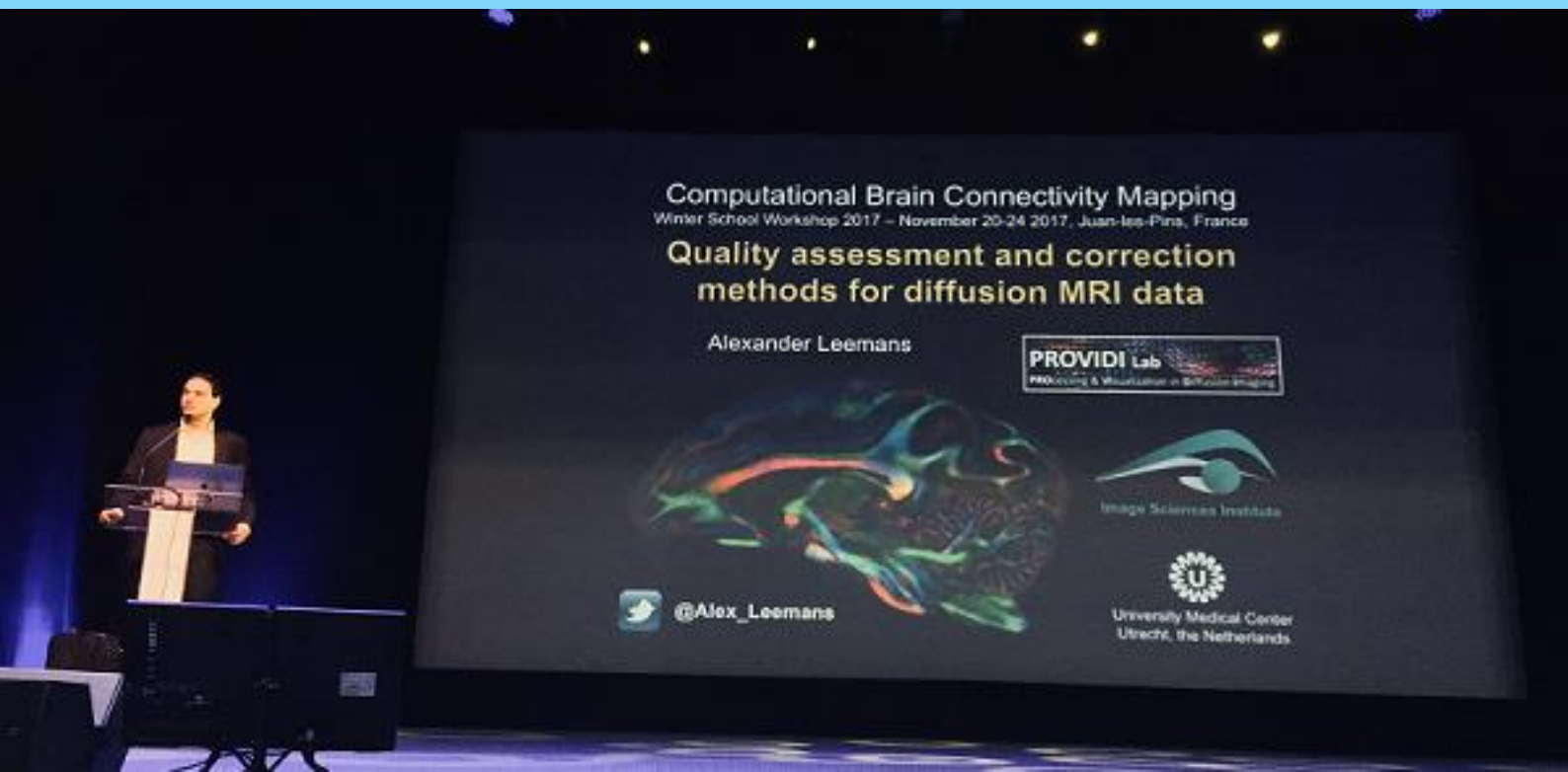
How do you like Computer Vision News? Did you enjoy reading it? Give us feedback here:

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!



A few images from the CoBCoM Winter School Workshop, held last week in Juan-les-Pins, France. The Winter School offered tutorials and advanced lectures from leading researchers and experts on computational brain connectivity mapping. This workshop is part of the ERC Advanced grant for the project CoBCoM "Computational Brain Connectivity Mapping", of which Rachid Deriche is P.I.: Rachid (pictured above) is Research Director at Inria in Sophia Antipolis.
Below: Alexander Leemans, Director at the PROcessing & VIsualization in Diffusion Imaging (PROVIDI) Lab, which is part of the Image Sciences Institute - University Medical Center Utrecht, The Netherlands.

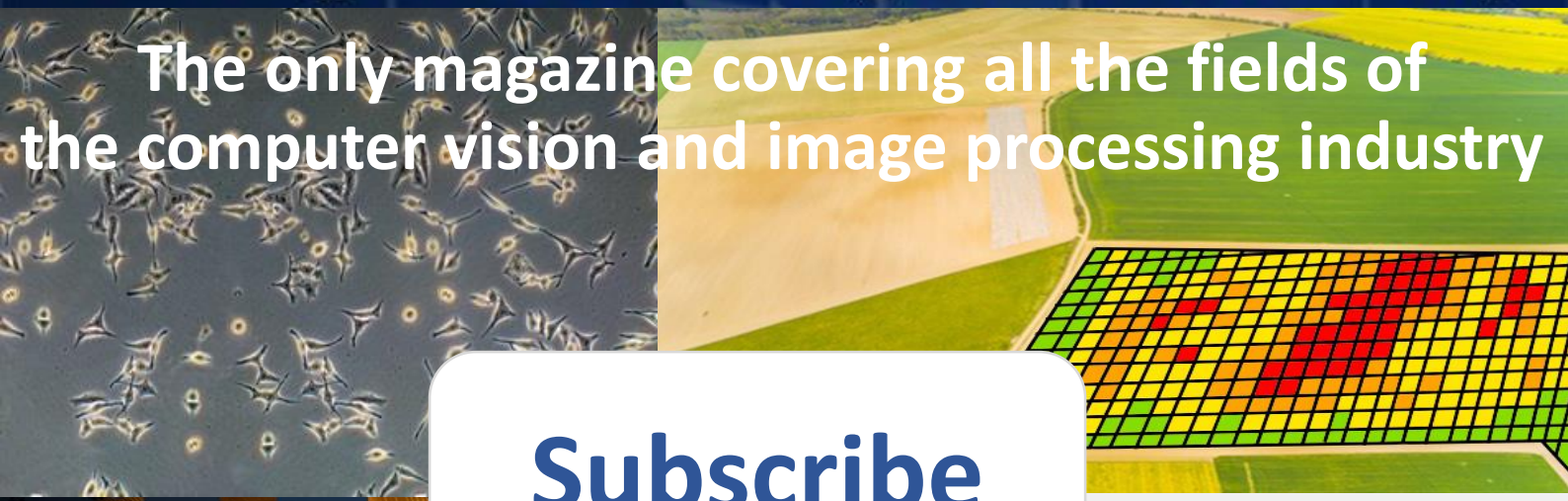


Maxime Descoteaux, head of the Sherbrooke Connectivity Imaging Laboratory (SCIL) at Sherbrooke University. Thank you Maxime for sending over these nice photos!

Improve your vision with

Computer Vision News

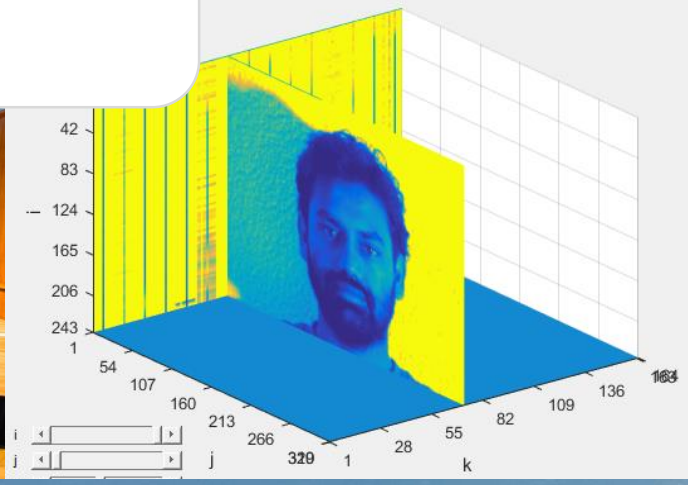
The Magazine Of The Algorithm Community



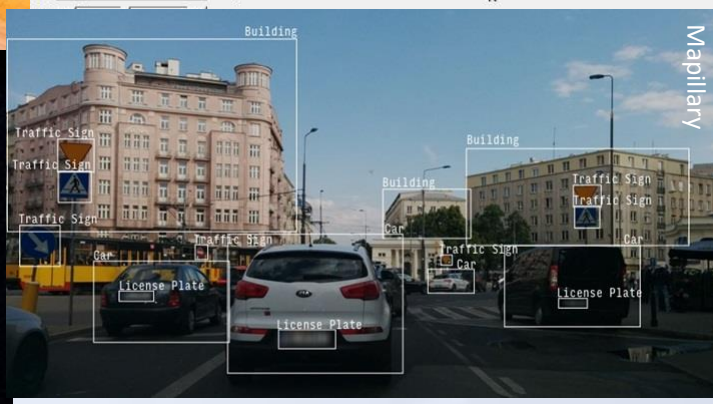
The only magazine covering all the fields of the computer vision and image processing industry

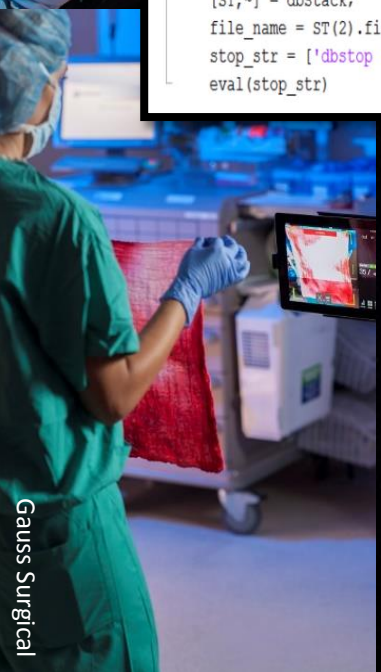
Subscribe

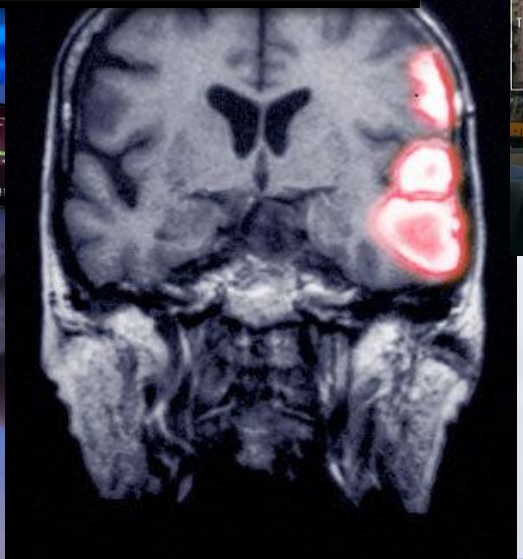
(click here, it's free)




```
% invoke the matlab debugger
function STOP_HERE()
    [ST,~] = dbstack;
    file_name = ST(2).file; fline = ST(2).line;
    stop_str = ['dbstop in ' file_name ' at ' num2str(fline+1)];
    eval(stop_str)
```







A publication by

RSIP VISION