Computer Vision News

The magazine of the algorithm community

A publication by RSIP VISION

January 2018

Women in Computer Vision: Noelia Vállez

Tool: Visual Question Answering - VQA

Spotlight News

Project Management by Ron Soferman: How to Build the Development Team

Challenge: Learning to Run

Application: MobileODT - Reach Every Patient!

Upcoming Events

Review of Research Paper: Breaking text-based CAPTCHAs

Image Processing: Augmented Reality - AR

Guest:

Raquel Urtasun - UofToronto and UBER ATG

"This is just the beginning!"

photo: Erica Edwards @Uber

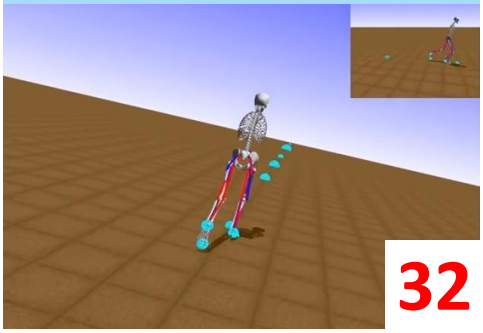
Research
Breaking text-based CAPTCHAs

04

Project Management
by Ron Soferman

15

Challenge
NIPS'17: Learning to Run



32

Guest
Raquel Urtasun - UBER

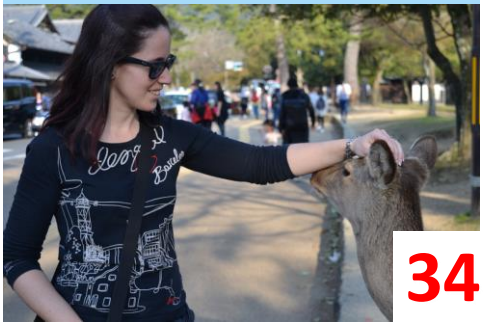
photo: Chris Sorensen

22

Applications
MobilODT

10

Woman in Computer Vision
Noelia Vállez



34

Project
Augmented Reality - AR

30

Spotlight News

14

Upcoming Events

39

03 Editorial
by Ralph Anzarouth

04 Research Paper
A generative vision model that trains...

10 Applications
MobileODT

14 Spotlight News
From Elsewhere on the Web

15 Project Management Tip
Lecture by Ron Soferman

16 Tool of the Month
Visual Question Answering (VQA)

22 Guest - Raquel Urtasun
University of Toronto and UBER ATG

30 Project - by RSIP Vision
Augmented Reality (AR)

32 Challenge - by S.P. Mohanty
NIPS'17 Learning to Run

34 Women in Computer Vision
Noelia Vállez

39 Computer Vision Events
Upcoming events Jan - Mar 2018

40 Israel Computer Vision Day
Hamlyn Winter School

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,

With the new year comes a content-rich January issue of Computer Vision News that caters to all tastes: if you like technical articles, you will enjoy the research, the project, the challenge and the tool of the month **VQA**; if you like community information, you will find it in the **Spotlight News** and the events section; if you'd like to read a case study about a go-to-market initiative, the application of the month **MobileODT** is for you; in addition, you can learn project management tips in **Ron Soferman**'s regular lecture. As usual, we shine a spotlight on a young female scientist: **Noelia Vállez** is the woman in computer vision of this month.

The guest of the month is **Raquel Urtasun**, a researcher and teacher at the **University of Toronto** as well as the head of the **Uber lab** there. In her long and fascinating interview, Raquel shares her vision about more subjects than I can mention here, from the secrets of her success to how Uber is going to save human lives on the road. Take a few minutes to read it, and you will admire her just like we do.

During 2017, our publications enjoyed great popularity among our community, with almost **one million pages of shared knowledge** viewed by our readers. As a global leader in computer vision, **RSIP Vision** is very proud of being at the center of this project, which of course, will continue in 2018 and beyond.

Enjoy the reading and happy new year!

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

A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs

by Assaf Spanier

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen to review **A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs**. We are indebted to the authors (**Dileep George, Wolfgang Lehrach, Ken Kansky, Miguel Lázaro-Gredilla, Christopher Laan, Bhaskara Marthi, Xinghua Lou, Zhaoshi Meng, Yi Liu, Huayan Wang, Alex Lavin and D. Scott Phoenix**) for allowing us to use their images to illustrate this review. Their work is here and their source code is here.

"All the methods the paper is based on are from before 2006"

We found this paper particularly interesting for a number of reasons: 1) It goes against the current: it does not use deep learning methods and still achieves impressive results. All the methods and articles the paper is based on are from before 2006. 2) The paper develops a method called **recursive cortical network (RCN)** – whose structure is an attempt to simulate actual neural structures from neuroscience research insights. The model's internal logic can be justified / understood; this is in contrast again to most current deep learning methods, where we are dealing with a black box, which - while achieving impressive results - it can only rarely (if ever) be understood why this is so. 3) The supplemental materials which come with the paper are explained in impressive detail, including ready to run code which you can try by yourself. 4) Finally, and most importantly, the paper demonstrates impressive results in **breaking CAPTCHAs**, with a method that achieves higher precision then deep learning methods, though requiring a relatively small dataset to train on.

Introduction:

Recent deep learning methods developed to break CAPTCHA required millions of images to train on, while a human being doesn't require a single image beyond the one presented, to understand a given CAPTCHA.

Using insights gleaned from neuroscience, the authors propose their recursive cortical network (RCN) model: A probabilistic generative model for vision, in which message-passing-based inference handles recognition, segmentation, and reasoning in a unified manner.

How does perception in the human brain function? Big question! Some of the perception mechanisms' capacity derives from our ability to use past experience

and the form of data storage in the brain, which can be accessed at the appropriate resolution for a given scenario. A great deal of this know-how is stored in the visual and motor cortices, which act for us as an internal representation of our world. Perception, to be efficient, must be able to handle a wide variety of scenarios concurrently. The more specific question -- what type of computer model is sufficient for simulating perception in the human brain? One of the means of approaching the answer to this question is asking: what would a model representing human vision look like? Then try to expand from there to a model for all perception. In this paper, the authors take a step towards answering these questions. They demonstrate how clues and hypotheses from neuroscience about the structure of the visual cortex can be combined to produce a computer vision model.

Here is a list of clues and hypotheses the authors used in building the RCN model:

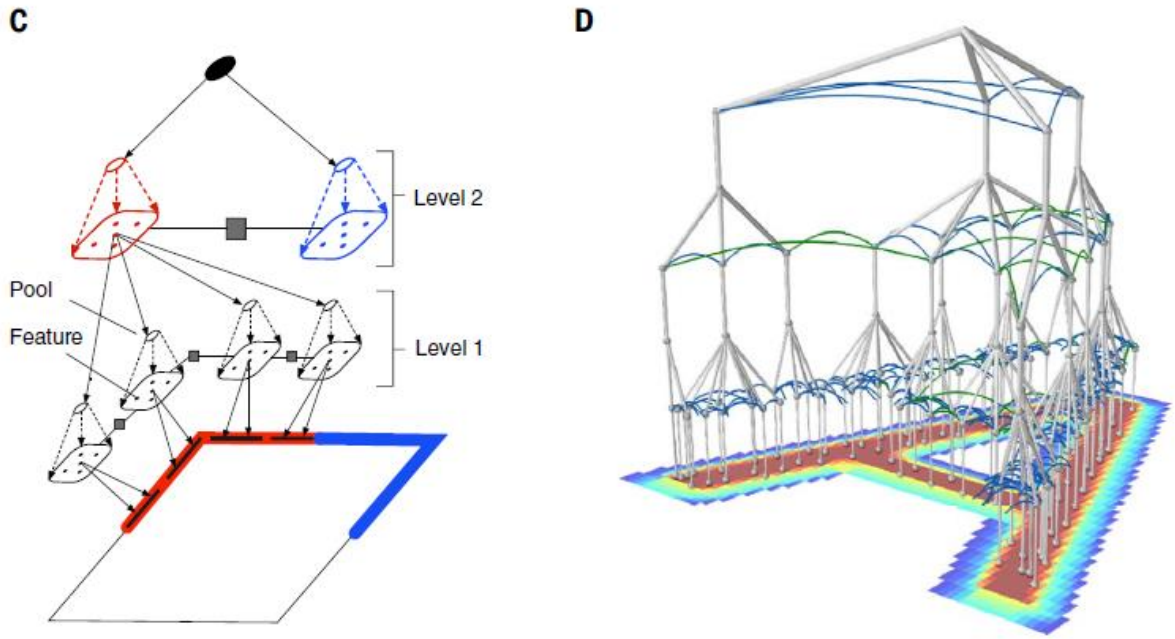
Biological observation	**Representation in the RCN model**
Evidence from neuroscience indicates that contour-lines and surfaces are represented in the brain as separate factors.	Surfaces are implemented in the model by Markov Random Field, which enforces continuity, except where interrupted by object contours.
Lateral connections are a predominant feature of the visual cortex. They have been amply observed and documented by research.	Lateral connections are implemented by pool variables being connected by factors that enforce compatibility between the choices made in different pools.
Research has shown the visual cortex has excellent capacity for distinguishing between object instances, even when they are highly overlapping and/or transparent. This quality is known as top-down object-based attention.	The model implements top-down object-based attention by a combination of non-negative weights and lateral connections.
Neuroscience evidence shows the visual cortex makes use of message-passing-based approximate inference and learning.	Representational choices were beneficial for message-passing-inference: include feature-specific lateral connections and sparsity of weights.

Method:

In RCN, objects are modeled from **combination of contours and surfaces**. Surfaces are modeled using a **conditional random field (CRF)**. Contours are modeled using a **compositional hierarchy of features**. This factored and flexible contour and surface representation enables the model to classify and detect various objects without training on every surfaces and contours combination. RCN is a hierarchical model and this hierarchy fulfils two roles: 1) enabling the representation of deformations through multiple levels; 2) efficient sharing of features between different objects.

While most **CNN models** developed in recent years, first, analyze whole images, and second, presume very little knowledge about the images and objects in them, RCN aims at modeling the contours and surfaces of objects and background in the image in a direct way.

The figure below illustrates the structure and internal workings of RCN. On the left is a toy example of a three level RCN representation of a square, with level 1 representing lines, and level 2 representing the four corners -- you can see the lateral connections within the network, as well as the pooling layers between levels of the hierarchy. On the right is a realistic four level RCN representation of the letter "**A**", built with the same structure.



The implementation consists of two steps:

1) PreProc -- a set of filters that extract low-level elements from the image.
2) Learns a hierarchical model from the data.

"RCN's performance suggests that incorporating insights from neuroscience can lead to highly data-efficient, generalizable and robust machine-learning models."

1. The PreProc (preprocessing) uses a bank of Gabor-like filters to extract edges from pixel values. It is structured as a pipeline consisting of 4 steps, which we will go into now.

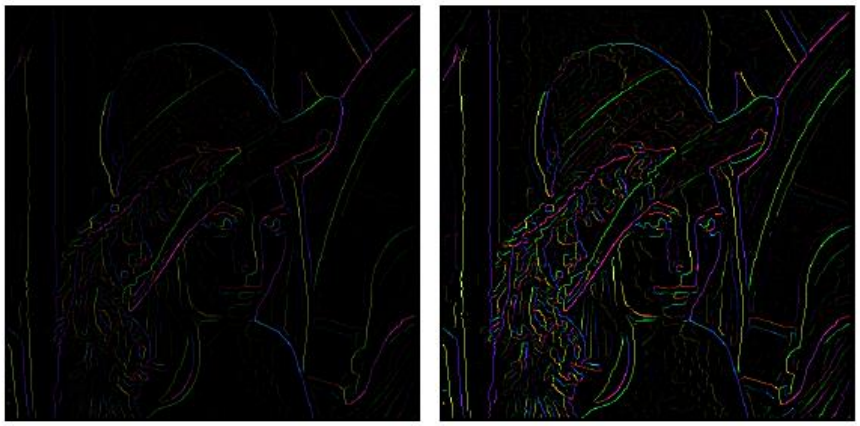
1.1 2D Correlation - A predefined set of grayscale edge filters using 16 edge orientations, where the filters consist of a positive and a negative Gaussian.

1.2 Localization - A variant of the Canny edge detection is used to prevent replacement of the evidence corresponding to a single edge.

1.3 Cross-channel suppression - Cross-channel suppression is computed (using Fpooled - for the exact equation see article) to overcome the risk of minor noise causing an image edge halfway between two filter orientations being suppressed.

1.4 Conversion to log likelihoods - Normalizes the maximum brightness, such that edges with high contrast do not have a higher score, while edges with very low contrast will have a brightness proportional to the score of the filter.

(a) Input Image (b) 2D Correlation (c) Localized



(d) Cross-Channel Suppression (e) Log Likelihood

2. Learning a hierarchical model from the preprocessed data.

The algorithm constructs the RCN representation layer-by-layer, from the bottom up. The model needs to represent the learning of two components: a) features and (b) lateral connections.

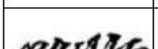
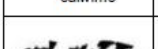
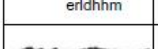
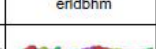
- **Features** - Unsupervised dictionary learning and sparse coding are used to extract contours from a training set of the preprocessed images produced by step 1. Both Intermediate-level and Top-level feature learning takes place during the Features step.
- **Lateral connections** - Pooling layers are learned from the input contours independently.

The learning process then uses forward and backward message-passing in the network to identify objects. It finds a complete MAP approximate solution by solving the scene-parsing with forward and backward pass. The upper bound on the log-probability of the top-level nodes is given by the forward pass, while the backward pass calculates the approximate MAP based on the forward pass' high scores. An additional outcome of the backward pass is the rejection of many of forward pass' false object hypotheses.

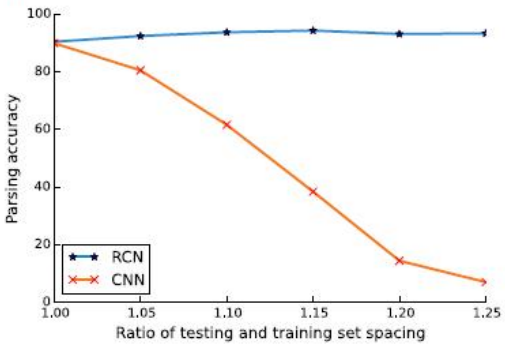
Results:

For a CAPTCHA to be considered broken, it's enough that it can be automatically solved at a rate above 1%. The authors' RCN method effectively broke a variety of different text-based CAPTCHAs using small training datasets and no CAPTCHA-specific heuristics (figure below), at an accuracy rate far exceeding this threshold. It achieved a character-level accuracy of 94.3% in solving reCAPTCHAs (which translated to a 66.6% accuracy rate for solving entire reCAPTCHAs), 64.4% for BotDetect, 57.4% for Yahoo and 57.1% for PayPal, rates incomparable to the 1% at which CAPTCHAs are considered ineffective.

A

Input & ground truth	RCN top parse	RCN second parse	Human labels
 calwlme	 calwime	 calwlme	canmme caiwime
 erldhhm	 erldbhm	 erldhhm	erldnhm erldhhm
 esterrow	 esterrow	 estenow	esterrow esterrow

B

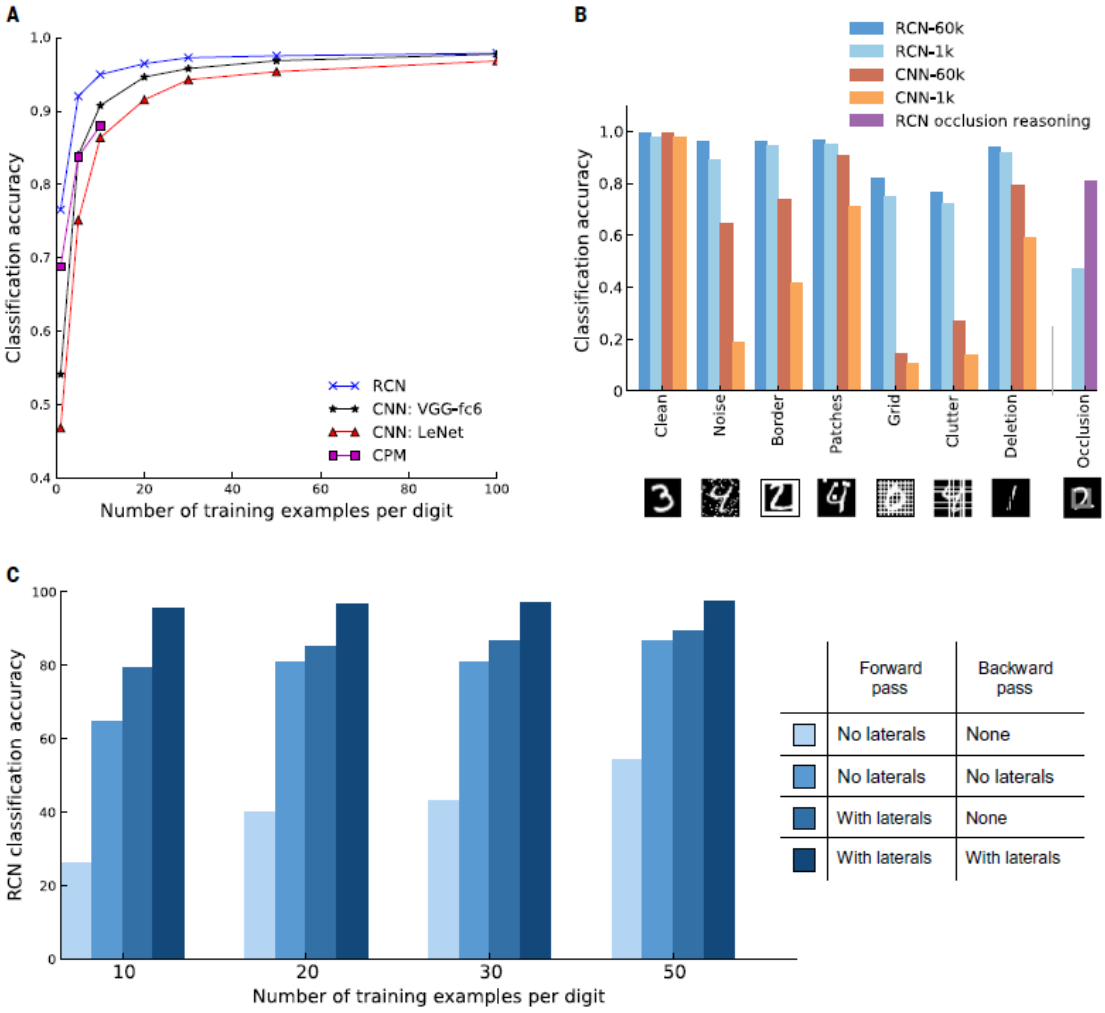


C

CAPTCHA name	Examples	Word accuracy	Character accuracy
reCAPTCHA		66.6%	94.3%
BotDetect		64.4%	91.6%
Yahoo		57.4%	92.5%
PayPal		57.1%	89.3%

D

(A) Representative reCAPTCHA parses showing top two solutions, their segmentations, and labels by two different Amazon Mechanical Turk workers. (B) Word accuracy rates of RCN and CNN on the control CAPTCHA data set. CNN is brittle and RCN is robust when character spacing is changed. (C) Accuracies for different CAPTCHA styles. (D) Representative BotDetect parses and segmentations (indicated by the different colors).



Comparisons against the compositional patch model (CPM) and CNN (Fig above) demonstrate RCN's advantage. RCN's recognition performance was 76.6% versus 68.9% for CPM and 54.2% for VGG-fc6. RCN proved robust to various types of clutter introduced for testing, without special training to learn those transformations. CNNs' generalization performance drops significantly when such out-of-sample test examples are introduced (B in the above figure). A lesion study was conducted in order to isolate the contributions of the forward and backward lateral connections: results, summarized in (C in the above figure), show that these lateral connections significantly contribute to RCN's performance.

Summary

RCN's performance suggests that incorporating insights from neuroscience can lead to highly data-efficient, generalizable and robust machine-learning models.

MobileODT is a company that has developed hardware and software to help detect and prevent cervical cancer in women by using smartphone technology.

"Cervical cancer accounts for a quarter of a million deaths every year"

This solution utilizes low cost hardware together with the power of mobile phones, **making medical systems accessible anywhere in nearly any condition**.

As the fourth most common cancer for women, cervical cancer accounts for a quarter of a million deaths every year, with 87% of those occurring in underdeveloped parts of the world such as Eastern and Central Africa. In the United States and other developed countries, doctors screen for cervical cancer by performing pap smears and HPV testing. However, in parts of the world lacking the resources and lab infrastructure, clinicians rely on visual inspection with the naked eye to screen cervical cancer.

Cancer screening in parts of Africa involves using a speculum and a long swab to apply a thin layer of acetic acid, essentially vinegar, on the cervix. Then, physicians use a headlamp or flashlight to observe the cervix through the vaginal canal to see if an area has turned white. If it has, the patient tests positive and receives immediate cryotherapy treatment.

David Levitz is **MobileODT**'s co-founder and CTO. He explains the ways to treat cervical cancer holistically. First, large scale screening efforts have an ability to treat the patients that test positively. Vaccination and patient education are also part of the solution. MobileODT focuses on the former treatment, screening and treating women at risk.

David and his team built a device based on smartphone technology that helps nurses **screening for cervical cancer in low resource settings**. Instead of judging with the naked eye, with the EVA System, clinicians can use a phone equipped with a lens, case, light, and app to look at the cervix of the patient. This gives them the ability to magnify and record the image before uploading

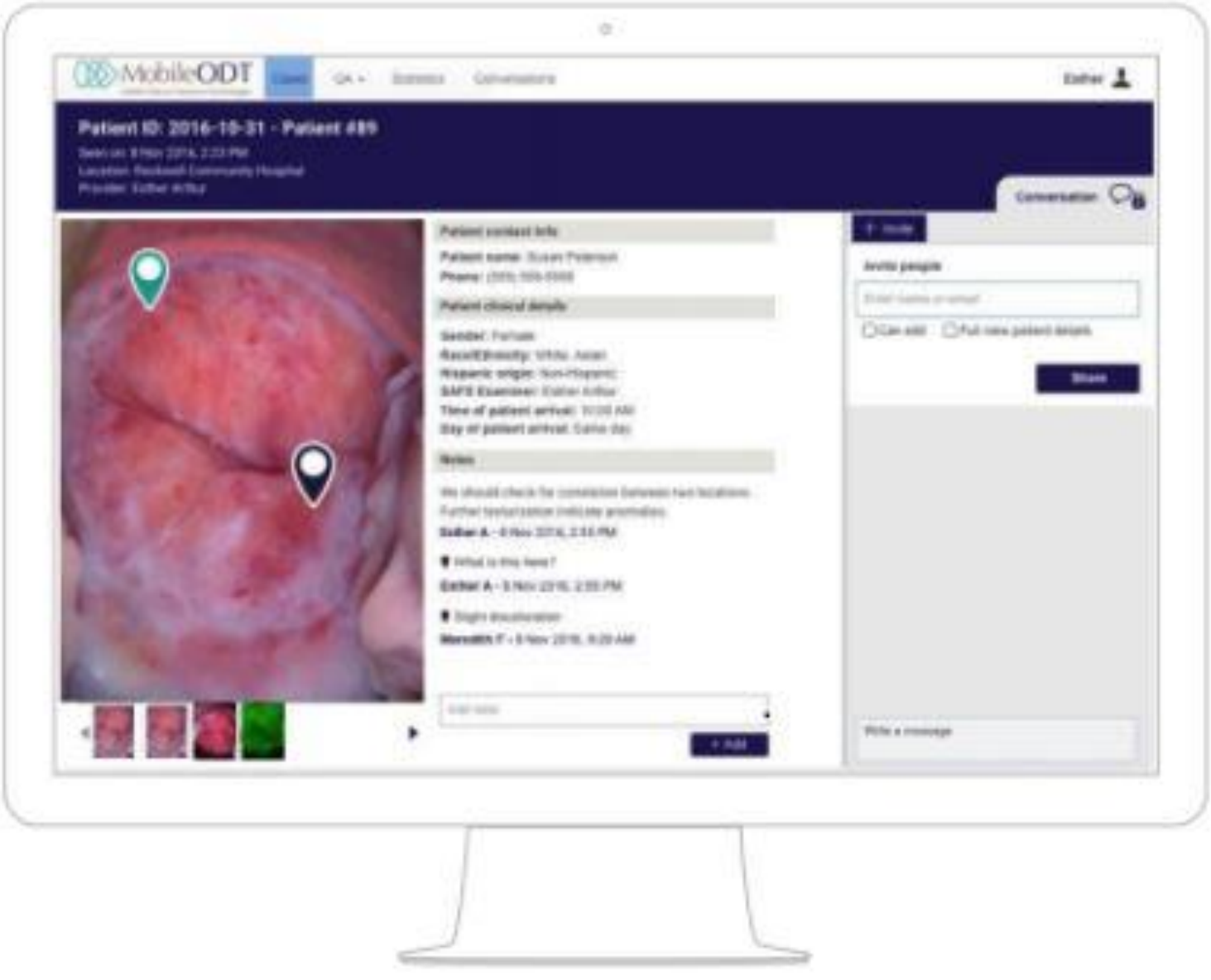
to secure online storage.

On the image portal, MobileODT has a primitive electronic medical reference system. The nurses can also record the decision they made at the point of care for later analysis. It provides simplified electronic medical records in low-resource settings lacking patient files or image portals to store information.

David explains that their main challenge in developing the software was meeting **HIPAA compliance**, the US privacy law which requires saving information from the clinician-to-patient session in a specific way so personal health information remains private. They needed to find a solution so that nurses could keep a record of the images and labels.

To solve this problem, they save pieces of information in different places on a secure online backup. Personal health information goes in one place while the image itself is stored on a second place. Meanwhile, the information on the provider is stored in a third place. **This solves the problem of privacy**. Although MobileODT doesn't have access to the personal health information, the clinician can still see and review the patient data.

Since the FDA does not want to regulate smartphones, MobileODT met challenges in obtaining **FDA clearance**.



Application

Typically with FDA filings, they need to know about every changing component. A smartphone makes that more difficult. In the end, MobileODT became the first smartphone-connected imaging device that received FDA approval.

The project still remains in its early phase, and MobileODT currently does not have a computer vision product. However, they have several projects on computer vision happening in parallel including a recent **Kaggle challenge** sponsored by **Intel**.

It must be said that a certain fraction of women has an extra fold in their cervix, making them untreatable. Even though this only occurs in 2% of the patients, treatment could potentially kill these women. They want to **flag these patients** so that the nurses at the point of care refer them to an expert rather than move forward with treatment right away. The Kaggle challenge, which ended a few months ago, involved identifying those patients. They want to take the information from the challenge and gradually turn it to something that can run on a mobile platform.

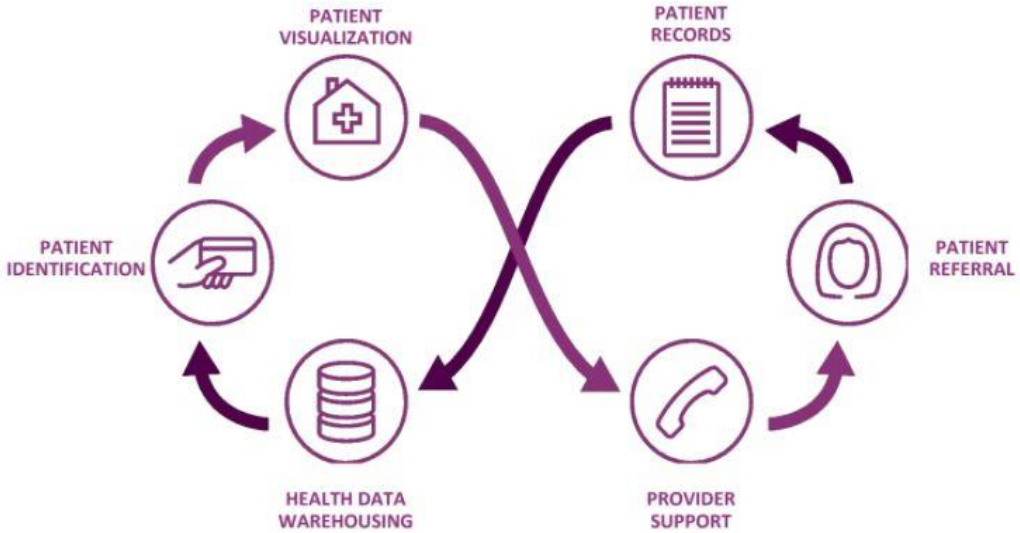
The various winners and their solutions took many approaches trying several different algorithms. Each of them had an incremental bump, and while no algorithm solved everything magically, together they performed quite well. The challenge validated that the methods work, although they have a partner that is still trying to figure out how to optimize these methods and put them on a mobile platform.

MobileODT has an ongoing project with **Global Good**, an organization funded by **Bill Gates** and **Intellectual Ventures** to give successful engineering ideas the chance to contribute to mankind. In this case they try to develop an algorithm that will allow nurses to detect cancer at the point of care.

In terms of computer vision, the project requires several things. First, they need to make sure that the images are good enough quality so that the algorithms can work well on them. Some of their users have never used a smartphone. They have to teach nurses how to use text messages and then the camera, in order to improve their ability to take high-quality images with the device.

The Global Good project developed a **focus score**. After a user takes an image, the focus score instructs them whether they need to take another image, and what to improve.

"MobileODT became the first smartphone-connected imaging device that received FDA approval"



This feature could also alert a supervisor if the nurse needs help with certain manual aspects of the procedure. In addition to this, they want to take annotations and refine them.

"A huge variability exists in the proficiency of the nurses screening for cervical cancer"

A huge variability exists in the proficiency of the nurses screening for cervical cancer. Some nurses get the answer right most of the time while others only get it a small part of the time. This makes it difficult to know which labels to trust. Labeling the data to train a set later makes it even more complicated. They need to actually take the images and send them through a sort of **QA process**, not only for image quality, but also for the decision that was made. They need to evaluate both aspects of the data before using the annotation and the label for machine learning.

The third issue is that the clinicians do not always trust another clinician's opinion, but they also refer to the golden standard of taking a biopsy. It is difficult to perform a biopsy in low resource settings for the same reason that it's difficult to do a pap in a low resource setting. It requires some infrastructure, a lab, and someone to process the tissue. The tissue goes then to a lab where it is dehydrated and put it into a paraffin. Then, they slice small sections of it and add ink before sending it to a pathologist. This pathologist does another subjective step that also needs automation. **That whole process does not exist in some countries in sub-Saharan Africa**.

MobileODT has an office in Nairobi where David learned that although East Africa has several hospitals, they really only reach a small fraction of the patients. With a pathologist, that problem gets even worse. A country like Botswana has only four pathologists to handle every disease. This creates an incredible shortage in solving these kinds of problems.

David believes that **AI has amazing potential** to take on some of these challenges. It could give nurses at the point of care an answer or indication, and even tell them exactly where to look. Even something like that would make a huge impact in healthcare. David insists: "*We want to save as many lives as quickly as possible. That's why we picked cervical cancer and other diseases because that's one way we can do just that!*"

No matter where they live, women around the world care about their health and want the ability to monitor themselves and stay healthy in order to take care of their children and family. "*We're here to actually help bridge the gap* - David concludes - *because there are 5 billion people with access to smartphones, but without access to a physician. That's just unfortunate, and we need to address it.*"

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!

Jürgen Schmidhuber: A New Type of Life is Going to Emerge from our Little Planet and Colonize the Entire Universe!

Our readers remember the fascinating interview that Professor Schmidhuber gave to Computer Vision News. Now you can hear his intriguing (and a bit scary) vision of future from his own voice in this exceptional **TED^{x} Talk**: don't miss it! **Click on the image to watch the video!**

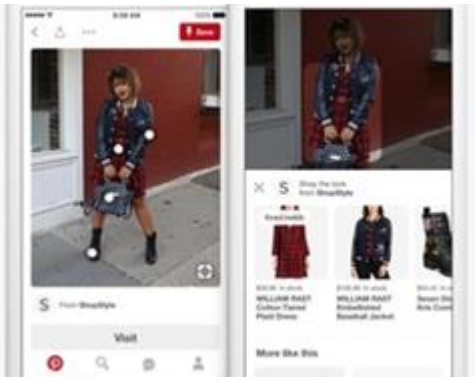
Haptic Technology: interact with the virtual world just by touch

A video recommended by **Michael Black** is one we cannot miss. Follow **Alexandra Cardinale's** infectious enthusiasm, while she explains with gusto how a virtual sense of touch enables humans and robots to interact with each other. It's called **Haptic Intelligence** and you must see what **Katherine Kuchenbecker** is doing with it at this newly established department of the **Max Plank Institute** for Intelligent Systems. **Click on the image to watch the video!**

One year in Computer Vision - by the M Tank

We wanted to review for you the exceptional work done by researchers in our community during **2017**. And then we saw that somebody did that already - an they did it very well. Have a look at this nice report by **Benjamin F. Duffy** and **Daniel R. Flynn**. We are not sure that everyone knows all the models which are presented here. **Read Now...**

Pinterest Sees The Future with Lens



Are you a **Pinterest** fan? We are not really into that, but they have 200 million monthly users and their **new AI features** are kind of cool. Take a picture of anything you want and ask Pinterest to suggest where to find it or what you could do with it. Let Pinterest discover ideas for your actual life: for instance, **Lens Your Look** tells you new ways to wear clothes you already own. **Read Now...**

New Sensor Gives Driverless Cars a Human-Like View

It's a new hybrid sensor combining a camera, solid-state Lidar, and onboard image processing to change how it sees in response to its surroundings. The startup is called **AEye**. **Read Now...**

Do you like reading about fooling tools against AI recognition?
Here is how to turn a pair of glasses into a sci-fi AI-fooling backdoor: Read Now

How to Build the Development Team

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 **How to Build the Development Team**, another tip for **Project Management in Computer Vision**.

"Transparency and openness are necessary in any organization"

The project manager needs much knowledge and experience to build the optimal development team in view of the R&D mission.

Many aspects of this task are common to all projects: today I will talk only about those which are specific to computer vision R&D.

Diversity of the team: since computer vision and artificial intelligence borrow many theories from a multitude of branches in exact sciences, it is recommended to build a team with members having heterogeneous backgrounds, able to take ideas from mathematics, physics, graphics, signal processing, computer science and of course deep learning. The combination of diverse sources might be the key to good results, since best practices can be chosen from an wide choice of activity fields offering an ample array of solutions.

Recruiting to the team: when you add new members to the team, you want the integration to be as quick as possible. Give them real tasks starting on day one, in order to enable them to learn and grow at the same time as they contribute their talent to the team. The project manager should find the optimal task with which the newcomer can start work and be integrated to the team, before getting into speed with other, more elaborated tasks.

Transparency and openness are necessary in any organization, all the more so in R&D and development. This is because research is a risky activity and experts should collaborate to give the proper aid whenever it is needed. When this happens, no time and resources are wasted, knocking your head against the wall for want of a solution.

In fact, any time a difficulty arises, the person noticing shares it as needed, reducing the time elapsing until a solution is found and implemented. Thus, when discussing algorithmic tasks in the project, transparency and open-mindedness play a crucial role in the building of the team, as well as in its performance.

"Build a team with members having heterogeneous backgrounds"

Visual Question Answering (VQA) networks

by Assaf Spanier

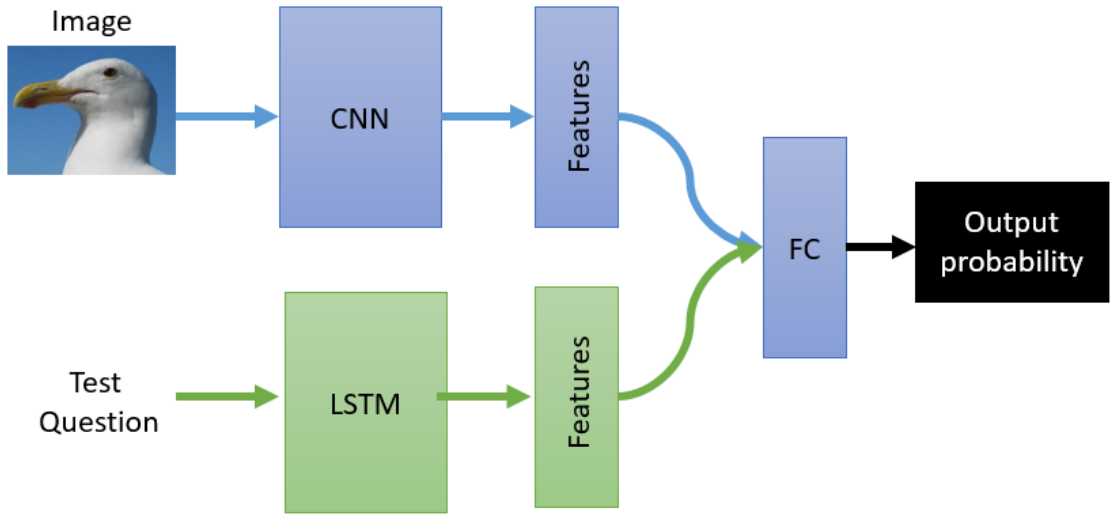
This month's we'll be looking at **Visual Question Answering (VQA)** networks. We'll talk about their overall structure, the type of input they get, typical databases used for training nowadays, and we'll present an implementation in Keras and some of the latest results from benchmark competitions in the field.

"Since its introduction in 2015, the VQA dataset of Antol et al. has been the de-facto benchmark"

First, what is **Visual Question Answering**?

A visual question answering (**VQA**) task includes an image and a free text question about the content of the image, where the goal is developing a model that is able to automatically answer the question correctly for that image. Models in this field combine computer vision, natural language processing and artificial intelligence. Since the questions tend to be visual and refer to specific areas of the image, such as background details, context, etc., VQA problem solving methods require much better and more detailed comprehension of the image, its content, logic and internal structure, compared with methods that just need to estimate an image's overall content to label it.

The models developed in recent years adopt the following overall structure: The network takes in input through a two-branch structure (see figure below) -- one branch for the image and one for the text of the question. The image branch uses CNN to extract image features, for example the activations from the last hidden layer of VGGNet. The text branch uses LSTM-type networks, taking the internal state of the LSTM as features. At this point both branches' outputs are combined in a single FC layer, whose output are the probabilities predicting the answer.



We'll present two similarly-structured, simple, elegant models for handling VQA, and show their implementation in Python, using the Keras and the torch deep learning libraries. Both are very easy-to-use deep learning libraries and written in Python.

Both models were developed to compete in The VQA Challenge (for more on the competition, see link). The first was developed back in 2015, based on a preliminary dataset prepared at that time. A number of problems arose: in retrospect it turned out that a non-negligible number of questions could be answered without referring to the image, such as "*what color is the sky in the image?*" These realizations led to the development of a new updated dataset in 2017, which we will describe in more detail below. We will also present an up to date deep learning VQA model developed this Summer, that achieved the best results on the new dataset.

"Since its introduction in 2015, the VQA dataset of Antol et al. has been the de-facto benchmark"

Dataset:

Several datasets have been published for VQA (for a survey, see Visual question answering: A survey of methods and datasets). Since its introduction in 2015, the VQA dataset of Antol et al. has been the de-facto benchmark.

To arrive at a set of good quality questions, the dataset's developers ran studies asking subjects for questions about a given image, that they believed a "**smart robot**" would have trouble answering.

However, in the aftermath of the 2015 benchmark competition, they discovered that many of the questions laymen proposed, such as "*what color is the cat?*" or "*how many chairs are in the scene?*" are too simple -- requiring only low-level computer vision knowledge. The VQA dataset developers' goal had been questions that require commonsense knowledge about the scene, like "*what sound does the pictured animal make?*" Furthermore, questions should also require the image to be correctly parsed and not be answerable using just commonsense, like "*what is the mustache made of?*" which combines identifying a location within the image with object identification.

This led to the development of the VQA v2 dataset, in which every question has two images associated with it. The images selected purposely so each leads to a different answer, discouraging blind guessing from the question alone. This new dataset was the basis for the 2017 VQA challenge.

The model of Agrawal et al. (2015):

1. Image features: the input image is passed through a Convolutional Neural Network (VGGNet) to obtain L2-normalized activations from the last hidden layer.
2. The text question features: 2048-dim embedding for the question is achieved by an LSTM with two hidden layers, followed by a fully-connected layer with tanh as non-linearity. The fully-connected transform of the 2048-dim embedding to 1024-dim (the question words themselves are encoded in the same way as in LSTM).
3. This image + question embedding are then passed to a fully connected neural network classifier with 2 hidden layers and 1000 hidden units. Dropout 0.5 layers is used between the hidden layers. Lastly a softmax layer is used to obtain a distribution over K answers.

The entire model is learned end-to-end using a cross-entropy loss. VGGNet parameters are those learned for ImageNet classification and not fine-tuned in the image channel.

We'll look at the core parts of the model (the full implementation can be found here, use the file model.py). This model uses three functions. The first (Word2VecModel) embeds the question text and extracts relevant features from it. The second (img_model) extracts image features. The third (vqa_model) fuses the text and image features into a single features vector, and adds an optimizer layer and fully connected layers.

1. The Word2VecModel function is built on a Keras sequence model that first encodes the question using the embedding function. The encoding is fed into 2 LSTM layers with a dropout layer between them -- all this is implemented in the short code snippet at the beginning of next page -- the beauty of the Keras library is that the code is simple, clean, minimalistic and self-explanatory.

```
def Word2VecModel(embedding_matrix, num_words, embedding_dim, seq_length, dropout_rate):
    print("Creating text model...")
    model = Sequential()
    model.add(Embedding(num_words, embedding_dim,
        weights=[embedding_matrix], input_length=seq_length, trainable=False))
    model.add(LSTM(units=512, return_sequences=True, input_shape=(seq_length, embedding_dim)))
    model.add(Dropout(dropout_rate))
    model.add(LSTM(units=512, return_sequences=False))
    model.add(Dropout(dropout_rate))
    model.add(Dense(1024, activation='tanh'))
    return model
```

2. The img_model function initiates a placeholder layer for the image weights (features). Actually, the code below doesn't implement the entire CNN network (in this case VggNet), the way this model was implemented, the network was pre-trained on all the images, and at this point the layer is prepared for loading the weights determined in pre-training.


```
def img_model(dropout_rate):
    print("Creating image model...")
    model = Sequential()
    model.add(Dense(1024, input_dim=4096, activation='tanh'))
    return model
```

3. The final function of the model, vqa_model, fuses the features extracted from the text and image into a single features vector, and adds 2 fully connected layers as well as setting the optimizer for training the network.

```
def vqa_model(embedding_matrix, num_words, embedding_dim, seq_length, dropout_rate, num_classes):
    vgg_model = img_model(dropout_rate)
    lstm_model = Word2VecModel(embedding_matrix, num_words, embedding_dim, seq_length, dropout_rate)
    print("Merging final model...")
    fc_model = Sequential()
    fc_model.add(Merge([vgg_model, lstm_model], mode='mul'))
    fc_model.add(Dropout(dropout_rate))
    fc_model.add(Dense(1000, activation='tanh'))
    fc_model.add(Dropout(dropout_rate))
    fc_model.add(Dense(num_classes, activation='softmax'))
    fc_model.compile(optimizer='rmsprop', loss='categorical_crossentropy',
        metrics=['accuracy'])
    return fc_model
```

The model of Tendy et al.:

The model described so far was developed for the first dataset from 2015 and published 2 years ago. Now let's turn to the model that won the competition last summer, by Tendy et al. (link). Its code too can be find in the model.py file. This model is implemented in torch. This method achieved better results than those of the 2015 competition on the improved (more difficult) 2017 dataset. We'll see that the model follows the same basic structure, but by implementing some minor changes it achieved improved performance on a task made more difficult.

The input for each instance - whether during training or test time - is a text question and an image.

1. Image features: the input image is passed through a Convolutional Neural Network (CNN) to obtain a vector representation of size K × 2048, where K is the number of image locations. K is computed based on a ResNet CNN within a Faster R-CNN framework. The resulting features can be thought of as ResNet features centered on the top-K objects in the image. It is trained to focus on specific elements in the given image, using annotations from the Visual Genome dataset.

```
# image encoding
    image = F.normalize(image, -1)  # (batch, K, feat_dim)

    # image attention
    qenc_reshape = qenc.repeat(1, self.K).view(-1, self.K, self.hid_dim)
    concated = torch.cat((image, qenc_reshape), -1)
    concated = self._gated_tanh(concated, self.gt_W_img_att, self.gt_W_prime_img_att)

    a = self.att_wa(concated)
    a = F.softmax(a.squeeze())
    v_head = torch.bmm(a.unsqueeze(1), image).squeeze()
```

2. The text question features: the given question is split into words using spaces and punctuation. Only the first 14 words of each question are used (almost no loss is involved, since only 0.25% of questions in the dataset are longer). Each word is turned into a 300-dimensional vectors learned along with other parameters during training. The resulting sequence of word embeddings of size 14 × 300 is passed through a Recurrent Gated Unit. The question embedding is the 512 dimension recurrent unit internal state.

```
# question encoding
  emb = self.wembed(question)
  enc, hid = self.gru(emb.permute(1, 0, 2))
  qenc = enc[-1]
```

3. Here too, as in the first model, the final set fuses the features extracted from the text and image.

```
# element-wise (question + image) multiplication
    q = self._gated_tanh(qenc, self.gt_W_question, self.gt_W_prime_question)
    v = self._gated_tanh(v_head, self.gt_W_img, self.gt_W_prime_img)
    h = torch.mul(q, v)

    # output classifier
    s_head = self.clf_w(self._gated_tanh(h, self.gt_W_clf, self.gt_W_prime_clf))
```

In all cases, the CNN is pre-trained and held fixed during the training of the VQA model. The features can therefore be extracted from the input images as a preprocessing step for efficiency.

Results:

For the purpose of comparison, to have a taste of the results, let's look at the comparison between those models' performance on the VQA v2 (you can find more results and analysis in the links below).

	VQA v2 test-std			
	All	Yes/No	Numb.	Others
Antol et al.	54.22	73.46	35.18	41.83
Tendy et al.	70.34	86.60	48.64	61.15

Code and installation instruction:

1. The first model:
   https://github.com/anantzoid/VQA-Keras-Visual-Question-Answering
2. The second model:
   https://github.com/markdtw/vqa-winner-cvprw-2017

Raquel Urtasun is Head of Uber Advanced Technologies Group (Uber's self-driving vehicle lab) in Toronto and Associate Professor at the University of Toronto.

Raquel, what makes you successful?

I think that I am very persistent. I don't give up easily when I have difficulties. I think that's what makes me successful.

Where do you find the motivation not to give up?

I have a lot of will. I learned this by playing team sports.

Which sports?

I played basketball for 15 years.

How was it growing up in Spain? What were you like when you were younger?

I was a rebel, not surprisingly.

What would you like to change in this world?

So many things. I would like a world where everybody has a chance, and there is no poverty.

I think what is very interesting right now is that, not just computer vision, but AI can actually change a lot in the world. We have a chance to change it for the better. I think that's super exciting. From my side, I'm really excited about bringing transportation to everybody and building cities that are better to live in. I think this is just the beginning.

Let's talk about these.

Transportation is one of the fundamental things in smarter cities.

Why do you have so much passion for

"I'm in a position where all those things are coming together right now..."

photo: Chris Sorensen

the subject of transportation?

There are two things. On one side, it's really great that we managed to build self-driving cars. We're going to touch the lives of everyone. This has many, many interesting potential benefits.

Are we going to save lives?

There are 1.3 million people that die every year from crash accidents. Every little percentage that we can save actually means many, many people. And we have people who get wounded… And also people that don't have mobility. The elderly and people with disabilities cannot have a normal life. We can bring them that. Pollution is an issue, the air condition in our cities. If you think about it, 20% of our cities are actually dedicated to parking. Imagine how many trees, grass, and parks you could build with that landscape.

So you belong to an organization that's trying to push things forward in this direction?

That's correct. Uber is going to bring cheap transportation to everyone, transportation of people, transportation of goods and transportation of food. I really like the idea of creating a platform where everybody can benefit, not just the rich. That was one of the reasons why I was super excited joining Uber.

The next thing is cities. What makes you passionate about this?

Transportation is just one way to think of how to improve our cities. There are many things that we can do, from understanding how our cities evolve to being able to create energy efficient houses to understanding how we move in cities. There are many, many things that we can do for our cities.

Did you grow up in a city?

Yes. In a small city north of Pamplona where they have the running of the bulls. There is no AI per se, there is just bulls.

[*laughs*] **How many bulls did you kill?**

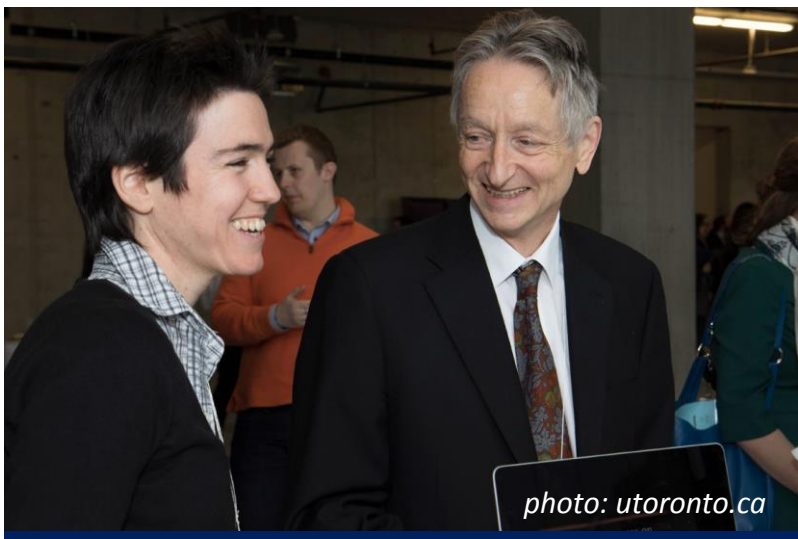
I didn't kill any. I don't want to kill the bulls. The bulls actually charge. People have to run, or they can be killed. I don't agree with bullfighting, but the running of the bulls, I agree with.

What was your main source of inspiration to go in the direction of improving our cities?

I have lived in cities my whole life. There is always this thought of "*what if we had this, what if we could change that?*" I think that now we have the tools to actually start to change those things and to make our lives better.

Can we expect big corporations to improve our lives?

Yes, of course. All of us citizens, we have a duty to the place where we live. The corporations should be the first ones to contribute to the government. It's not about a war between corporations and government. It's really about how government, corporations



Raquel with Geoff Hinton

and academia together can actually create a better place.

Do you feel more like a teacher or more like someone working at a corporation?

I don't feel like a person working in a corporation actually. I have a research lab in industry where I have most of my PhD students working. They are full-time employees and full-time students. I built my lab really as a copycat of what I had in university because I believe when you have a really nice environment, where you can be free to work on what you like, to wake up every day and think "*I want to know the answer to this*", that's how everybody should work.

When and how did you decide that would be your direction?

I think it was very clear from the beginning. Shame on me because actually Uber is my first job in industry, which I started eight months ago. It was very clear that I really always wanted to know things that we didn't know. That's the research spirit. I like the challenges.

What would you like your students to remember the most?

That's very interesting. As a PhD advisor, what one hopes for is that each student develops. They are like a diamond that needs to be polished. The hope is that you help polish so that when they graduate, they are at their peak

"I really like the idea of creating a platform where everybody can benefit, not just the rich."


thestar.com

Life · Fashion & Style

U of T scientists create software to analyze outfits

New program, which they hope to turn into an app, determines whether an outfit is stylish and offers suggestions.

University of Toronto researchers Raquel Urtasun, left, and Sanja Fidler are creating an app that assesses clothing and recommends how to be more fashionable. (STEVE RUSSELL / TORONTO STAR) | ORDER THIS PHOTO

for the beginning of their careers. At the same time, I also hope that I can teach them how to interact with peers in a way that is respectful and where everybody has a place.

One of your students told me that you are the teacher she learned from the most. Was there a time when you were really impressed by a student?

[*surprised*] My students are brilliant. They surprised me many, many times. I'm very proud of each one of them. I think as long as they grow as researchers and as people as well, that's what is really important to me. It's not really about my success. It's really about their success and how they can see their progression.

A friend of yours told me once that she's much happier when a student's paper gets accepted than when her own is. Can you relate to that?

[*both laugh*] Yes! It is a satisfaction. Sometimes papers are not the end goal. They're a way to show what you have done in a particular point in time. Students' progress is more important.

Yes, but seven papers with your name were accepted at ICCV2017. That must be saying something about the quality of your work.

Numbers don't matter. It's not about the number of papers that you write. If you write one paper where people remember what you did, then you have succeeded. That's what you should aim for. These days, students try to make the deadline, and there are so many deadlines with so much stress. That's just the problem. I think it's much more important to try to do one fundamental thing that will really put you on the map. Now it's hard to do that. You cannot always put all of your eggs in a single basket because that's very risky. You have to strike a balance between the two things.

"An uncomfortable number of hours in the lab..."

What do you miss the most about Spain, while living in Canada?

Food! [*we both laugh*] Well, I have lots of friends still there. What I think I miss the most is when I was living in Spain, I was young and having a good time. That's not the case anymore. It has nothing to do with Spain per se.

Why did you put so much on your shoulders and sacrifice all that?

When your work is the same thing as your passion, then it's really, really nice. It's really rewarding, but at the same time it's dangerous. Sometimes you just forget about the rest of the things in life. In my case, I still try to strike a balance between my personal life and my work. You can argue whether I succeed or not! [*laughs*] But I try my best. It's true that I spend an uncomfortable number of hours in the lab. I always tell my students that I expect everything from them, but I give them everything as well.

If I gave you back two hours a day for the last ten years, what would you do with all that time?

I would probably start a new research project! [*both laugh*]

If you didn't become a researcher, what would you have done?

I love sports and I'm actually pretty good at it.

Which position did you play in basketball?

I was a shooter, particularly a 3-point

shooter. I played when I was little, and I played in university.

Do you still play?

Ah, no. I broke my knee many years ago. I play soccer. I love soccer.

Football! The other word does not exist!

[*both laugh*] I've been in America for too long…

In which position on the field?

Anything. I don't like defense too much, but anything.

Do you score many goals?

Depends on the game. I play with my students in the summer weekly. I like doing that. I used to play with a team in Boston a while ago when I did my postdoc at MIT. Now I cannot commit to regular practices, but I definitely like sports a lot.

Team sports also can teach a lot of things. They are very useful for research, sacrificing for the team and always having team goals instead of individual goals. This is super important. I hope that my students learn the same.

"Then you can go really far!"

What did you learn in Canada that you couldn't learn in Europe?

Canada has a very good trade-off between the North American research style and the European style of living. Also Canada is very welcoming to immigrants. Toronto is a fantastic city and very diverse. I think 50% of the population is from a different country, and what they really appreciate is that you don't need to change the way you are. Actually, people like to see diversity and different opinions. They expect you to continue with your traditions, which is really interesting.

What does it mean to you when you discover new things every time?

I think it's a really interesting part about research: to do something that nobody has done before. I really drive my curiosity these days when I wake up in the morning, and I run to the lab to see what happened, to see the answers to some of the questions I asked my students, to see what is the new thing that we have done.

You have many drives: curiosity, passion and cooperation. At the same time, you have all of the things that you enjoy: a team that works together, people that care about each other, a society and corporations that care about people. How are you able to fulfill all of these goals and drives?

I don't know, but I'm in a position where all those things are coming together right now. It's not necessarily something that I planned. It just turns out to be.

You are a lucky girl!

But at the same time, it is important for everyone to see that the opportunities are there and not be passive about, but try to get these opportunities. When I grew up, I studied undergrad in a university that nobody knows about. I also had to grow from nothing to what I am today. It's just a matter of trying hard and discovering what you're good at, then have your passion and your expertise match. Then you can go really far!

What would you advise to a young student who does not have these opportunities?

Don't stay passive. Always look for opportunities. Without that, I would never have had many of the opportunities

I've had in the past. I think you have to be active and also try to ask around. Try to use every opportunity you have. If you go to a conference, meet people. Seek advice from people.

You gave three tips. One is networking, one is going to conferences, and one is seeking for advice.

In seeking for advice, if you have mentors, that's the best way. Now these days in conferences, for example, there is a nice setup. There are things like the Doctoral Consortium where you can ask for some feedback from somebody that you'd like to work with or if you value their opinion. People are very approachable. I have to say that I'm giving this advice, but I was not very good at doing this when I was younger. I was just very shy.

What is it like to be a female in a male dominated field? Do you think females deserve a better chance?

To start with, everyone deserves a better chance, in the sense that everybody should have chances, the same chances. Now, it's not the case, right? It depends on your upbringing, where you were born, how rich your parents were, all sorts of things. As a female, it has not been easy, and it's still not easy. Unfortunately, we see a lot of examples of discrimination throughout our lives, and it's much more common than people think. What I always say is that we shouldn't give up. If we give up, we will never change this. It's really about example. We should show the world that women are actually as valuable as men.

Did you ever feel discriminated against?

Absolutely! Many times…

Can you give us examples?

An interesting one is discrimination from the press. When an article is written on some topic, and both males and females are interviewed, typically only the males get to the final edition. If the females gets there, it's never with the proper titles nor at the beginning of the article. There are many ways to impose a bias. It's subconscious most of the time. It's not necessarily on purpose. Then the immediate reaction is that there are not enough women that can appear on this topic, but that's simply not true. That's one example.

photo: Johnny Guatto

What do you do when this happens? Do you fight back?

Sometimes. I think these days, we're seeing the power of social media where if something really bad happens, and you want people to know about it, you can actually share it. People are actually going to share that experience. Suddenly, the public is aware of what it means to be a woman. I have seen many, many cases out there.

That happens to men too!

Let me give you one more example:

when they introduce the committee, it's Professor/Doctor Something. Then they arrive to the female faculty and suddenly it's Mrs Blahblah.

Was she a professor or doctor?

Of course she was a professor or doctor, but suddenly it's Mrs! This has happened a lot to me: they don't introduce you as professor or doctor. Yes, it's very typical that they suddenly remove the titles.

Do you correct that?

In these particular cases, I do not do it. I always regret it. Depending on certain settings, you also think about the consequences of you making a scene that aren't necessarily as good as what you are going to cause. The important thing is creating awareness about this. Then people might change.

Can we go back to the subject of confidence? You told me recently that it's not easy to be a woman. But here you are such a strong example of success. Where does the lack of confidence come from?

I think almost all women suffer from impostor syndrome. I am just one of them.

Where does it come from?

I don't know where it comes from: if we knew, we would have solved one of the biggest problems.

When we tell you that you are perfectly okay, do you trust the person who tells you?

That's the thing, right. We're just insecure.

You prefer to hear all of the negative rather than listen to the positive?

Even if insecure, we all need to fight our impostor syndrome! This is how I've managed to be successful, by never giving up, no matter how unconfident I felt.

Give us a tip on how to fight it.

Think about the other people that you know and their abilities. Think about what they would be doing in your case. Then you will see the world from a very different perspective because the glasses that you put on to see others are not the same that you use to see yourself. Maybe that will help you overcome this lack of confidence.

How can others give support and help

photo: Erica Edwards @Uber

"We all need to fight our impostor syndrome! This is how I've managed to be successful..."

someone feel more confident?

If you see somebody around you that is lacking confidence, you can very subtly showcase all the things that they have achieved. That might help put things in the context of other people.

I think that a lot of the education of women when they are young is about being pretty and not necessarily about being successful. We should change that right away. We really see these things with kids when they are 3 or 4 years old. We should stop it! We should go to the kindergarten or wherever, and stop it!

Now, it happens a lot that someone is actually really good, but because of a lack of confidence, they say, "*I'm not good enough. I shouldn't apply for this job because I'm going to be rejected*". I think that people around them should help by discussing it and putting into perspective what one has to achieve, so that they can actually gain the confidence necessary to try to reach their goals in the first place. Sometimes a little push can bring us really far.

"It takes so little. If everyone was like this, the world would be a better place..."

What can good friends do to help one feel stronger?

I think that it's not just about friends. The environment in the group is also very important. Everybody should be friends at work. I think that we should remove barriers, and everybody should be treated the same as a researcher, independent of their gender, religion, point of view, whatever. When somebody lacks confidence, regardless of any of these, we should just help that person feel better. While at the same time, what is also important is that people do not become arrogant, which is an issue as well. This happens also sometimes with good students who are very young. We should also open their eyes to the fact that this is not how the world is going to work. You should value your qualities as much as everybody else's qualities. It's important to have a balanced environment.

"Look around you! What are they feeling?"

Can you give us one take-home tip?

Improve your empathy. Look around you. What are they feeling? Given that, try to help. This is a tip for at work, at home, with your friends. This is the most important.

If we were better to one another then this world would be much better.

Yes, and it takes so little. If everyone was like this, the world would be a better place.

Augmented Reality - AR

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 approach to Augmented Reality and our experience in AR projects**. RSIP Vision's engineers can assist you in countless application fields. **Get effective R&D consulting from our experts now!**

"Many problems in the field of vision can be solved using Augmented Reality techniques"

Many problems in the field of vision can be solved using **Augmented Reality techniques**. We shall try to explain the meaning of this and what kind of problems can be solved through **AR**.

The task of Augmented Reality is to **overlay rich media on a picture taken in the real world**. This overlay is virtual and it is designed to create or emphasize virtual objects in a real environment, with the goal of generating an interaction between the user and these objects.

The main problem to be solved when dealing with augmented reality tasks is understanding the real environment: for instance, scene understanding and detection of the surfaces on which the virtual objects will be overlaid, without compromising the veracity of the generated scene.

Real objects must be correctly detected and tracked as well. This is particularly true in medical applications: when we apply an overlay on the patient, we need to track the region which will be examined or operated. When surgeons use AR glasses, we need to track their gaze as well, as it moves around the scene. In this case, tracking follow the surgeons eyes and the scene can change dramatically, as a function of the gaze moving.

A continuous registration process is necessary to perform this tracking: an object which is detected in a specific frame of the scene (and upon which the overlay is displayed) needs to be registered again in the following frame.

The team of **computer vision experts at RSIP Vision** has many years of successful experience in solving problems of **detection**, **registration** and **tracking**: we run several projects in ophthalmic imaging, accurately tracking eyes and pupils in real time. We also developed general applications of augmented reality for user guidance, guiding users according

to their need.

In addition to experience in the field, RSIP Vision offers clients its technical proficiency: for instance, solving registration problems using models and algorithms which allow to keep track of the interesting object in the following frames. Another registration technique is performed through feature extraction and tracking accompanied by machine learning algorithms: in this way, the software learns to adapt to any changes in the shape of the object due to the shift in gaze. The role of machine learning in this case is that of, given the registration of the following frame, learn the features seen in the new scene and use them in the next registration.

Since a virtual object is located or moves in the scene following the coordinates system of the real object, tracking of this real object plays a significant role in **AR applications**.

Tracking is an easy task when the tracked object has markers, but this case is often unfeasible or not practical.

Tracking of markerless objects can be model-based or image-based: state-of-the-art methods utilize Machine Learning approaches for ongoing learning of the tracked object's shape, since it could change during the video. The main Machine Learning tracking approaches are two: a) CNN-based and b) Kernelized Correlation Filters - KCF-based. State-of-the-art CNN-based approach includes **Recurrent Neural Networks** for best absolute object location prediction and Reinforcement Learning for tracking instructions.

"Software learns to adapt to any changes in the shape of the object due to the shift in gaze"

Algorithms can also offer a solution in cases of partial occlusion of the tracked objects, boosting the reliability of the resulting view.

Another key issue is dealing with changes in distance between the camera and the object: proportions need to be respected if we want to correctly identify the object in the new frame. That means that the algorithm needs to understand the dimensions of the objects in the scene, so that each can be rendered in the proper way.

"Proportions need to be respected if we want to correctly identify the object in the new frame"

NIPS'17: Learning to Run Challenge

by S.P. Mohanty

Every month, Computer Vision News reviews a **challenge** related to our field. If you do not take part in challenges, but are interested to know the new methods proposed by the scientific community to solve them, this section is for you. This month we have chosen to review the **NIPS'17: Learning to Run Challenge**, organized around NIPS 2017. The website of the challenge, with all its related resources, is **here**. Read below what **Sharada Prasanna Mohanty**, a Doctoral Assistant in the team of Marcel Salathé at the **Digital Epidemiology Lab of EPFL in Geneva, Switzerland** and one of the organizers, tells us about this **Learning to Run challenge**. We would like to mention co-organizer **Lukasz Kidzinski** from the lab of Scott Delp, who was also a major driving force from the **Stanford** team.

The **Learning to Run challenge** came together because we wanted to explore the feasibility of using deep reinforcement learning on high dimensional biomechanical systems to "learn" complex tasks like walking and running. The participants were provided a **human musculoskeletal model** in a physics-based simulation environment (**OpenSim**), and the task was to **design a real-time controller to navigate through an obstacle course as quickly as possible**.

Our initial baseline solutions based on **DDPG** and **TRPO** showed that it is indeed feasible to learn policies which are represented by feasible gaits and also good cumulative rewards (distance covered by pelvis). The challenge managed to attract a total of **2154 submissions from 471 participants all over the world**. Some of the top submissions were capable of running at ~4.5m/s (interesting sidenote: a speed of 4.5m/s qualifies you for the Boston Marathon).

In the initial phases of the competition, participants on the top of the leaderboard mostly used TRPO, which generated promising policies, but in many cases, the policies learnt were stuck in local minimas, and the model finally only ended up hopping, or had unusual gaits (unusual when compared to the way humans walk). While TRPO does guarantee monotonic increment, the presence of a large number of local minimas hindered the learning of more efficient policies after a particular threshold of cumulative reward.

Another major issue in the challenge was that the simulations were approximately 1600 times slower than other commonly used Humanoid-based reinforcement learning environments based on MuJoCo physics engine. The root of this problem lied in the fact that OpenSim was designed for accuracy and high precision, and some participants came up with interesting hacks to trade a little bit of the precision for faster simulation times; and also showed that the models do learn efficient policies even when trained against this alternate build of OpenSim.

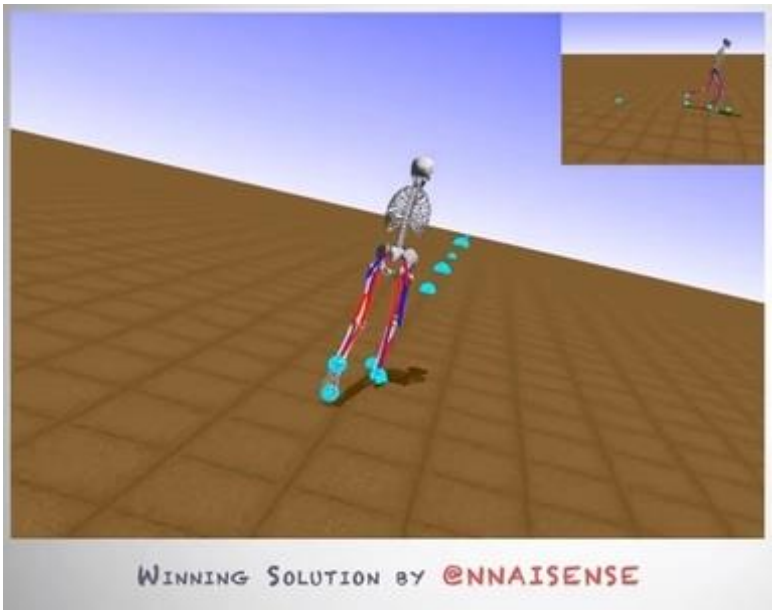
In **future version of the challenge**, we might consider having an alternate build of OpenSim which is more optimized for speed of the simulations and not on the precision of the contact force calculations. In any case, this constraint did incentivize the participants to rely on more sample efficient approaches like DDPG, which ended up dominating the leaderboard

towards the end of the challenge.

The participants in this challenge were very diverse, with expertise in reinforcement learning, biomechanical systems, etc.; and only a small percent of the participants were actually well versed in both the major subdomains this problem belonged to. A setting like this, usually ends up with participants assuming one component or other to be a black box.

In the case of this challenge, it was great to see participants with expertise in reinforcement learning developing a decent familiarity with the internal workings of OpenSim, and also around the basic workings of the musculoskeletal models, to come up with relevant adaptations to their approaches. For instance, a common local minima was one where the model learnt to walk with just one leg while only dragging the other leg; to account for the same, some participants exploited the inherent symmetry in a human musculoskeletal model to do action replays of every episode by swapping the actions for the muscles associated with the left and the right legs.

There were many other such examples where participants used **enrichment of the observation vector**, reward shaping, etc. exploiting the domain specific information they already had about the environment. This helped collectively reiterate the fact that **blind application of blackbox approaches from domain A to a problem of domain B does not go a long way**; and a decent understanding of problems and the constraints of the approaches are necessary to make good progress.



"Blind application of blackbox approaches from domain A to a problem of domain B does not go a long way..."

In conclusion, running this challenge was a great experience, from trying to explore the feasibility of the problem, to helping shape the final problem with considerable input and support from the community of participants. By the end of the challenge, we collectively demonstrated that many reinforcement learning algorithms have reached a stage where we can start exploring their use on a range of high dimensional real world problems in more expensive real world simulation settings; and in the process, also learn to look at the classical problems in different domains in a new light.

"Towards the end of the challenge, DDPG ended up dominating the leaderboard..."

S.P. Mohanty


“I like to achieve new results or share my understanding of things with others…”

Noelia Vállez

Noelia Vállez is an assistant professor at the University of Castilla-La Mancha, Spain and a PhD with experience in medical image processing and machine learning.

Noelia, can you tell us about your current projects?

I am currently involved in a few European projects: in Eyes of Things we are developing our reference platform for computer vision applications with the idea of obtaining a very small, powerful device at a low cost. It gives the users the ability to develop computer vision applications with free software. In the Bonseyes project we are creating a marketplace for deep learning, which is not really complex, but it's a tedious task. We try to make it more visible to the users.

What do you like about these projects?

These are very challenging for me. I like to achieve new results or share my understanding of things with others. That's why I think I don't just enjoy programming. I wanted something more challenging.

How did you end up in your position?

I was studying for my degree in computer science. Then I started working part time with the VISILAB Group. There I did my final degree project before I did my Master's degree project. Two years ago I also finished my PhD with this group.

Would you prefer to work in the industry or stay in the academia?

Well, I don't know because I've never worked in the industry, but I have friends who do. I think I prefer academia.

Do you see yourself staying in academia all of your life?

I hope so.

Do you want to teach?

Yes, I like teaching. I also like researching. I would like to be a professor.

What is the thing about teaching that most attracts you?

I think sharing my knowledge and experience with the students. It depends on the students, of course. It's not the same teaching a first level course and advanced courses.

Why did you choose science?

I used to hate history.

That's a very good reason!

[*we both laugh*]

I like studying language and literature, but I don't like history. I would prefer something more technical, something that I can control.

Is there anything in technology that you don't enjoy?

No... I like it. With technology, I like it all.

Does it run in the family?

My husband also studied computer science.

When did you discover that you liked these kinds of subjects?

I was thinking of working in a bank or something like that because I wanted to work with computers. At that time, I only saw computers in banks or these type of places.

How old were you at that time?

Maybe four or five.

When did you come across computers again?

At school.

Now you can find computers everywhere.

Yes, with your smartphone, you're connected everywhere.

Your husband works in technology as well as many of your friends. Was it a deliberate choice? [*we both laugh*]

I don't know why. Most of my friends are in technology, yes. It just ended up that way.

Can you name a teacher that impressed you?

I would say Maura González, my chemistry teacher during my last years in secondary school in Spain. She was very encouraging. She not only taught the subject, but also explained how to deal with difficult situations and personal conflicts. She was a good example for us. When I will be a teacher, I would like to be like her.

Did you see any of your friends give up a career in science because it was too difficult or they didn't have a good role model?

I did see friends quit their studies because they couldn't get the grades. Others persisted though, and then they finished.

More male or female students?

More male because we are a small number of female students in computer science.

What is the most difficult part about being a woman in a male-dominated profession?

I think it's more difficult working in companies instead of academia. Here, we all work together, and it doesn't matter if you are a boy or a girl. In some companies, people tend to be less open minded.

What is the advantage of being a woman in a predominantly male profession?

Well, I don't know if this is an advantage.

So for you, there are no advantages or disadvantages?

No, not really

Of all of the technologies that you have seen or learned about during your studies, which impressed you the most?

Smartphones because you have the power of a computer in your pocket. That impressed me.

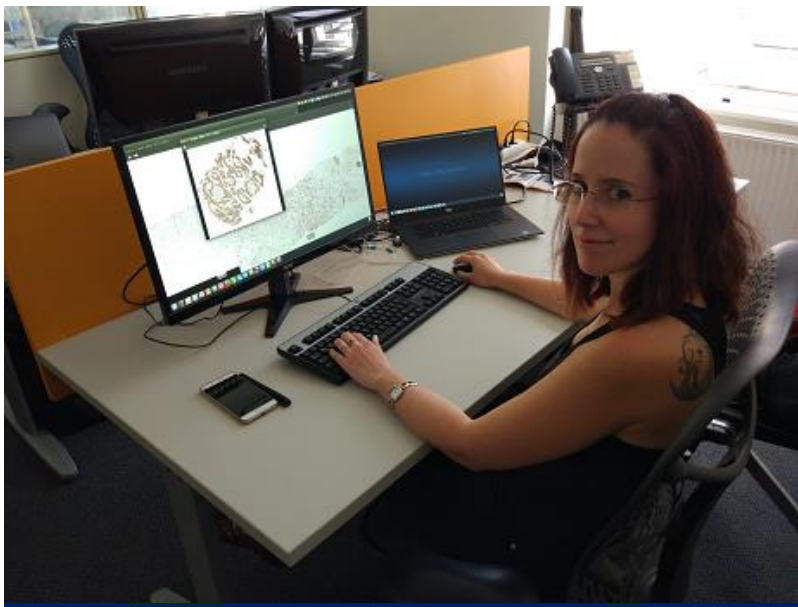
Michael Black gave the same answer! What would you like to achieve?

I would like to be a professor. I would like to be an example to my students, and also continue in the field discovering new technology or new knowledge.

Would you like to discover something for a particular field?

As I'm working now in computer vision and artificial intelligence, I would like to contribute something great to these fields.

Are you afraid that someone might use the technology in the wrong way?



During a secondment in Vienna, Austria

Noelia teaching during the First Computer Vision Week event in Ciudad Real, Spain

We shall see. [*we both laugh*] For example, internet was invented for the army. It can all be used in the wrong way.

So you say we'll see when we get there.

Yes, of course

In your opinion, what will be the next big invention?

I think quantum computation, computers based on quantum physics. I think this will be revolutionary.

Why do you think so?

Because it will change all of the schemes we are following.

When will that happen?

Don't know… Let's see in a few years. There are people working on that.

Is there anything that scares you about the future?

The economic crisis… we'll have to see what happens in the future in Spain.

Do you have kids?

No.

When you have kids, do you hope that they become scientists?

I hope so [*laughs*] I will try! If they want... if they don't, I will advise them to be a therapist or go into the field they want.

So if they like history, you will not send them away?

No [*laughs*]

It seems that you go with the flow very easily. What challenges you in life?

Human relationships... I'm not very good at that.

Why not?

I prefer my computer. [*we both laugh*]

I'm sure your computer also prefers you, but we also like you very much. [*we laugh again*] **What is the most difficult part about human relationships?**

Trying to say something in a way that doesn't make other people worry.

Is it easier to talk to a computer?

Yes, the computer will only do what you said.

But it's less funny! You know what the computer will do. You can't predict what people will do. That's the spice in everything.

When we add more intelligence to the computer, let's see what happens!

Noelia at Ueno Park in Tokyo, Japan

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

Global Artificial Intelligence Conference
Santa Clara, CA Jan 17-19 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

Robotic Vision Summer School - Australian Centre for Robotic Vision
Kioloa, Australia Feb. 4-9 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

WACV Winter Conference on Applications of Computer Vision
Lake Tahoe, NV Mar 12-15 Website and Registration

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

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.

Last minute - 2017 in pictures:
The best science images of the year
Find them here

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)

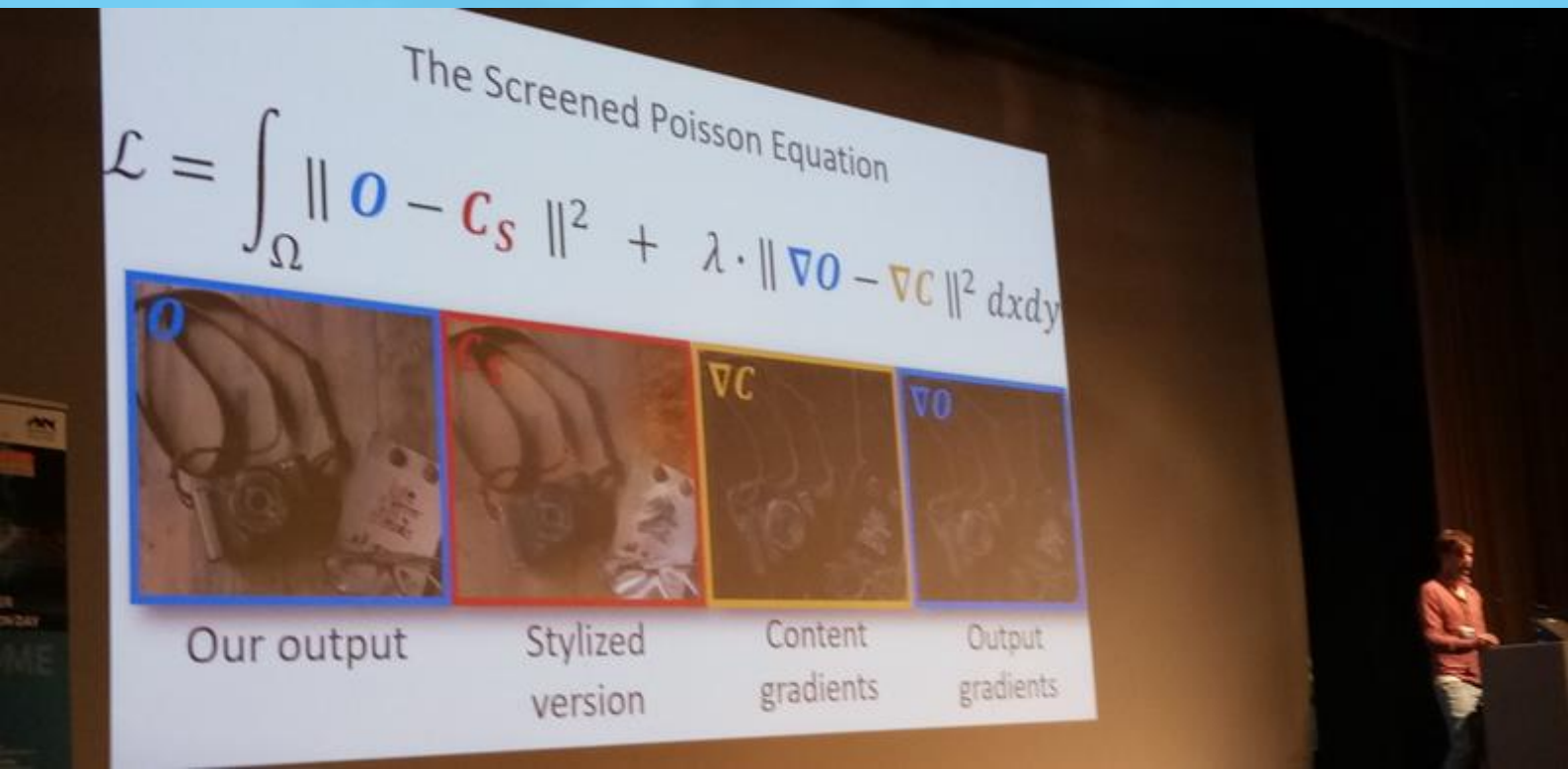
It will take you only 2 minutes to fill and it will help us give the computer vision community the great magazine it deserves!

A few images from the 15th Israel Computer Vision Day, held a few weeks ago. The local computer vision community is expanding and the event hosted more than 600 participants. Kudos to the organizing committee: Ronen Basri (WIS), Michal Irani (WIS), Yael Moses (IDC), Yacov Hel-Or (Amazon & IDC) and Hagit Hel-Or (Haifa).

Below: Hadar Averbuch-Elor presents "*Bringing Portraits to Life*", a method by Tel Aviv University and Facebook which automatically generates photo-realistic videos that express emotions. If that reminds you Face2Face from our issue of May 2016, bingo!

Top: Tali Dekel presents "*On The Effectiveness of Visible Watermarks*", a work by Google demonstrating the vulnerability of watermarks that are typically overlaid on digital images provided by stock photography websites. They propose to make watermarks more effective by adding random geometric perturbations to the watermark when embedding it in each image: this warping makes it more robust and difficult to attack.

Below: Roey Mechrez presents "*Photorealistic Style Transfer with Screened Poisson Equation*", a model by Technion/Adobe that enables to transfer painterly style to images, maintaining the fidelity of the stylized image while at the same time securing a photorealistic output. Other speakers were Elad Osherov, Gautam Pai and more...

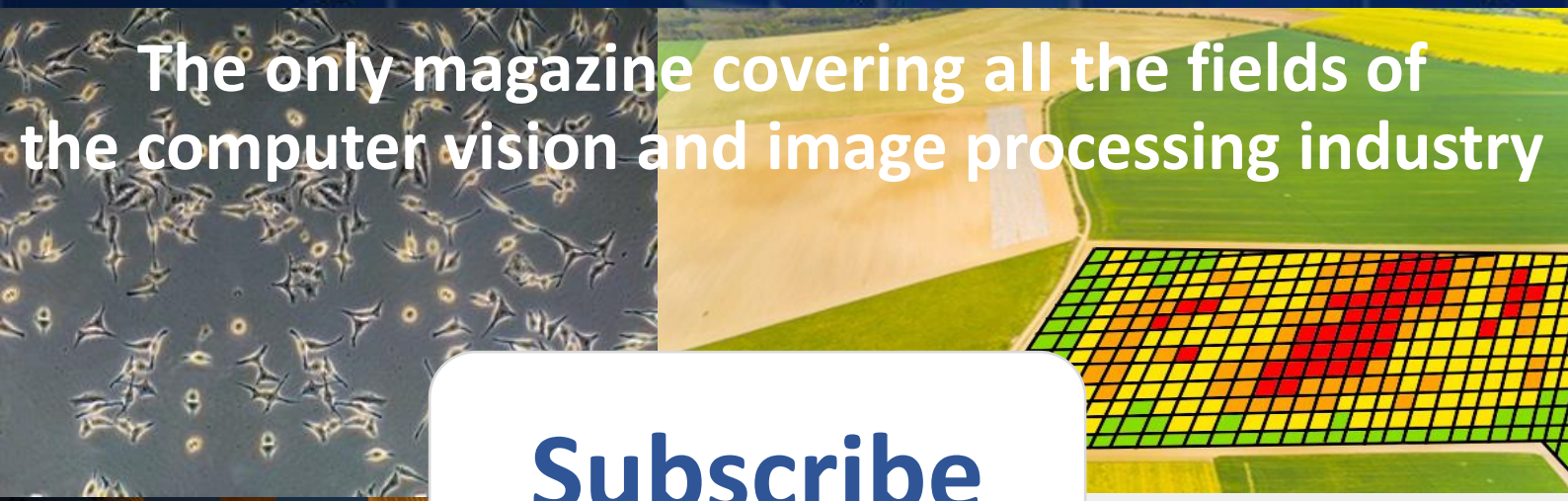


A few images from the Winter School on Surgical Imaging and Vision, held a few weeks ago at the Hamlyn Centre for Robotic Surgery, Imperial College London.
The focus of the Hamlyn Winter School is on both technical and clinical aspects of Surgical Imaging and Vision. Participants enjoyed lectures, hands-on demonstrations, workshops, and mini-projects.
Below left: the visit to the Lab, including an introduction to the Da Vinci robot.

Improve your vision with

Computer Vision News

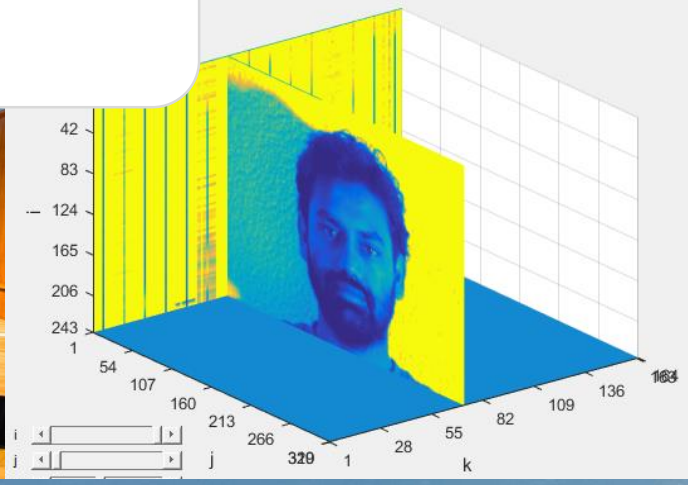
The Magazine Of The Algorithm Community



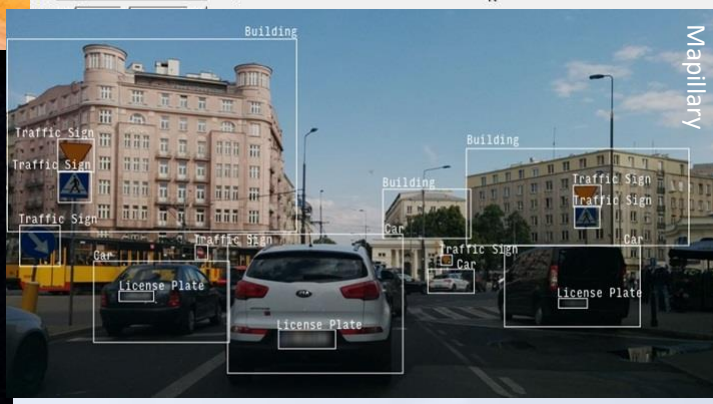
The only magazine covering all the fields of the computer vision and image processing industry

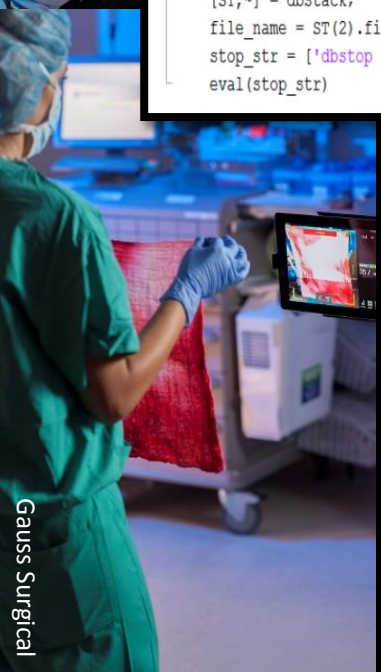
Subscribe

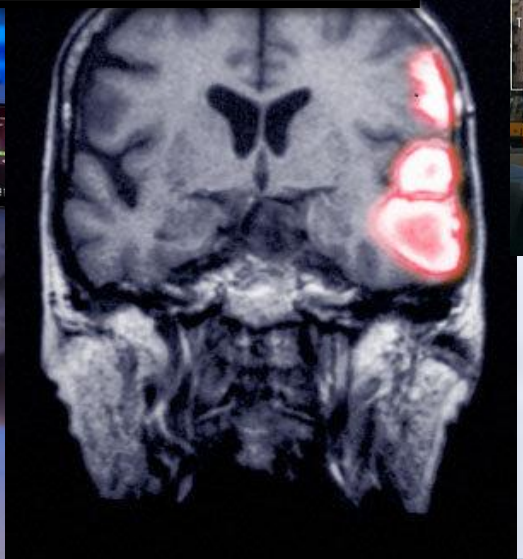
(click here, it's free)




```
% 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

RSIP VISION | Global Leader in Computer Vision

JOIN THE AI AGE IN OPHTHALMIC IMAGING

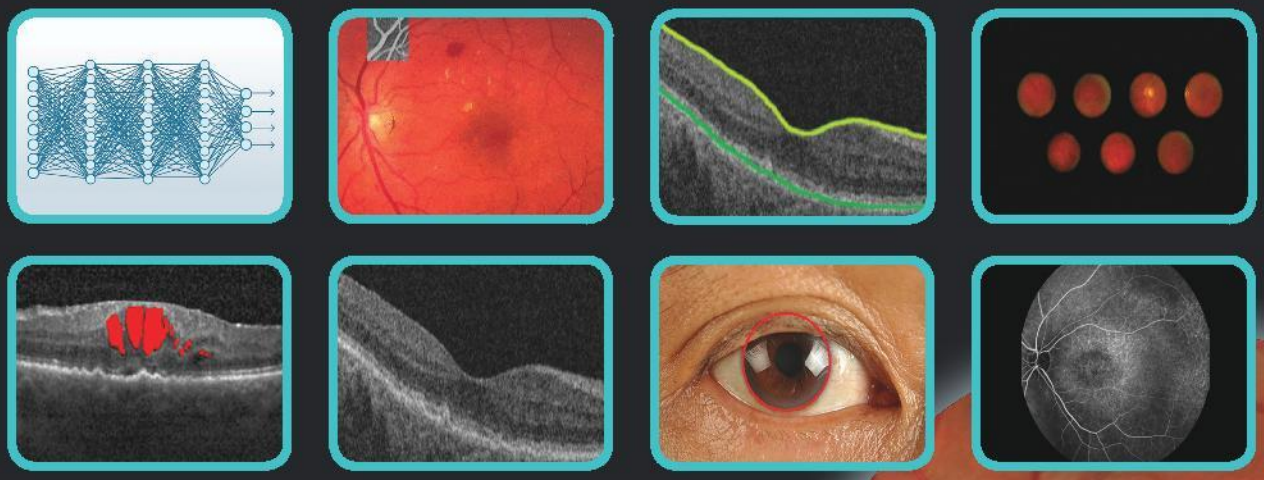
Be part of the AI revolution in ophthalmic imaging. RSIP Vision, a global leader in computer vision and image processing, has completed numerous ophthalmology projects involving development of advanced algorithmic software. In order to achieve state-of-the-art results, the company has invested in breakthrough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