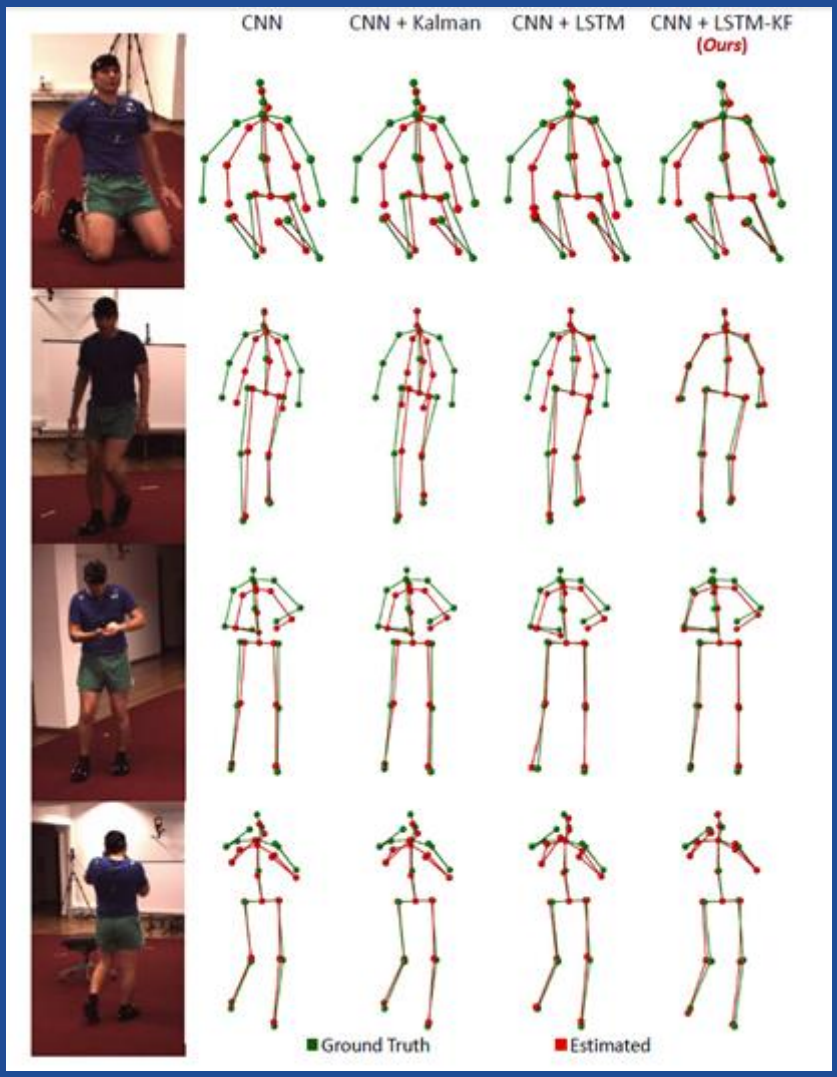
Computer Vision News

The magazine of the algorithm community

A publication by

RSIP VISION

April 2019



Upcoming Events

Women in Science:
Parvin Mousavi

Event:
ADAS SENSORS 2019

Project Management:
When data keeps coming...

Application:
Lifelong Learning Machines

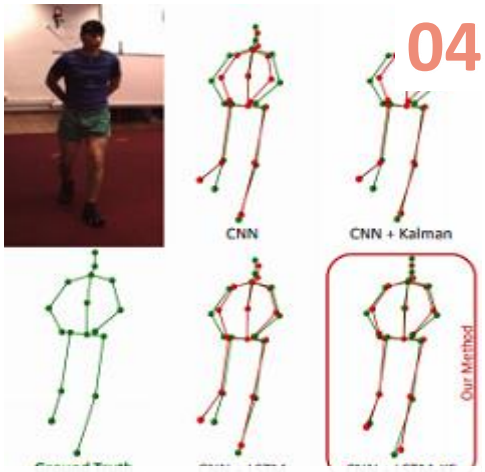
Artificial Intelligence Spotlight News

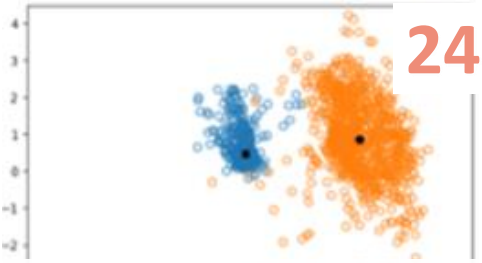
Computer Vision Project:
Automated RECIST Measurement

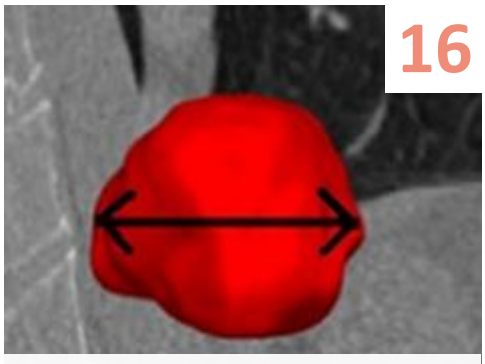
Research Review:
LSTM Kalman Filters: Recurrent Neural Estimators for Pose Regularization

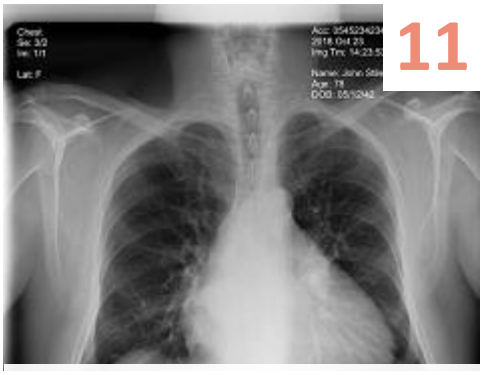
Challenge and Workshop at CVPR2019:
SkelNetOn: Deep Learning for Geometric Shape Understanding

Contents









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

Computer Vision News

Editor:
Ralph Anzarouth

Engineering Editor:
Assaf Spanier

Publisher:
RSIP Vision

Contact us

Give us feedback

Free subscription

Read previous magazines

Follow us:

Dear reader,

First of all, an important invitation goes out to our readers in the **Bay Area**: **The Bay Vision group** will host another **Computer Vision, Deep Learning and Artificial Intelligence Meetup**, this time centered on **Medical Imaging**. I will personally attend and hope to see you all in Cupertino on April 4. **The event will feature superb speakers and registration has opened; plus the pizza is on us!**

The field of Medical Imaging is not covered only by this Meetup. It was also the subject of our **Webinar** that we broadcasted a few days ago. Did you miss it? The **RSIP Vision YouTube channel** features the full recording (with all the slides).

Meetups and Webinars represent two more programs offered by **RSIP Vision** to achieve our main goal: to inform all members of our community about the latest advancements in AI technology and solutions. With **CVPR 2019** (and our **CVPR daily**) just two months away, combined with this **Computer Vision News** monthly magazine, we strive to do our best to reach this target. **Please let us know what you think**.

Enjoy the reading!

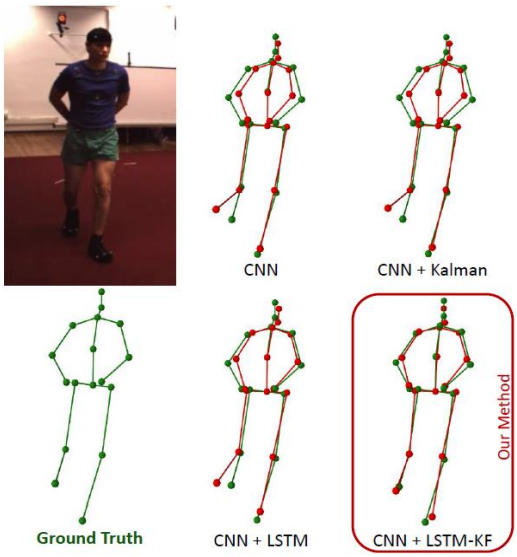
Ralph Anzarouth
Editor, **Computer Vision News**
RSIP Vision

Three fathers of AI were named recipients of the Turing award. Read our exclusive interviews with Yoshua Bengio and Yann LeCun!

by Assaf Spanier

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen **Long Short-Term Memory Kalman Filters: Recurrent Neural Estimators for Pose Regularization**. We are indebted to the authors (**Huseyin Coskun**, **Felix Achilles**, **Robert DiPietro**, Nassir Navab, **Federico Tombari**), for allowing us to use their images. The paper was first published at **ICCV2017** and it is found here.

Camera localization, object tracking and pose estimation (for example, localization of joints such as shoulders, knees, hips, wrists, etc. as in illustration) are very changing tasks in computer vision, mainly due to their high dynamic noise.



The authors' innovation is the LSTM Kalman filter (LSTM-KF), a new architecture capable of learning a motion model and all the parameters of a Kalman filter

To overcome these tasks/challenges, regularization is often handled using temporal filters: the Kalman filter is the simplest and most general, enjoying widespread use. However, Kalman filters require pre-determining a motion model and measuring model, which raise modeling complexity and are often only a rough approximation of real world behavior. For pose estimation or object tracking, for instance, the motion models used most often assume constant velocity or constant acceleration -- these unrealistic, simplified representations make it very difficult to arrive at a precise solution.

The authors propose constructing a network to learn dynamic representations of a motion model and of noise. Specifically, they propose using an LSTM network to learn these models from the data. The proposed method has the advantage of arriving at a representation model which takes account of all previous observations and previous states. In the paper, the authors evaluated their method using the three most popular datasets in this field and achieved state of the art results on each dataset.

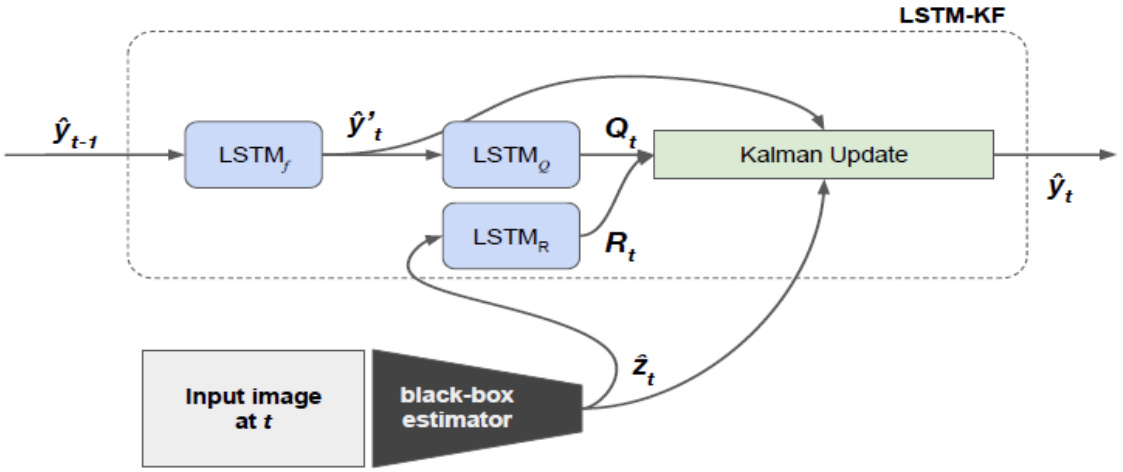
Introduction

Pose estimation is a challenging task in the field of computer vision. It is needed for tasks such as camera localization, body pose estimation (human and animal), object tracking and more. There are a number of approaches trying to solve these problems using one-shot pose estimation -- that is, each image is classified independently of other images, meaning this type of modeling completely ignores the data available due to the sequential nature of video images. This disregard of temporal data can result in very noisy estimates and a confusion of image features which may be visually similar, but are in fact spatially significantly dissimilar -- such as confusion between the right leg and the left -- in the case of body part localization. Temporal filters are a widely used approach to improving classification in cases such as these. Kalman filters are the most popular, due to their simplicity and applicability to the widest variety of cases. Moreover, the extended Kalman filter can handle nonlinear system motion models in both its measuring models and in transition between states -- that is, from one point in time to the next.

… to achieve all the advantages of machine learning, using a much smaller amount of data.

In many cases and tasks, however, such measuring models cannot be predefined, and in these cases the application of Kalman filters is severely limited. In common computer vision tasks, objects and body parts don't conform to simple motion models. In such scenarios, Kalman filters, which use a constant velocity or constant acceleration model of motion can only arrive at a rough estimate of real-world motion. To overcome these limitations, researchers attempt to have networks learn motion models directly from the training data, using methods such as SVM or LSTM. These machine learning methods can indeed help the model and at the same time enrich the basic motion model. However, using learned motion models requires the network to arrive at models based on the consistent movement constraints observed in the training data over time, which means massive training data is needed to cover all possible movement scenarios of an object type.

The authors' innovation is the **LSTM Kalman filter (LSTM-KF)**, a new architecture (shown in the illustration below) capable of learning a motion model and all the parameters of a Kalman filter. Thus, making it possible to **achieve all the advantages of machine learning, using a much smaller amount of data.**



Let's take a closer look at the LSTM-KF -- long short-term memory Kalman filter, which the authors also define as a model for the temporal regularization of pose estimators.

At its core, its idea is to leverage Kalman filters without pre-defining a linear state function A or any constant process, and various measuring matrices Q and R. Instead, the network learns and models a non-linear transition function F, along with matrices Q and R, using 3 different LSTMs - in this way the authors make the method capable of learning a rich and dynamic representation of Kalman filter parameters directly from the data.

Architecture

Kalman Filter:

A quick recap of Kalman Filter equation - is an optimal state estimator if the linearity and Gaussian noise assumptions hold. Where y_t represents the state and z_t our measurement:

$$y_t = Ay_{t-1} + w, \qquad w \sim N(0, Q)$$
$$z_t = Hy_t + v, \qquad v \sim N(0, R)$$

A, *Q*, *H* and *R* are known matrices

The Kalman filter is trained and achieves optimality using an iterative loop with two update steps: the prediction step and the update step.

The prediction step estimates the mean and covariance independently of the measurement at time *t*:

$$\widehat{y'}_t = A\widehat{y}_{t-1}$$
$$\widehat{P'}_t = A\widehat{P}_{t-1}A^T + Q$$

The update step computes the optimal Kalman gain K_t , along with the observed measurement z_t , to estimate the mean and covariance of y_t :

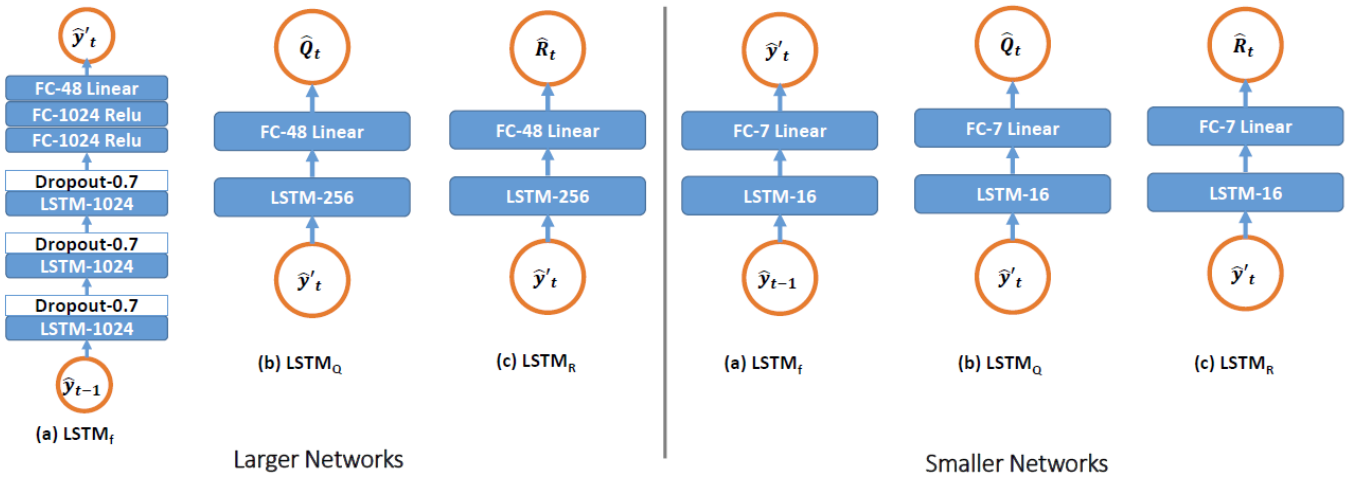
$$K_t = \widehat{P'}_t H^T (H\widehat{P'}_t H^T + R)^{-1}$$

$$\widehat{y_t} = \widehat{y'}_t + K_t(\hat{z}_t - H\widehat{y'}_t)$$

$$\widehat{P_t} = (I - K_t H_t)\widehat{P}_t$$

The model uses LSTM components: you can review details about LSTMs in this previous article of ours.

Three LSTM components are used: $LSTM_f$, $LSTM_Q$ and $LSTM_R$ respectively modeling f , Q_t and $\widehat{R_t}$. Each component is described in the illustration below. Note there are two different configurations for each LSTM component -- for larger networks (on the left) and smaller networks (on the right). At any point in time t, $LSTM_f$ takes the former prediction $\widehat{y}_{t-1}$ as input and produces Q_t (the covariance matrix). At the same time, z_t is taken as input by $LSTM_R$ to produce the matrix R_t as output. Then, $\widehat{y'}_t$ and z_t , as well as matrices Qt and $\widehat{Rt}$ enter as input into a standard Kalman filter, undergoing the process described in equations of the Kalman Filter above. And finally, the network produces a new estimation $\widehat{y_t}$. Note that *Q* and *R* are diagonal positive definite matrices.



Loss:

Initially, the authors used standard Euclidean loss for all training stages, but they found that the $LSTM_f$ module failed to learn a reasonable mapping. They therefore added an additional loss term to improve the gradient flow to the $LSTM_f$ module for improving the training process, resulting in the following overall loss function:

Research

$$L(\theta) = \frac{1}{T}\sum_{t=1}^{T} \left|\left|y_t - \hat{y}_t(\theta)\right|\right|^2 + \lambda \left|\left|y_t - \widehat{y_t}'(\theta)\right|\right|^2$$

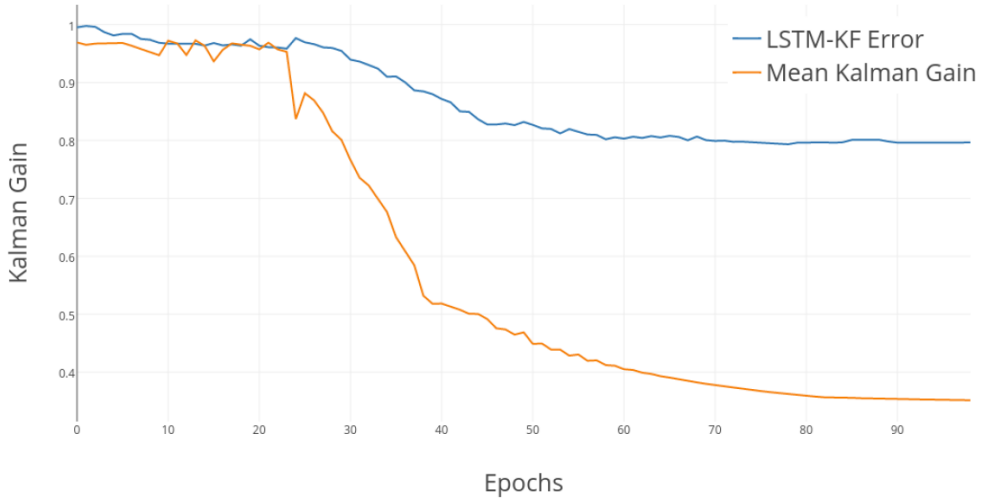
Where the value of the parameter λ is 0.8. The goal, of course, is to find the optimal parameter values, such that they minimize the loss function given by the equation $L(\theta)$. In this model, they are the concatenation of all the weighting matrices from all 3 LSTM modules. Note that these modules combine LSTM layers with linear layers, as described in the illustration below. The model can be trained end-to-end; the method was implemented using TensorFlow and Adam optimization.

Results

The authors evaluated the LSTM-KF architecture's performance by comparison to a variety of temporal regularization methods: 2 standard Kalman filter implementations, assuming constant velocity or constant acceleration (denoted as Kalman Vel and Kalman Acc, respectively), exponential moving average filter (EMA), and a standard LSTM module (Std. LSTM).

The evaluation was run on 4 different datasets, one for evaluating 3D human pose, two for evaluating camera pose, and one for object pose estimation, all using RGB images as input.

The illustration below presents Kalman gain values as a function of training time (epochs) for the LSTM-KF error and mean Kalman gain. Note that the gain (as well as the error) is high at first, which is an indication that at this stage the method relies almost entirely on measurements. As training progresses, Kalman gain drops significantly, indicating that the Kalman filter is significantly relying on both measurements and the $LSTM_f$ module.



Human Pose Estimation Results:

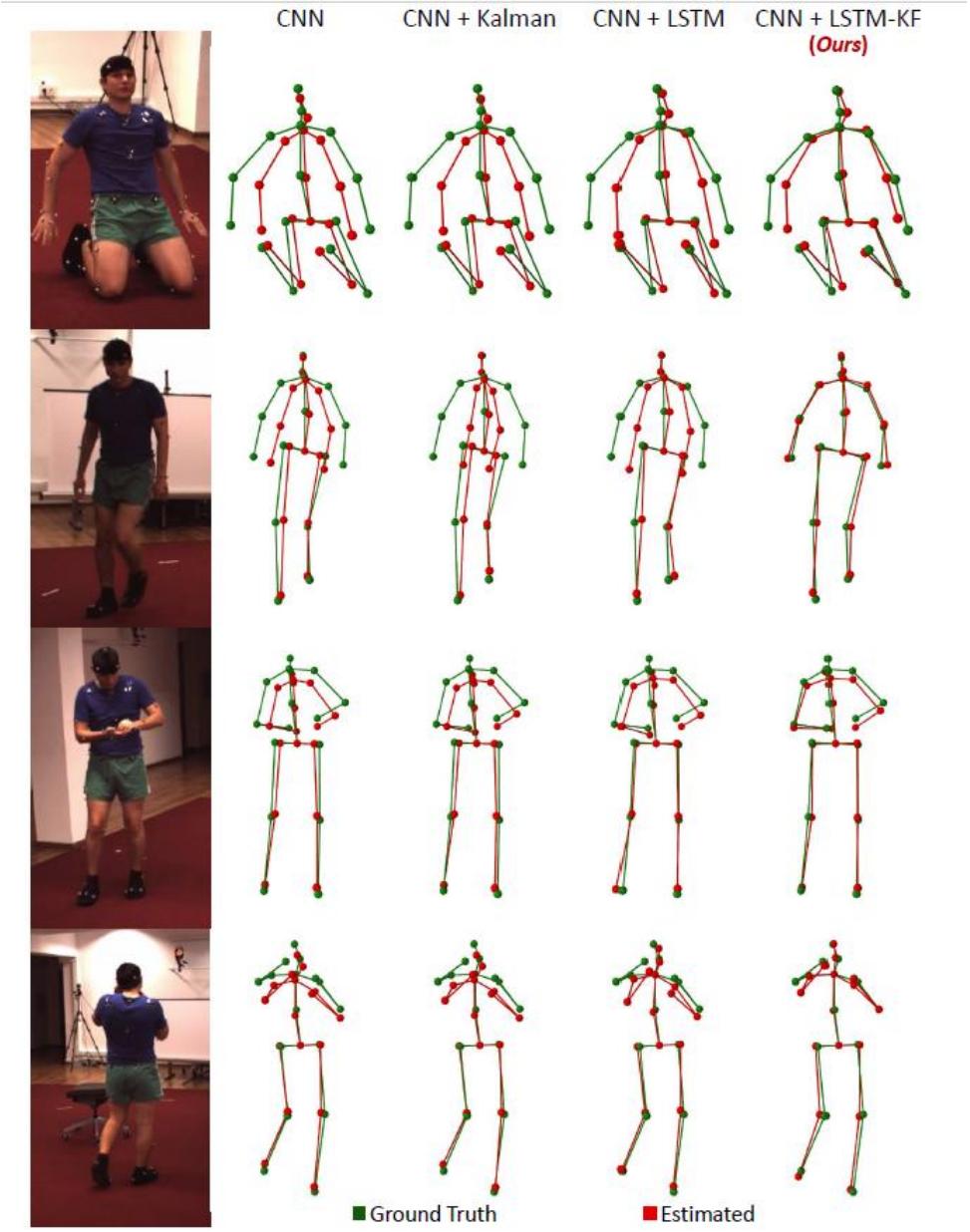
The Human3.6M database consists of 3.6 million RGB images from video sequences. Each video sequence includes 7 actors, each of them performing 15 activities of varying movement complexity. Each activity takes between 3,000

and 5,000 frames. For this paper, the authors use for training the sets of movements marked as S1, S5, S6, S7 and S8 and for testing the sets S9 and S11. For more details about the those sets, see the Human3.6M website.

The table below provides detailed quantitative results:

	Street		K. College		S. Facade		St. M. Church		Old Hospital		**Mean**	
	tran.	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*
PoseNet [16]	3.35	6.12°	1.97	5.38°	1.65	8.49°	2.88	9.04°	2.60	5.32°	2.49	6.87
+ Kalman Vel.	3.16	5.93°	**1.85**	5.29°	**1.48**	8.20°	2.94	9.29°	2.53	5.07°	2.39	6.75°
+ Kalman Acc.	3.14	5.92°	1.88	5.29°	1.49	8.33°	2.95	9.33°	2.45	5.07°	2.38	6.79°
+ EMA	3.33	5.63 °	1.95	**5.28°**	1.62	8.35°	2.82	8.99°	2.68	5.10°	2.48	6.67°
+ Std. LSTM	9.56	11.2°	4.24	7.95°	1.87	7.04°	3.34	11.52°	4.03	6.46°	4.61	8.83°
+ LSTM-KF (ours)	**3.05**	**5.62°**	2.01	5.35°	1.63	**6.89°**	**2.61**	**8.94°**	**2.35**	**5.05°**	**2.33**	**6.37°**

Below follow the qualitative results for the evaluation on the Human3.6M database. The ground truth is in red and the model's estimate is in green. Based on the initial CNN estimate, we compare outputs of standard Kalman, standard LSTM and the authors' LSTM-KF model. LSTM-KF shows significant improvement over the other methods, especially for the arm joints and leg joints.



Research

Results of temporal regularization on object tracking estimates - errors in translation as [mm] and rotation in [degrees]:

	Kinect Box		Tide		Orange Juice		Milk		Mean	
	tran.	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*
Tan *et al.* [24]	1.70	0.30°	1.17	0.44°	1.29	0.35°	1.27	0.41°	1.36	0.37°
+ Kalman Vel. al.	1.69	0.29°	1.84	**0.38°**	1.27	0.35°	1.27	**0.35°**	1.52	0.34°
+ Kalman Acc.	1.69	**0.28°**	1.84	**0.38°**	1.28	0.31°	1.79	0.42°	1.65	0.35°
+ EMA	1.71	**0.28°**	1.17	0.39°	1.50	**0.28°**	1.49	0.37°	1.47	**0.33°**
+ Std. LSTM	41.03	6.30°	32.23	8.31°	30.16	7.42°	18.3	7.95°	30.43	7.49°
+ LSTM-KF (ours)	**0.86**	0.35°	**0.77**	0.49°	**0.59**	0.37°	**0.66**	0.43°	**0.72**	0.41°

Results on camera pose estimates provided by PoseNet on the Cambridge Landmarks dataset. Values are in [m] of translation and in [degrees] for rotation:

	Street		K. College		S. Facade		St. M. Church		Old Hospital		Mean	
	tran.	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*	*tran.*	*rot.*
PoseNet [16]	3.35	6.12°	1.97	5.38°	1.65	8.49°	2.88	9.04°	2.60	5.32°	2.49	6.87
+ Kalman Vel.	3.16	5.93°	**1.85**	5.29°	**1.48**	8.20°	2.94	9.29°	2.53	5.07°	2.39	6.75°
+ Kalman Acc.	3.14	5.92°	1.88	5.29°	1.49	8.33°	2.95	9.33°	2.45	5.07°	2.38	6.79°
+ EMA	3.33	5.63 °	1.95	**5.28°**	1.62	8.35°	2.82	8.99°	2.68	5.10°	2.48	6.67°
+ Std. LSTM	9.56	11.2°	4.24	7.95°	1.87	7.04°	3.34	11.52°	4.03	6.46°	4.61	8.83°
+ LSTM-KF (ours)	**3.05**	**5.62°**	2.01	5.35°	1.63	**6.89°**	**2.61**	**8.94°**	**2.35**	**5.05°**	**2.33**	**6.37°**

See the full article for more detailed camera tracking and object tracking results.

...a tool that makes motion modeling and estimation problems much easier, and makes it possible to learn rich models from difficult-to-model real-world data.

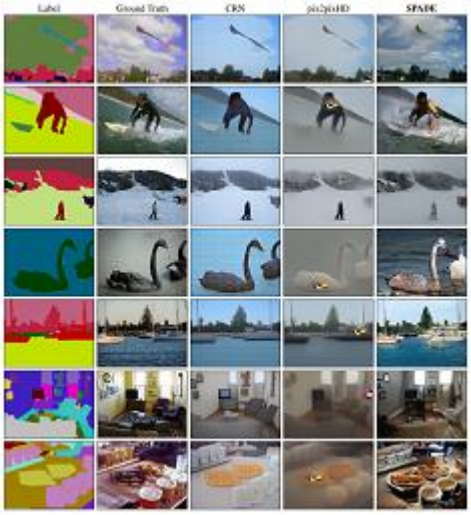
Conclusion

The paper presented long short-term memory Kalman filter (LSTM-KF) model, a tool that makes motion modeling and estimation problems much easier, and makes it possible to learn rich models from difficult-to-model real-world data. Utilizing a wide array of experiments, the authors showed LSTM-KF outperforms the standard Kalman filter and standard LSTM, achieving state of the art performance on 3 widely divergent tasks -- just to illustrate, their model reduced the error in localization of body joints on the Human3.6M database by 13.8%, from 82.3 mm to 71.0 mm.

Computer Vision News has found great new stories, written somewhere else by somebody else. We share them with you, adding a short comment. **Enjoy!**

Amazon Explains How CV Can De-identify Medical Images:

Let's start once again with **Amazon:** How to de-identify **millions of medical images** that researchers need to solve all kind of pathologies? They propose **Amazon Comprehend Medical** and the controversial **Amazon Rekognition** as a two-step automated solution that solves privacy issues and regulatory obligations like **HIPAA** (Health Information Portability and Accountability Act). This blog post displays how these two services, plus some Python code, can inexpensively and quickly detect, identify and redact **PHI (Protected Health Information)** from within medical images. **Read More…**



Semantic Image Synthesis with Spatially-Adaptive Normalization:

It's another great paper from **the Berkeley folks**, together with **NVIDIA** and **MIT**. You will not be surprised to learn that BAIR's **Alyosha Efros** and NVIDIA's **Jan Kautz** gave advice on this work. It's a simple but effective layer for synthesizing photorealistic images given an input semantic layout. Their model allows users to easily control the style and content of synthesis results as well as create multi-modal results. We look forward to their oral presentation at **CVPR2019**, in June at Long Beach, CA! **Enjoy!**

Reinforced Cross-Modal Matching & Self-Supervised… :

Another oral presentation that we will follow at CVPR2019 is **Reinforced Cross-Modal Matching and Self-Supervised Imitation Learning for Vision-Language Navigation**. This paper at the intersection of NLP with Computer Vision proposes a new method for vision-language navigation (VLN) tasks that combines the strengths of both reinforcement learning and self-supervised imitation learning. **Read More**

The Improving Picture of Artificial Intelligence in Healthcare:

RSIP Vision's CEO **Ron Soferman** writes great articles on other magazines too. Here is another one. What he forgot to tell is that RSIP Vision can help you acquire almost **all technologies and applications listed in the article!** **Read!** **Contact Ron**

More Great Articles to Read:

Will Artificial Intelligence Steal Your Fingerprints? **Help Chemists with Disabilities**
Estimation of Free Space on Car Park Using Computer Vision Algorithms

When Data is Collected During 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 **Yael Zak** tells us about another aspect of **Project Management with Deep Learning: When Data is Collected During the Project**. It's another handy tip by **RSIP Vision** for Project Management in Computer Vision.

"Data is everything in this game!"

Let's see what happens when data is collected during an on-going project; in other words, when input images or videos keep coming at the same time as you develop your network.

In such a case you might find yourself in a feedback loop, requesting for specific data and changing your network as the data grows. Two main things that would help you with this would be flexibility in network architecture, and a careful follow up of the training procedures.

We'll start with code flexibility: Every deep learning development project implies experimenting with varied deep learning methods and tricks. Therefore, you'd like to be able to easily change network architecture, without the need to modify your entire previously written procedure.

When data is growing, it might be that the project is still changing its shape. Some of its challenges are yet undiscovered and perhaps the network's objective will change a few times before it will mature. In addition to changes in the internal architecture of the network, this could entail changes of network interface: the loss function, the inputs and outputs, and correspondingly the format of the ground truth. You want to be able to quickly go back and forth between network architectures in order to lunch trainings and compare their results. Thus, in order to make life easier, consider a code structure that allows flexible network interface, a generic prediction output, and a modular network architecture.

The process of better understanding the network's problems and objective could be efficiently driven by iteratively training the network on the existing data, and then inspecting its false predictions. By analyzing the resulting errors, you could identify specific cases in which the network is not performing well. What common features and characteristics do these cases share? This could teach you what the network has actually learnt, and what are the challenges it is still facing.

At early steps of the project you might discover that the network hasn't learnt anything sophisticated, but it is predicting based on degenerated features that happen to fit your ground truth. In this case, this means that your data is not varied enough with regard to the degenerated feature, and it is not reliable enough in its representation of

In deep learning, once your data is well organized, you have done a big slice of the work!

reality. When the project is ongoing, you can sometimes guide data collection about the data that is required. This is a huge advantage, since the major condition for the network's success is a varied and representative dataset. If you are lucky enough to direct the data collection, you could point out the faults of the data. When you retrain your network with the improved data, you can now analyze the resulting errors, and better understand the network's faults. You could ask for the data collection's help in supplying more examples of those peculiar cases in which the network fails.

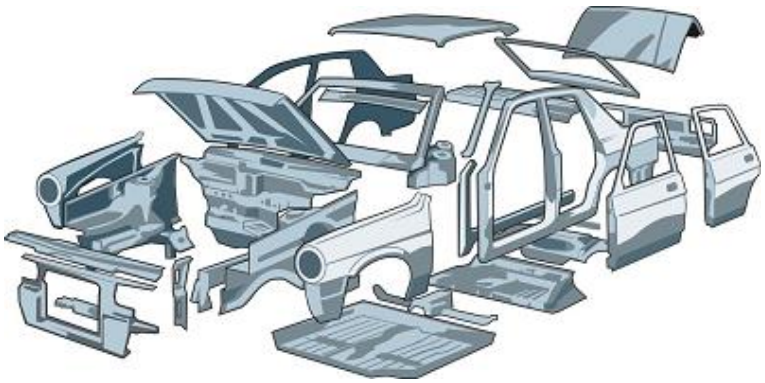
When data keeps coming...

Here, it's important to go into details and not content ourselves with the numeric results of the network's grade. A project manager could easily be excited about a high scoring network. He believes it to be very well trained - while in reality, validation set was limited and degenerated, and perhaps it did not evolve together with the ever growing and more complex training set. It could also happen that the validation dataset is varied, but the measurement you use to grade your network is not well representing its achievement, now that the dataset has changed. Remember to verify, by sampling and inspecting actual examples, that your validation dataset and its quantitative evaluation are still informative and reliable.

Assessing improvements with an ever growing data require a system of code that will allow a very clear follow up of the launched trainings. Comparing network achievements makes sense only if both networks were trained on an equally complicated dataset. We'd like to avoid the unfortunate moment of staring a list of weights files, without having any clue about what do their quantitative scores mean and what is the conclusion of all this. For this sake, I recommend to keep for each weights file a clear follow up of the training process it went through, linking it or saving it in a directory with all the relevant information: its network architecture, network hyperparameters, a list of data items it was trained on, a list of data items it was evaluated on and its score.

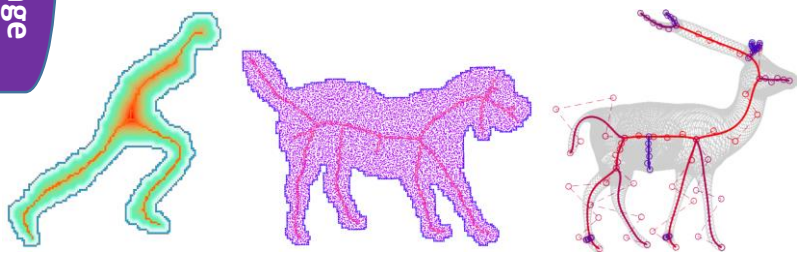
Finally, when your dataset is quite mature, and you inspect the errors and ask yourself: what is the story behind these errors? You may find out that the network has some difficulty in understanding the core question you are asking it, despite the varied data. At this point, you might consider changing the question you are asking, which might entail modification of the network's interface and thus require some bravery. These are exciting moments in a project, since they invite you to see the problem with new eyes. They hold within them the opportunity for a step forward in the project's achievements.

More articles on Project Management



General shape understanding tasks have not been explored to the same extent as segmentation, recognition and detection tasks...

Ilke Demir, Kathryn Leonard, Géraldine Morin and Camilla Hahn are hosting a workshop at CVPR 2019 – which takes place in Long Beach, California in June – about the use of deep learning approaches for geometric shape understanding. Deep learning approaches for segmentation, recognition and detection tasks are already very popular; but general shape understanding tasks, such as how we can extract abstractions of shapes, have not been explored to the same extent.



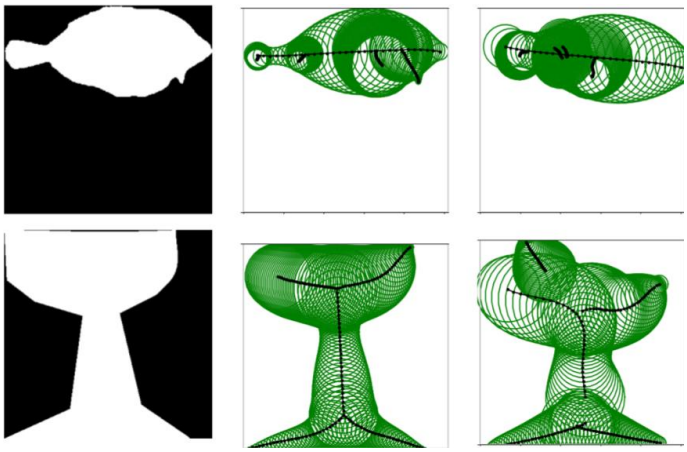
In the lead up to the **workshop**, the team are running a **challenge** looking at one general shape understanding task: **skeleton extraction from a given shape**.

The communities of skeletal modelling and deep learning don't have a huge amount of overlap, so this is something that has not been done before. There are existing methods for extracting skeletons from shapes; however, there are drawbacks to those approaches in the sense that they end up producing branches that capture meaningless information such as pixelation and other kinds of noise on the boundary. The team are hoping to circumvent this by **using deep learning to extract clean skeletons**.

This will be a challenge because extracting a skeleton is not a robust process. If you have a noisy shape, the skeleton is going to be very different. Small changes on the boundary can produce large changes to the skeleton.

These geometric shapes have a particular structure that is embedded in the skeleton.

The team are proposing to explore the problem in three domains: image, point, and parametric representation. They just released a dataset paper with some very simple baselines for participants on how to approach the challenge, and people can go from there and develop their own solutions. If the challenge itself isn't enough to tempt people to join, there are prizes on offer! The winner in each track will receive a **Titan GPU** sponsored by **NVIDIA**.



Where did it all begin?

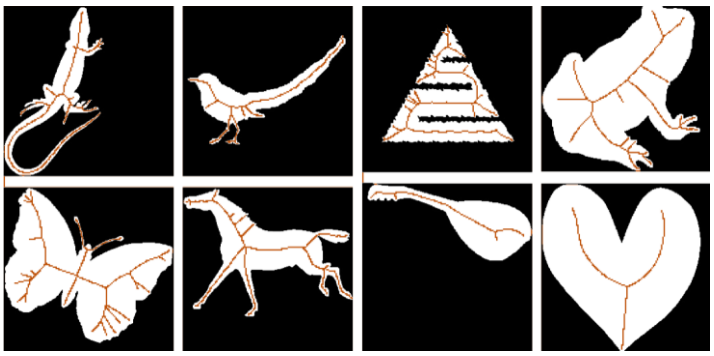
The idea for the challenge was conceived when the team met at a workshop last summer in Germany,

organized through a research network called **Women in Shape Modelling**, sponsored by the **Association for Women in Mathematics**. Every few years it brings together women who are interested in shape modelling to work on open research problems together. At the workshop, there was one group working on medial axis and skeletons, and another working on deep learning, which gave the team the idea to join the two together.

Kathryn, who is a professor at Occidental College, tells us that she has really enjoyed working on the challenge: *"The challenge has been challenging for us as well, but it's been a lot of fun getting a peek behind the scenes of how these things get put together and all of the different moving parts. We have people in France, Germany, and across the United States, so just working with this really diverse team of creative and good-humored people has been great."*

Camilla, M.S. from Bergische Universität Wuppertal, adds: *"I'm in the very early stage of my career, and probably family as well, so it's been a great opportunity to meet so many great women in so many different steps of their careers, and work with them together on this great project."*

Géraldine, a professor from the University of Toulouse, echoed this sentiment: *"It was so much fun to start this working group and this challenge, and I think we've been ambitious. We are very happy to get some new ideas on this topic."*

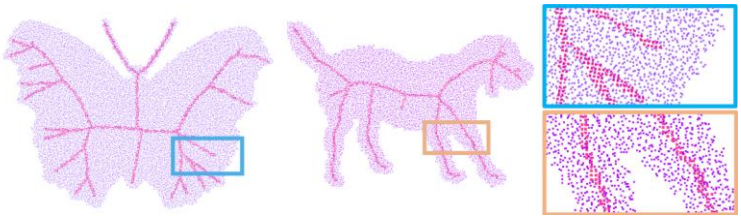


"If this challenge can output a network that is able to learn a lot of things on this 3D data that we get through all of these devices, that would be really great!"

Challenge

What do they hope to achieve?

Ilke, a research scientist at DeepScale, explains that geometric deep learning is an area that is not yet explored to its full potential: *"I think the ultimate outcome is some groundbreaking deep learning approach that has its own representation for shapes, and has its own convolutions and pooling and everything to understand shapes. What we expect from this workshop is* ***to find that breakthrough deep learning architecture that can be used for shape understanding****."*



Géraldine agrees: *"I think it's a real challenge. There are all these devices to capture the world around us, like LiDAR scanners, reconstruction, set of points, and we know that deep learning is doing great for images. If this challenge can output a network that is able to learn a lot of things on this 3D data that we get through all of these devices, that would be really great."*

Pharma - Automated RECIST Score

Computer Vision News

by Miki Haimovich

Every month, **Computer Vision News** reviews a successful project. Our main purpose is to show how diverse image processing techniques contribute to solving technical challenges and real world constraints. This month we review a Medical Image Analysis project by **RSIP Vision** for the **Pharma** industry: **Automated RECIST Score**.

Project

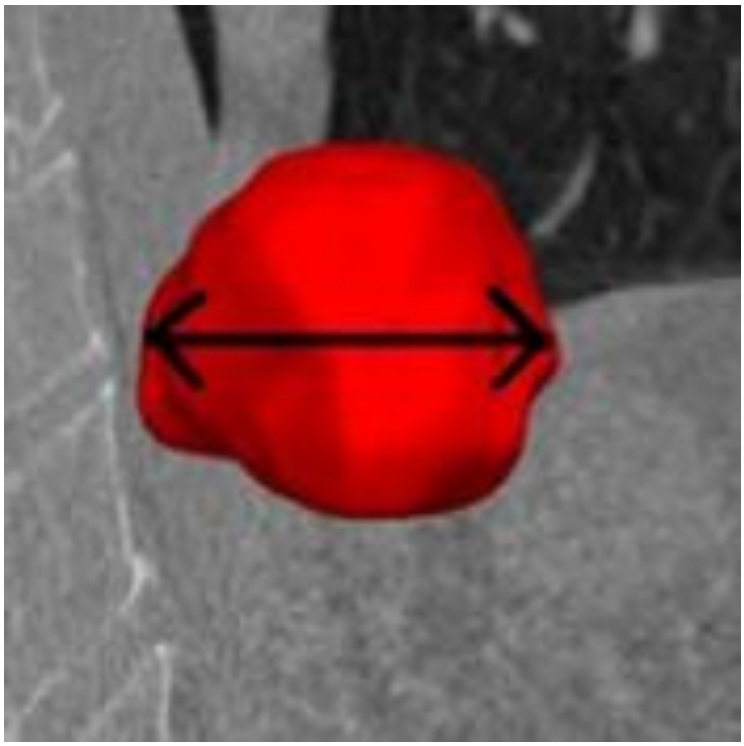
One of the most common practices in clinical trial is: **tracking the effect of a new drug on cancer tumors**. The Pharma industry constantly needs a fast and accurate system for measuring tumors and tracking tumor progression along the treatment.

The golden standard is the **RECIST score (Response Evaluation Criteria in Solid Tumors)**, used by the medical community to assess solid tumors (generally on CT scans) for almost 20 years. The RECIST score is based on the measurement of the diameter of the tumor, calculated as the longest line that can run through the tumor under any angle.

To this day, radiologists measure the RECIST score **manually**, which is problematic for mainly two reasons:

the first reason is **the long time** it takes the expert to view the tumor from all possible angles; the second reason is **the chance of mistake** (like skipping the angle with the longest diameter), together with the large inter- and intra-variability of human observations. Tumors vary in shapes and sizes, and they represent **a severe challenge to the human observer**.

RSIP Vision developed a software and procedure to solve this problem. This solution supports the radiologists in their decision (without replacing them). **The algorithm** assists them by offering a faster and more accurate detection and measurement of the longest diameter in each tumor.



The algorithm assists radiologists by offering them faster and more accurate detection and measurement of the longest diameter in each tumor.

The algorithm operates following several steps:

The first step is the **automatic detection and segmentation of the tumors in the scan**. Then, according to the standard, the algorithm chooses the two largest tumors as targets and it measures the RECIST score for them.

In the second step, **a sophisticated registration process locates the same tumors in the later CTs** that are scanned along the treatment period. It makes it fast and easy to compare different time points and state with certainty whether the tumor has grown or recessed.

Tracking the RECIST scores of those tumors along time gives the researcher the necessary data for a very accurate clinical evaluation of the medicine and its effect on the sampled patients.

In addition, the software can measure **the volume of the tumor** (not only the longest diameter) and its progression over time, providing more meaningful insights regarding the severity of the pathology and its response to different drugs and therapies.

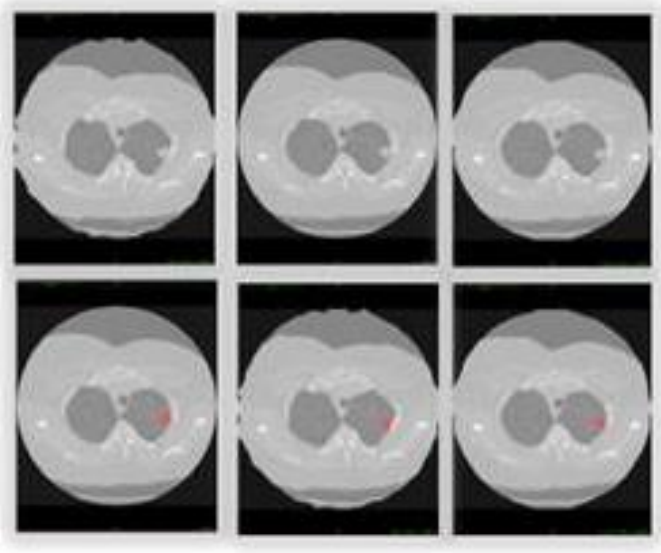
Once the Automated RECIST algorithm is properly trained by RSIP Vision engineers, it can make the radiologist's work easier and less error prone.

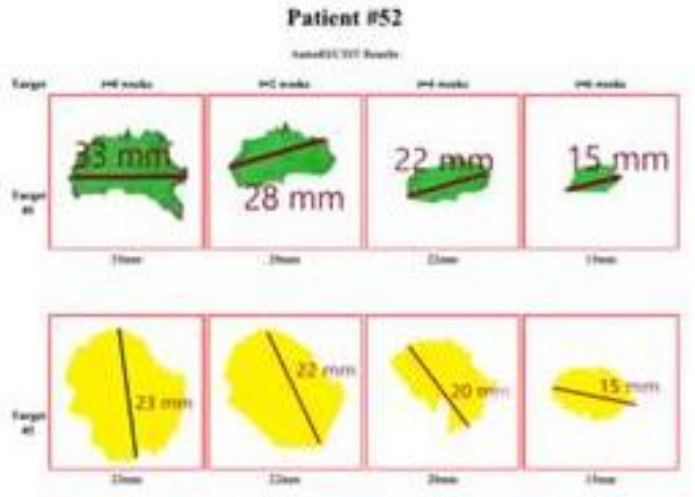
The results are presented to the radiologists in a sequential mode, in order to enable them to give a final approval, after viewing the tumors progression or regression along the treatment.

This fully transparent and fully logged procedure gives the Pharma company whole traceable data along with the final results - all of which is performed in a much more efficient way. RECIST algorithms allow to review very high volume of scans, enabling to make the most of large databases.

[Read about more projects for the Pharma industry]

Tool allows for the automated registration of multiple scans to follow patient across





Take us along for your next Deep Learning project!
Request a call here

Parvin Mousavi

Parvin Mousavi, PhD, is the Director of the Medical Informatics Laboratory and a professor at the School of Computing at Queen's University in Kingston, Ontario, Canada.

Read more interviews with women scientists

First of all, Parvin, happy 8th of March!

Thank you very much!

Today it's International Women's Day so it's a great day to speak to you. When we were kids, we used to go in the streets of Milan on the 8th of March and celebrate.

Really? When I was a kid, I wasn't aware that there was such an occasion. I'm very impressed to hear that!

We used to bring flowers to women on the 8th of March, a yellow flower called mimosa.

That is so lovely to hear. Italians always do everything right!

Thank you for saying that! I will ask you first about your work, Parvin. Can you tell us what you do?

Yes, I am the Director of the Medical Informatics Laboratory at Queen's University in Canada. I am a professor at the School of Computing as well. The focus of my lab is on the applications of

Photo: Bernard Clark

"We can better impact patient care if industry is with us!"

machine learning for computer aided diagnosis and interventions. A lot of the focus is on cancer diagnosis as well as guidance of needles for biopsy in cancers.

"... you need a community to work with you and support you to help make everything a success!"

What about your students?

All of the work, in fact, is done by students, post-doctoral fellows, researchers, both undergraduate and graduate. All of the success is attributed to the contribution of these people. I could never do what I do in isolation. My work requires a lot of hands-on, in the clinic data collection, working with hardware, building software, as well as algorithm design. For this, you need a community to work with you and support you to help make everything a success.

What is the main quality that you admire in your students?

I would say that there are many qualities. Everyone is so different and every student has their own qualities that I admire. I was just writing, before speaking to you, a few paragraphs about the training of students for my Dean. I was truly impressed by my students. Every single one has a scholarship. Every single one has done something that has won prizes in their very own different ways. I think the quality that I'm most proud of is how aware they are of the role they have to make a difference in the world. They are really keen to learn more. They are really keen to do something that makes a difference. I think that's what really brings them to my lab and my work. They don't want to just work on e.g., deep learning or hardware because it's cool. They want to advance themselves, but they want to make a difference. I think that is what I really admire in them.

What do you think that they admire about your lab?

It would be very selfish to think that it is me that they admire. I think mostly it is the ability and the focus of the lab in actually doing interventions in clinic. Our health, our lives, and our mortality is something that we have dealt with as humans for thousands of years. We are always fascinated by this and the reason why we strive to make a difference in people's lives and patient outcomes. I think the other thing is working with like-minded people. I feel that, in the lab, good students, students that are cooperative, excited, compassionate, and committed, bring in like-minded people. When I want to interview a student or if a student is interested to come in and work with us, usually I get them to speak with other students in the lab even if they don't want to. I ask them to think about it because it gives them a good feel for who their colleagues and friends will be. That seems to work


Photo: Dave Dove

well. I think that's one of the reasons they want to come work for my lab.

My boss did the same when I started working at RSIP Vision, so I can relate. How old were you when you left Iran?

I was very young. I left Iran after my first degree. In my year of graduation from high school, I was the youngest student in my country to enter university I think, I was 15. I went to Imperial College and then I came to the west coast of Canada, the University of British Columbia, to do my PhD.

Have you ever thought of how your career would have been different if you had stayed in Iran?

[*laughs*] You ask good questions! You ask me about my parallel universe, if I had lived a different life. I think it would have been very similar. I don't think that there would have been very many differences in what I wanted to do or where I was going to go. I wanted to be an academic. Likely I would have ended up somewhere in academia working with students and trying to make a difference. It would have just been in a very different environment focused on other things. I'll give you an example. As an undergraduate, I worked in a research lab focused on the control of prosthetic arms. It was an important topic at that time, following several years of war. There were casualties, young people that needed help using their limbs again. This would have made a difference to those people. I think I would have done very similar things, just in a different context, maybe focusing on different priorities in healthcare that every region or country has.

Probably you would have arrived to IPCAI anyway. Let's talk about IPCAI, which is coming up in a few months. Do you want to tell us what makes IPCAI important and what will happen there?

IPCAI is a conference that started when people felt a need for like-minded people to get together and focus mainly on information processing for computer-aided interventions. It is a very high-quality conference. Papers are thoroughly reviewed. In fact, all the papers that are accepted are ready for journal publication. We have put in place a dedicated team of program chairs and area chairs that work really hard to provide a program that is unique, that is novel, that is exciting. Something else that is quite exciting is for every paper that is accepted at IPCAI, the authors get an opportunity to have an oral presentation and poster presentation. At our posters, we get a chance to have in-depth discussions. Others like the chance to be on the podium and impress their colleagues about something exciting. At IPCAI, we give them the best of both worlds. They get to be on the podium.

Photo: Dave Dove

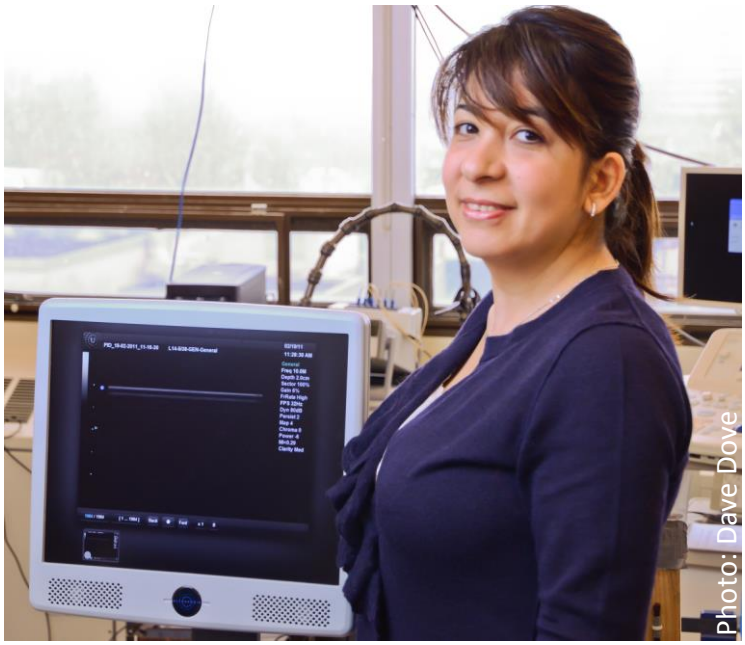


Photo: Dave Dove

They get to also have posters. For students, in particular, we have many awards given the size of IPCAI. I think it's a great opportunity for them to show, to shine, to get recognition and feedback, and hopefully to help them network and move on to great positions and improve the IPCAI community. We are also fortunate to have the support of CARS, Computer Assisted Radiology and Surgery. We are co-located with CARS which, as you know, was headed and chaired by Heinz Lemke for many years. This year CARS will be hosted for the first time by Pierre Jannin in Rennes. Both of them have been incredibly supportive and essentially key to the success of IPCAI. Pierre has been one of the original people that started IPCAI. He has been extremely supportive of me and of IPCAI for many years. This year he is chairing CARS. I have no doubt that we will see an outstanding conference. IPCAI, as a side effect, will benefit immensely as it has over the years from Pierre's leadership.

A last word about IPCAI: if some of our readers are thinking about coming there and are hesitating, what do you tell them?

I think you will not regret coming there. First of all, the size is relatively small. You get to interact with the community. The community is extremely supportive. They all show up. They engage. They provide feedback, but it is all very positive. They do realize that our job is to support students and their research and to provide feedback in a constructive manner. If any students or faculty researchers are interested in IPCAI, I encourage them. Maybe this is the year to start attending. Our call for papers happens in the fall, and that is a great opportunity for them to consider. Their papers, if accepted, are most likely to be published as a journal paper. That is also an excellent, added benefit.

"... to provide a program that is unique, that is novel, that is exciting!"

Are you happy about the industry's involvement in CARS, IPCAI, and MICCAI? Don't you think we should see more industry involvement in the coming years?

That's very perceptive of you Ralph! I wonder if you would consider coming to advise us! [*laughs*] You see it from a different perspective! That's a very smart observation and question. We are lucky that the industrial supporters that we have, for instance, NDI, Siemens, Philips, and other supporters that we have had over the years at IPCAI. For example, Andrew Wiles from NDI has provided supported IPCAI for many years in terms of sponsorship, reviewing, and attending when he can.

As much as we love the sponsorships and the support of sponsors, we encourage them to participate in the conference and interact with the students. Other sponsors such as Intuitive also participate in IPCAI in addition to sponsorship and we are looking forward to more of this in 2019. I think, overall, it improves the quality of the conference. We are also interested in seeing more of scientific paper submissions from industry. I realize that not all of the companies in this area focus on research. They often concentrate on developing products and sales and service but if they have research, it would be very nice to have that presented as well. It would be helpful for the community to realize how we can better impact patient care if industry is with us.

Do you have a message for PhD students? What would you like to tell those students who have difficulties or experience crisis and sometimes doubt about themselves? What is your best advice to succeed?

I would say that if you ask any single one of us, we would say, "I have been there. I have gone through similar experiences." First, reach out for support. Do not go into crisis alone. Your supervisors, your colleagues, your department, your university, they all share a purpose: to help you succeed in your studies. You are not alone. Every other student has most likely either gone through similar things or will go through similar things. For PhD, in particular, it is a long run. There are many unknowns. Otherwise, it wouldn't be research. I would say, you have to be persistent. You have to focus on the end goal. Think about why you came here. What was your motivation to start your PhD? That is always a good support. Reach out. In conferences, look for people to mentor you. If you feel like you are lacking some aspect of mentorship, reach out to your supervisor. Reach out to your colleagues and other faculty members. It will help.

"We are also interested in seeing more of scientific paper submissions from industry"

Would you like to tell us more about your research?

I want to talk about some of our deep learning methods and our machine learning approaches that have helped us guide interventions for prostate cancer. Prostate cancer is a prevalent health issue in the western world. In many countries in the western world, it's either the second or the third cause of cancer-related death in men. The key here is to be able to accurately diagnosis high-grade cancer. Right now, ultrasound is used to guide biopsies for the prostate. The final diagnosis is made based on the analysis of the biopsy. What usually happens is that ultrasound is blind to the existence of cancer. It's only used to guide the needle

to the anatomy. You need additional information to tell you where to biopsy. That additional information is starting to come from MRI (Magnetic Resonance Imaging). However, MRI is done prior to biopsy and in order to integrate that information into the biopsy process, image registration is required. Your readers will know all about registration of preoperative and intraoperative images, which is challenging to start with. Otherwise, we wouldn't have registration tracks and papers at MICCAI for so many years. Also, during biopsy, patients are awake. Every poke you make with a needle, the patient will jump. Then the registration will be lost. It is very important that what you have guidance technology that can make up for this. Most of my research in ultrasound in prostate cancer is focused on a new technology called Temporal enhanced Ultrasound (TeUs). With this technology, we take consecutive images for a short time of the tissue without intentional motion of the tissue or the imaging probe. From this time series, we are able to detect signatures of tissue that are associated with pathology. How could this help the biopsy? You could still use MRI information that you have taken pre-operatively. Now, with the information that we create in real-time during biopsy from TeUS, we can provide a color map. The color map shows how likely the patient is to be cancerous. If you were to biopsy, and even if the patient moved, the color map can be regenerated in real-time. You can actually compensate for the dissociation of your pre-operative images because you have real-time description of the tissue. We are really excited about this work. We are now doing multi-center feasibility trials. I am really looking forward to see if this can make a difference in patients' lives, and that is what my lab strives for.


Photo: Bernard Clark

How long will it take to turn this into a clinical reality?

It takes a long time to see anything, research-wise, get adopted. For now, we are starting our prospective studies. We have done, so far, retrospective studies. It's important to show that prospectively this can work. Another important item to show is that it works in different centers equally well. Our first trials were at the National Institute of Health, their clinical center in Bethesda, Maryland. Now we are extending this to the west coast and east coast of Canada. It is very important to show that it works equally well with different machines, in different centers, in real clinical conditions, not in a research center. I think we have a few more years, but hopefully not too far, and hopefully before I retire!

Read more interviews like this

"I think we have a few more years, but hopefully not too far, and hopefully before I retire!"

Aswin Nadamuni Raghavan

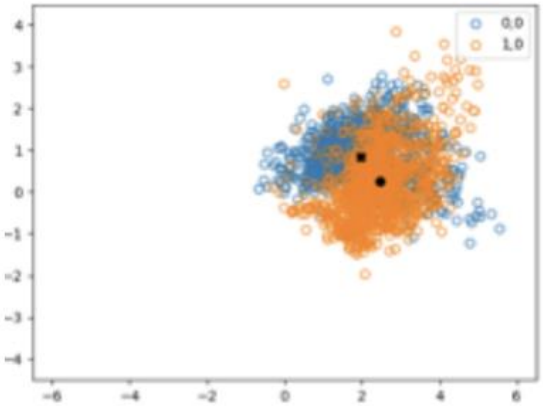
Jesse Hostetler

"Reduce the forgetting problem and allow for forward transfer, which means taking knowledge from earlier tasks and using that knowledge to help solve later tasks or tasks that the program hasn't learned yet"

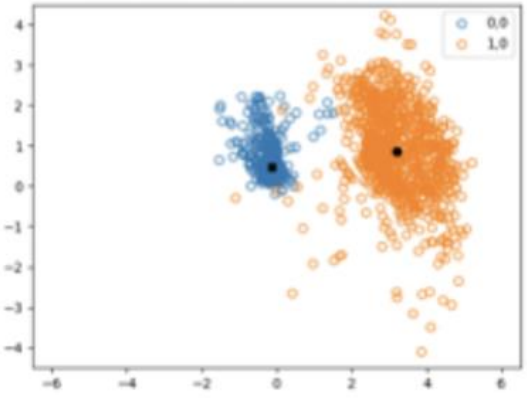
Headquartered in Menlo Park, California, **SRI International (SRI)** is a non-profit research institute that conducts research and develops innovations for government and industry. SRI's mission statement is focused on technology innovation and it has many years of innovation including interactive computing and medical ultrasounds to cancer drugs and beyond. SRI's team of over 2,100 scientists, engineers, technologists, and policy researchers from over 20 countries work together to create world-changing technology.

Advanced computer scientists, **Aswin Nadamuni Raghavan** and **Jesse Hostetler**, both work as a part of the Center for Vision Technology group at the SRI, located in Princeton, New Jersey. They focus on developing a next-generation artificial intelligence (AI) system that relies on continuous learning to become better and more dependable at performing new tasks. The team at SRI supports a number of projects from commercial and government agencies, including **DARPA (Defense Advanced Research Projects Agency)**.

Aswin's work covers various fundamental research problems mostly related to deep learning. He explores problems that deep learning cannot yet solve as well as how to make deep learning networks be smaller and operate faster to function on edge



Overlapping tasks in latent space (blue/orange) makes Lifelong RL training prone to forgetting



Our generative memory and task encoding separate the latent space to ensure efficient memory use to avoid forgetting.

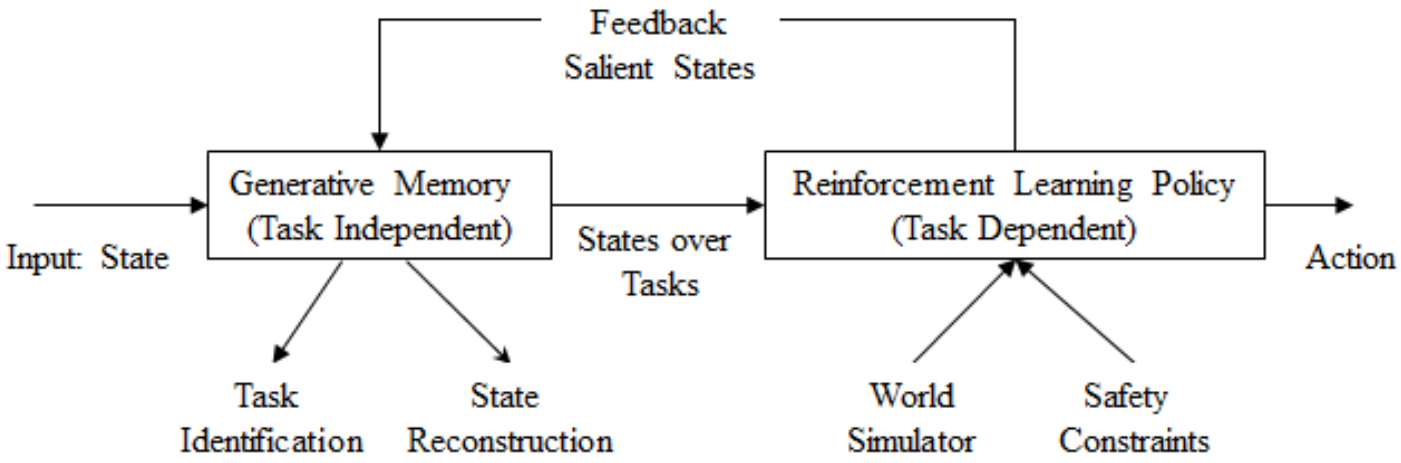
devices. For a long time, he has worked on the challenging problem to makes neural networks more compact and faster using low-bit precision techniques. He also writes papers and codes as well as grant proposals. In addition, Aswin works on a more applied project called **AEOSP** for **DARPA Communicating with Computers (CwC)**, which functions as a visual storytelling platform. During the interactive process, the user types instructions and uses natural language. Then the system turns the story into a visual movie.

Aswin and Jesse enthusiastically speak about their work on **Lifelong Learning Machines** or **L2M**, a program funded by **DARPA MTO (Microelectronics Technology Office)** and led by **Hava Siegelmann**. The program looks at machine learning in a fundamentally different way that addresses the many drawbacks of deep learning and assumptions surrounding machine learning.

Currently, deep learning can solve one task but then forgets all of the previous tasks, a process referred to as catastrophic forgetting. The L2M project aims to reduce the forgetting problem and allow for forward transfer, which means taking knowledge from earlier tasks and using that knowledge to help solve later tasks or tasks that the program hasn't learned yet.

Because of the difficulty of lifelong learning and solving the forgetting problem in neural networks, researchers have put aside the problem for several decades. The team at SRI are tackling the high-dimensional problem using a strategy video game, by Blizzard Entertainment called **StarCraft2**, as a simulation environment for lifelong learning.

StarCraft requires players to switch between solving different tasks. First, the player has to understand how to play the game, a process more difficult than simple image classification. Then, the user must learn to control the game units. The game also involves multiple feature maps that is similar to computer vision and image processing professionals are accustomed to seeing and similar to what the human actually sees when playing the game. Although a Google DeepMind bot recently defeated



the best human players, deep learning still falls short in efficiently shifting between tasks, and especially when the bot is faced with new surprises that is not presented in the training data.

The SRI team is approaching the AI challenges with deep reinforcement learning (RL) problem. In terms of algorithmic techniques, the research team uses memory as a way of addressing the problem. They rely on generative modeling techniques to model memory since storing the experiences doesn't scale up to the size of the problem or the number of different tasks experienced. They encode memory into separable latent space, and then generate those experiences (e.g. like memory recall) to retrain the deep RL algorithm. They are using the StarCraft2 game simulation environment to provide a lifelong learning training environment.

As Jesse explains: *"Our approach to lifelong learning is based on the idea of memory, remembering past experiences, and using those to help in new situations and to re-experience past situations so that you don't forget about them. Our approach is based on the way that the biological memory systems have been found to replay past experiences in a mammal brain. We're creating an analogous bio-inspired system that learns how to encode experiences and then reuse them in a process called pseudo-rehearsal which allows us to transfer that knowledge to new tasks and to re-*

experience previous tasks."

Aswin further elaborates: "*One key technique we use is generative modeling to learn and model experiences related to a task. The generative memory supports recall of the most relevant memory to your current situation. The second big technique is reinforcement learning, specifically multi-task reinforcement learning. It lets us transfer these memories to an actionable representation of what the agent has to do in a new, unseen task.*"

Currently, the project is currently at an early research stage, but eventually, it could have many commercial applications for creating a technology that would not require retraining. Lifelong learning has several applications especially now that deep learning applies to all kinds of devices and applications. As Aswin explains: "*Our lifelong learning technique extends beyond StarCraft. StarCraft is a very hard problem that gives a faithful representation of the capability that we are developing.*"

Typically, scientists train a deep learning model and apply it to a device such as the iPhone, Siri, Alexa, or an autonomous car with the hopes that the model will capture all of the situations that will arise in the device. Lifelong learning strives to continue the learning process even after deploying the device in order to handle any surprises.

Jesse adds, "*I'd say that lifelong learning is part of a broader goal of making learning systems have more human-like capabilities, especially when it comes to making sense of complex information, learning over time, building a knowledge base. I'd say that the way that this fits into broader machine learning is that so far, machine learning mostly focuses on solving particular tasks, but when we have tools like lifelong learning, we start to think about much more human-like capabilities in terms of building up skills, knowledge, etc. over a lifetime.*"

Feedback of the Month

The team at RSIP has been great to work with, in supplementing some of our internal computer vision engineering efforts and meeting aggressive deadlines!

Modar Alaoui
Founder & CEO, eyeris.ai

by Shmulik Shpiro

Shmulik Shpiro is RSIP Vision's EVP Global Business Development & Strategy. Shmulik visited this year's ADAS SENSORS show in Detroit. Here is what he found out regarding the autonomous driving landscape and the Advanced Driver-Assistance Systems (ADAS) field as a whole.

Among the many things which I have seen at the show, there was a very interesting mix of exhibiting vendors: first, the famous tier-1 and tier-2 manufacturers; second, some smaller and younger companies, among which were quite a few startups presenting cutting edge technologies around automotive sensors, many of them working in LiDAR and advanced cameras or innovative lenses; third (perhaps the most remarkable part), there were service providers, service centers and even car dealers. All this field is going to change with gigantic step. How each segment of the market will be impacted? Everyone wants to have a part in shaping the new landscape. All of this was surprising and exciting! One of the presenters was a service center for replacing glasses and windshields and they talked for almost half an hour about the challenge of recalibrating the sensors after a glass replacement! How do you make sure that the sensors are still working after that? Even the most sophisticated and expensive sensors might have a tough time working with the new car geometry. More generally, how do you make sure that all ADAS features keep working after even a small accident?

The last few years we have seen huge resources being invested in this field: there are more than 70 companies developing LiDAR technologies and an almost infinite number of companies working on sensors in general. All this necessarily means that something big is happening. Even if only a share of these companies is going to stay in business and succeed, the movement which is going on is grandiose! Another thing which I want to point out is: what kind of sensors are we going to use? What capabilities will they have? How will we consolidate these sensors? In fact, the number of sensors must be finite, since all of them need electricity in order to function, they have a cost and they need to be maintained. At ADAS SENSORS 2019 we could see this consolidation going on. Probably it's not for tomorrow and even not for next year, but I bet that in a couple of years we'll see the landscape considerably changing.

Calmcar Demo Car, presenting cutting edge AI capabilities, embedded on different kind of sensors

I have seen visitors coming from all around: from the West Coast and the Silicon Valley, from Canada, from Europe and also a considerable group from Asia, mainly Chinese, Japanese, Korean and more, including several manufacturers of components, chipsets wires and connectors. There were representative coming from all kind of research institutes: all the automotive ecosystem was there!

I would also like to talk about current and upcoming trends. The show made it clear that we are going to experience a two-stage transformation: the first is already happening, as we saw a couple of months ago at CES 2019, with autonomous taxis driving along the Strip. All this is going to be the robotaxi: like Uber, Lyft, Zoox and others. They do not give much importance to the look of the car as much as they care for its functionality. Indeed, they have specific needs. Let's say it's a huge bulk with many sensors around. Their business is to drive you from point A to point B as safely and as quickly as possible. That's the first stage. The second stage is going into the private and personal cars and that's the biggest challenge: you will need to have all those capabilities, but mixed in an endless variety to fit the tastes and needs of an almost infinite multitude of car owners. How is the look and feel of the car going to change, if it's going to be autonomous? What is the owner going to do, if some of the sensors and capabilities will not be available?

The first stage is the real change, because as soon as we will have millions of robotaxis driving around, everything will be different forever. Following that, we will see autonomous trucks (like TuSimple) and truck platoons. That's what we are going to see, maybe not in our own garage, but all over our cities and roads.

"Everything will be different forever"

RSIP VISION

Global Leader in Computer Vision & Deep Learning

COMPUTER VISION PROJECT MANAGEMENT

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

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

- How to implement Deep Learning
- Team Leadership and Management
- Validation and Test Techniques
- How to solve all kinds of challenges
- What are the best practices?

"Even the biggest hammer cannot replace a screwdriver!"

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

FREE SUBSCRIPTION

Dear reader,

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

Subscription Form
(click here, it's free)

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

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

Bay Vision Meetup: Artificial Intelligence in Medical Imaging
Cupertino, CA Apr 4 Website and Registration

ISBI 2019 - IEEE International Symposium on Biomedical Imaging
Venice, Italy Apr 8-11 Website and Registration

European Symposium - Artificial NN, Comput. Intelligence and ML
Bruges, Belgium Apr 24-26 Website and Registration

AI & Big Data Expo Global
London, UK Apr 25-26 Website and Registration

Automatic Face and Gesture Recognition
Lille, France May 14-18 Website and Registration

ICRA - International Conference on Robotics and Automation
Montreal, Canada May 20-24 Website and Registration

RE•WORK - Deep Learning Summit (+ Healthcare)
Boston, MA May 23-24 Website and Registration

Digital Pathology & AI Congress: USA - 2019 meeting
New York City, NY Jun 13-14 Website and Registration

CVPR - Computer Vision and Patter Recognition
Long Beach, CA Jun 15-21 Website and Registration

CARS - Computer Assisted Radiology and Surgery
Rennes, France Jun 18-21 Website and Registration

AI & Big Data Expo Europe
Amsterdam, Netherlands Jun 19-20 Website and Registration

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

Did you read the Feedback of the Month?
It's on page 27!

FEEDBACK

Dear reader,

If you like Computer Vision News (and also if you don't like it) we would love to hear from you:

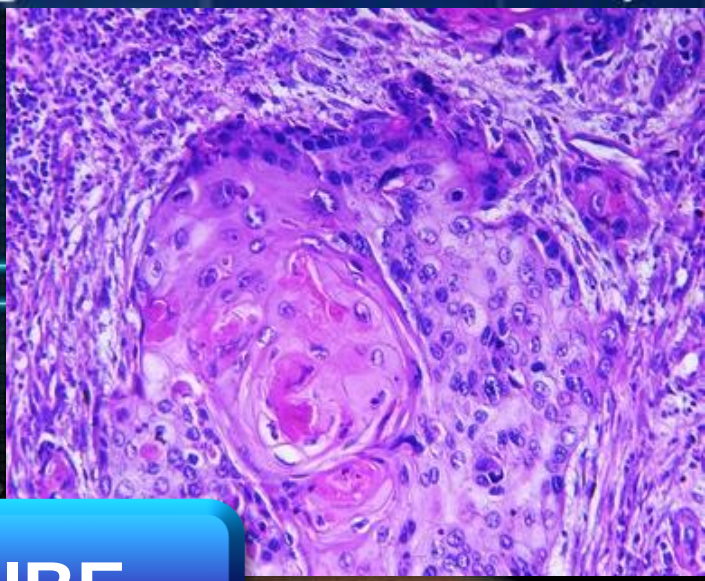
Give us feedback, please (click here)

It will take you only 2 minutes. Please tell us and we will do our best to improve. Thank you!

IMPROVE YOUR VISION WITH

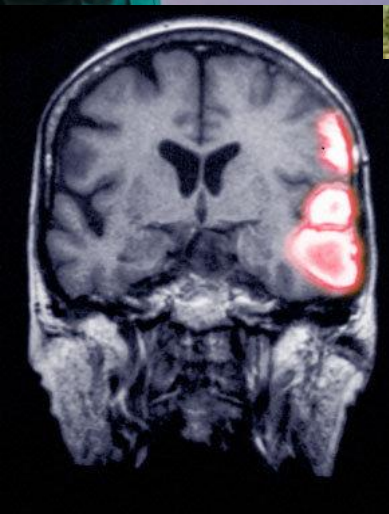
Computer Vision News

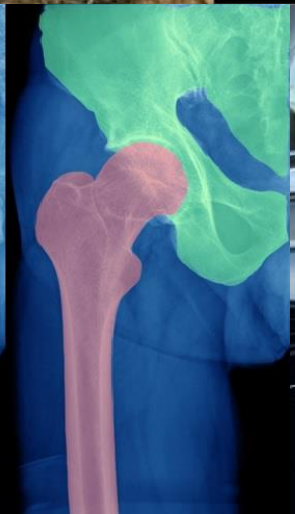
The Magazine Of The Algorithm Community



SUBSCRIBE

CLICK HERE, IT'S FREE





A PUBLICATION BY

RSIP VISION

Global Leader in Computer Vision and Deep Learning