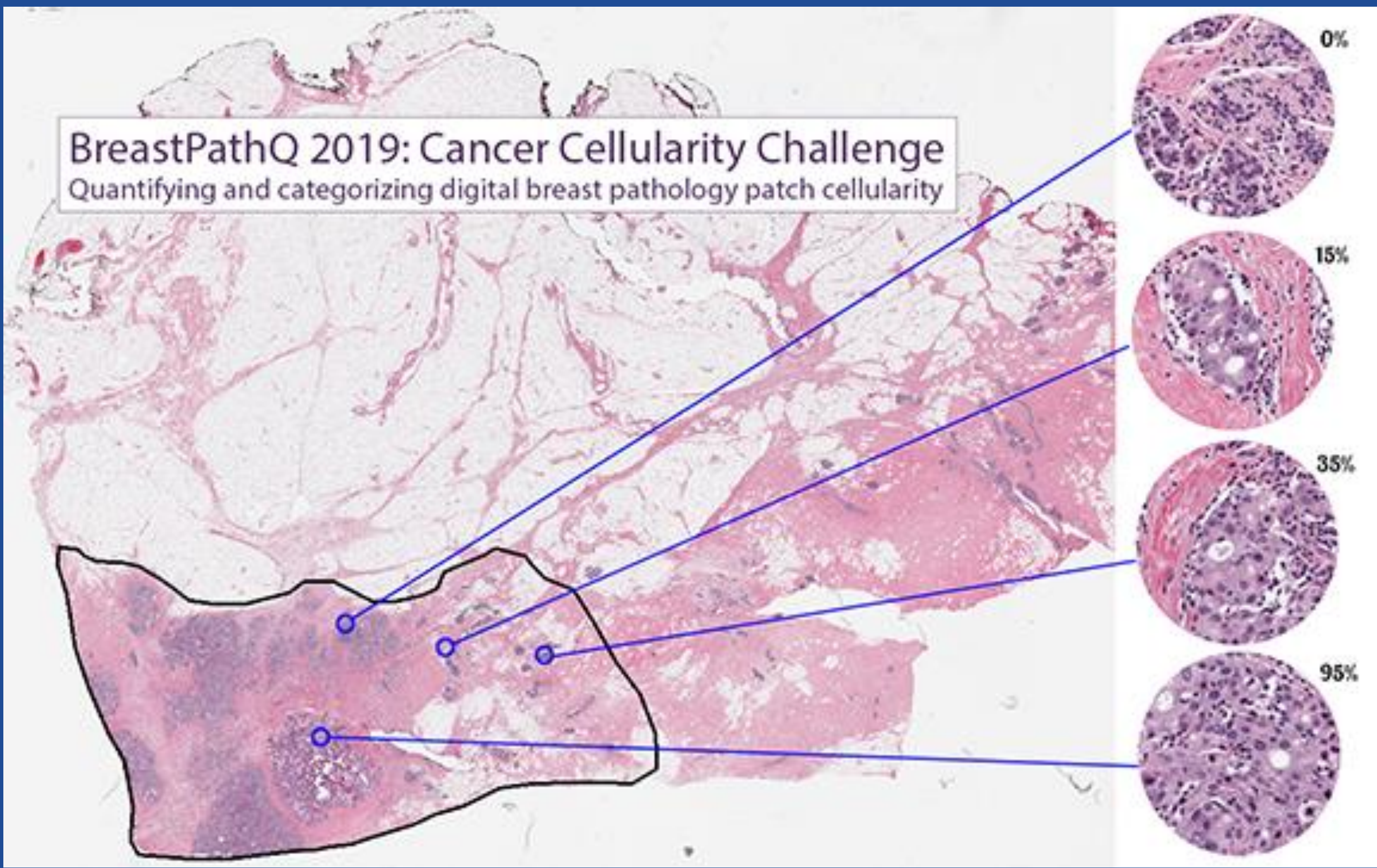
Computer Vision News

The magazine of the algorithm community

A publication by RSIP Vision

March 2019



Artificial Intelligence Spotlight News

Bay Vision Meetup

Women in Science:
Chip Huyen - NVIDIA

Research Review:
Kalman Filter Based Classifier Fusion for Affective State Recognition

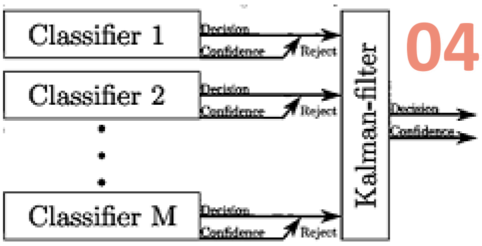
Focus On:
PyOD (with codes!)

Project Management:
When Client Says: “Do the Best You Can!”

Upcoming Events

Computer Vision Project:
3D Stereo Vision Calibration and Triangulation

Contents



04
Research Paper Review
... of Kalman Filters for...

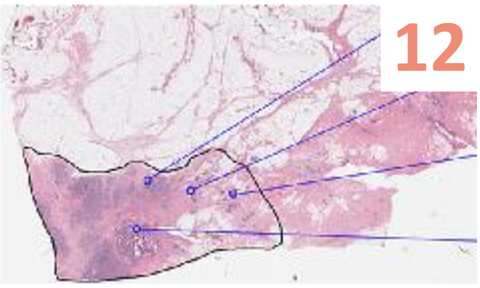
15
AI Spotlight News

24
Women in Science
Chip Huyen - NVIDIA

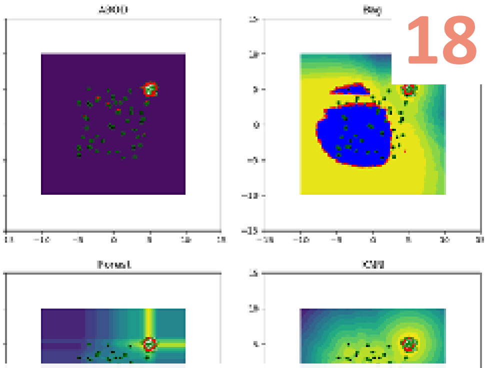
10
Project Management Tip
"Do the best you can!"

16
Project by RSIP Vision
3D Stereo Video Calibration

30
Bay Vision Meetup



12
Challenge - BreastPathQ
with Nicholas Petrick



18
Focus on: PyOD
Python open source toolbox

33
Upcoming Events

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,

Following our issue dedicated mostly to the field of **Autonomous Vehicles** (February 2019), we keep informing you about the most interesting developments in that area. For instance, the **Bay Vision Meetup** that we sponsored on February 27 in Cupertino: see the photos and the full video of the discussions on pages 30-31 of this magazine.

We did not forget the other key subjects normally featured on **Computer Vision News**: we reviewed for you a very important **medical segmentation** challenge on page 12, thanks to help from **Nicholas Petrick** of the **FDA**: the quantitative task of the **BreastPathQ challenge** is to determine cancer cellularity from H&E patches of **breast cancer tumors**. Read also what the **winners** of the challenge say.

Finally, we know how much **our readers love codes**! If you relate, don't miss the technical articles published in this issue, in particular the "**Focus on: PyOD**" on page 18. As usual, lovers of **Artificial Intelligence** will find plenty of code to work with.

Enjoy the reading!

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

by Assaf Spanier

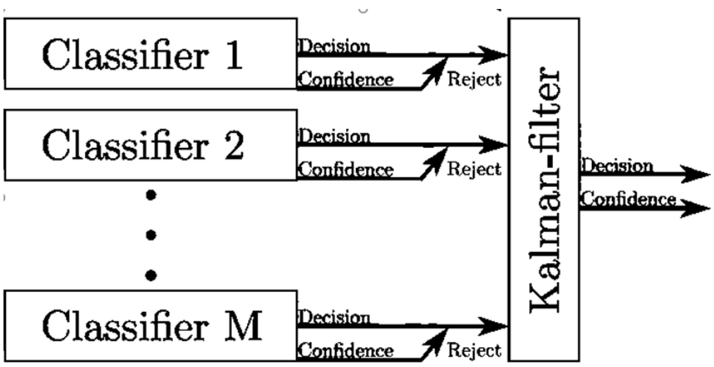
Every month, Computer Vision News reviews a research paper from our field. This month we have chosen **Kalman Filter Based Classifier Fusion for Affective State Recognition**. The authors are **Michael Glodek, Stephan Reuter, Martin Schels, Klaus C. J. Dietmayer** and **Friedhelm Schwenker**. The paper is here.

An implementation of Kalman Filters for fusing the decisions of a number of classifiers

Introduction

Kalman Filters are a method in widespread use in the fields of object tracking and autonomous driving (navigation). Kalman Filters are highly efficient at fusion of measurements, due to their Markov chain based design (Markov assumption holds). The idea of Kalman Filters is to **reduce measurement evaluation "noise" by fusing measurements from a variety of sources, even if some have missing values in certain instances, and computing what weight to give each source in each instance**. The model can handle missing measurements by raising the level of uncertainty. Kalman Filter works in two stages: the prediction stage and the update stage. The prediction stage estimates a scalar -- the fusion of classifier outputs. The update stage (also known as the correction step) combines this estimate with the current measurements z.

The classifier fusion method proposed in this paper can be used for any classifier type, as long as the following assumptions hold: 1) Markov assumption: future states are independent of past states. And 2) the data is sequentially structured.



The figure above illustrates the model proposed in this paper: classifier fusion of a number of simple classifiers (base classifiers), with a reject option for each classifier. At every time point, the classification decision and confidence measure of every classifier are collected and fused by a Kalman filter, which then outputs a fused classification decision with a confidence measure for that decision.

Method

As mentioned, the Kalman Filter consists of two stages: the prediction stage and the update stage. The prediction stage computes an estimated scalar $\widehat{x_t}$. The update stage (also known as the correction step), fuses this estimate ($\widehat{x_t}$) with the latest measurements z_{mt} , where $t \in \{1,...,T\}$ is the time and $m \in \{1,...,M\}$ is the m-th classifier.

The prediction stage: the new prediction $\widehat{x_t}$ is estimated based on the last prediction $\widehat{x_{t-1}}$, according to the formula $\widehat{x_x} = f \cdot x_{t-1}$. The original Kalman Filter formula includes an update element *u*; however, it is meaningless here, as it refers to a command or instruction given to a robot or autonomous vehicle, and the paper deals with a classifier.

$\widehat{p_t}$ is the covariance of the prediction and is equal to $\widehat{p_t} = f \cdot p_t \cdot f + q_m$: it is obtained by combining the a posteriori covariance with an additional covariance q_m , which models the process noise:

The update stage is performed for each classifier *m* and it consists of the following three sub-component computations:

$$y = z_{mt} - h \cdot \widehat{x_t}$$
$$s = h \cdot \widehat{p_t} \cdot h + r_m$$
$$k = \widehat{p_t} \cdot h \cdot s^{-1}$$

where *h* is the observation model mapping the prediction to the new estimate, and r_m is the observation noise.

The outcomes of this stage are used to update the estimate and covariance of the next stage, based on the following formula:

$$x_t = \widehat{x_t} + k \cdot y$$
$$p_t = \widehat{p_t} - k \cdot s \cdot k$$

When there are missing values, that is for some classifiers there is no decision -- the output of that classifier (z_{mt}) is set as 0.5 and the observation noise r_m of that classifier is raised. The authors expanded the classifiers' decision range to include the possibility of rejection, that is, classifiers may decide to return a no-classification output, rather than their estimate, based on a too-low confidence measure. Due to the temporal structure of the method (the fact that there are a number of classifications at every point in time), the missing classification results can be estimated based on the results of the other classifiers. The architecture proposed in the paper is as follow - at every point in time, each classifier *m* produces a classification decision and a corresponding confidence measure, which serves for making rejection decisions.

Dataset

The paper used the AVEC dataset, presented in 2011 as a benchmark for user emotional states. The data are audio and video recordings of a human communicating with a virtual agent and trying to stimulate one of four emotional states: Arousal, Expectancy, Power and Valence. Ground truth was determined by having eight human evaluators evaluate each recording on a continuous scale, with the final binary labeling determined by applying a threshold to the average evaluation.

The Base Classifiers

Below are the features used as input by the base classifiers. Each classifier was trained either on audio recordings or video recordings, and optimized through cross-validation with other classifiers for the same recording type.

Audio:

To arrive at a fixed length input vector (from a long continuous varying input), HMM-based transformation is used. Classification is done by five bags of random forests, and the final decision is determined by averaging the five trees, with the standard deviation used to compute the measure of confidence.
Audio classification is conducted per-word using 3 bug-of-words composed of the following features:

- Fundamental frequency, the energy and linear predictive coding (LPC)
- Mel frequency cepstral coefficient (MFCC)
- Relative spectral transform - perceptual linear prediction (RASTA-PLP)

Video:

Video channel features were acquired from the computer expression recognition toolbox (CERT), dedicated to facial expression recognition. Four models from the CERT toolbox were used: Basic Smile Detector, Unilaterals, FACS 4.4 and Emotions 4.4.3. The outputs of all four models were concatenated to create a length 36 vector for every video frame.

The overall classification decisions and confidence measures were determined similarly to the audio by using five bags of random forests. In 8% of cases facial recognition failed leading to missing results for the base classifier, which the Kalman Filter handled. The hyperparameter settings of all base classifiers underwent optimization using the training and validation datasets. To optimize the hyperparameters of the classifier fusion algorithm (Kalman Filter) the same training and validation datasets were used, this time treating the classification decisions and confidence measures of each base classifier as features.

The model fuses a large number of measurements and can handle missing values by increasing the noise level for that classifier

Results

Audio and video classification pre-fusion performance. The results are percentages with standard deviation. The video results are per frame. Table shows performance of the audio and video classifiers separately, without classifier fusion.

Audio	Arousal	Expectancy	Power	Valance
↑Acc.	61.8±3.6	59.0±6.3	57.5±9.4	57.5±7.9
↑F_1	65.8±3.8	16.4±7.1	69.6±9.3	70.1±6.8
↑$\overline{F}_1$	56.7±3.4	72.6±5.2	24.7±6.6	24.9±8.4
Video	**Arousal**	**Expectancy**	**Power**	**Valance**
↑Acc.	57.0±4.2	54.8±4.0	55.7±2.9	59.9±7.4
↑F_1	60.8±5.1	49.6±9.4	57.5±11.3	67.1±11.5
↑$\overline{F}_1$	51.3±9.3	56.6±10.7	48.8±12.3	43.5±7.1

Images without a classification decision, such as when the person is not speaking or facial recognition failed, were excluded for the purpose of this evaluation. The table presents precision and the F1 rate

$$F1 = \frac{2 \cdot P \cdot R}{P + R}/$$

(where P is the precision and R is the recall) for the four emotional state categories. Compared to the best performance previously achieved on the benchmark these results are already impressive (for instance, the best precision for predicting 'Arousal' was 61%). The next two tables present performance for the four categories with uni-modal classifier fusion -- that is fusion of the audio base classifiers separately and the video base classifiers separately.

Audio classification performance after fusion using Kalman Filter:

Audio	Arousal	Expectancy	Power	Valance
↑Acc.	74.3±6.6	57.5±6.2	56.5±11.2	59.7±11.0
↑F_1	77.4±6.5	12.5±6.2	69.1±10.2	72.9±8.5
↑$\overline{F}_1$	69.4±8.2	71.8±5.3	22.7±8.9	16.2±14.1
Reject	10%	90%	90%	50%
q_{audio}	10^{-7}	10^{-7}	10^{-5}	10^{-4}
r	0.1	0.1	0.1	0.75

Video classification performance after fusion using Kalman Filter. The table below presents video-channel results. Here, all categories show improvement compared to the pre-fusion results. Note, however, that Arousal achieved better results using the audio channel:

Video	Arousal	Expectancy	Power	Valance
↑Acc.	64.5±2.9	59.1±4.7	58.9±5.0	65.4±9.8
↑F_1	67.9±4.2	56.7±11.1	60.6±13.8	73.3±11.4
↑$\overline{F}_1$	57.3±12.7	56.4±17.5	48.2±17.1	44.2±8.5
Reject	50%	90%	0%	90%
q_{video}	10^{-7}	7.5^{-6}	5^{-5}	10^{-4}
r	0.1	0.1	0.1	0.75

Reject indicates percentage of decisions rejected, q_{AUDIO} and r correspond to the process noise and the observation noise, respectively. z_{mt} is set to 0.5 and observation noise r_m is set to a higher rate than originally assigned to r.

At first glance, comparing the results of the second table ("audio fusion") to the first table -- no-classifier-fusion -- it may seem only Valence and Arousal show improvement. However, the most important improvement in evaluation is that now there is a classification decision for every video frame.

The table below presents results for the fusion of audio and video channel data. This multi-modal fusion improves performance for Power and Expectancy. However, Arousal performance was best fusing only the audio data, and Valence performance was best fusing only the video data. The lowered performance of Arousal is likely caused by an imbalance between the audio and video. Audio channel decisions aren't always available, while the smart platforms of the video channel are almost always available. In the case of the Valence category, the lowered performance can be attributed to the relatively low F1 of the audio channel.

Audio-visual	Arousal	Expectancy	Power	Valance
↑Acc.	68.5±5.7	62.5±4.9	61.8±6.6	64.2±9.3
↑F_1	72.6±4.2	42.2±15.7	69.1±7.6	72.6±10.9
↑$\overline{F}_1$	59.7±15.1	71.1±5.8	43.5±18.0	43.7±3.0
A. rej.	0%	0%	0%	90%
V. rej.	50%	50%	0%	10%
q_{AUDIO}	10^{-6}	5^{-6}	10^{-7}	5^{-5}
q_{VIDEO}	10^{-5}	10^{-5}	10^{-5}	10^{-4}
r	0.75	0.1	0.1	0.75

Conclusion

The paper presents an implementation of **Kalman Filters** for fusing the decisions of a number of classifiers. The authors show the feasibility of **fusing multi-modal classifier outputs**. The model fuses a large number of measurements and can handle missing values by increasing the noise level for that classifier. The authors used the audio/visual emotional challenge (AVEC) 2011 data set to evaluate the performance of the fused classifier. Fusion clearly improved performance for all four emotional state categories measured in the challenge. Moreover, using the Kalman Filter enabled the system to estimate the missing classification results, so that it was able to classify all data (all video frames). Despite the fact that the Kalman Filter can be considered the simplest instance of a time series (with no control matrix and assuming an identity matrix for dynamics), **the results presented in the paper are excellen**t.

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**

When the client says: “Do the best you can!”

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 **Aliza Minkov** tells us what to do **when the client says: “Do the best you can!”**. It’s another tip by **RSIP Vision** for Project Management in Computer Vision.

“…a good way of knowing if you are in the good direction…”

Besides scientific and algorithmic challenges, computer vision work presents also project management challenges: we are going to discuss in this article the specific circumstance of projects that do not have well-defined quantitative goals. Since this is a research project, neither client nor supplier can predict results *a priori*.

One of the first tasks which needs to be dealt with in every project is the need to define our goals, such as would be accepted and understood by both parts, the client and **RSIP Vision**.

In projects of this kind, customers start with some idea of what they want to achieve, but for lack of knowledge of what is possible to achieve, they often expect the service provider to supply the specific operative goals of the project: in other word, how to measure achievements and success at the end of the work. When the objectives are clear and understood, it makes it easier to verify whether we are getting closer or not.

When this does not happen, the lack of operative goals has an effect on the whole process. For instance, if you define clearly the goals that you want, you can define the constraints that are upon your work. Some of them are straightforward and some are not: time, amount and quality of data needed. Ideally, the goals of a deep learning project in computer vision should include the level of accuracy that we want to achieve. Sometimes, at the time of the **POC**, the client has no clear idea of the desired accuracy. That’s equivalent to saying: “**Do the best you can (with the given constraints)!**”. That absence of ground truth might lead to overshooting, i. e. working more than necessary and spend more time than necessary to further improve the results.

Without it, gaps may be found in the design: for instance, we have the deep neural network, but how do we translate the output of the neural network to exactly what we need? When you have little time for experimenting, you cannot try everything: you need instead to **focus your effort**. You define an end-to-end procedure and can admit some errors in the procedure, if the final result is accurate enough. Again, the lack of ground truth denies the chance to evaluate the quality of the result. With

… you need to develop some tools that give you an objective assessment!

no early estimate of the performance, the team might not know how they are doing until it's almost too late to rectify the project and refocus the efforts.

In that case, you need to have **a regular contact with the client**: exposing what you do with all the steps that you have taken is key to keep the focus as close as possible to what the real expectations of the client are. Measuring performance as you progress and reflecting it to the clients is a good way of knowing if you are in the good direction to provide what they really want.

There's something that we learn from most computer vision tasks: you look at the image, you look at the network output and you decide if it's good enough or not. But sometimes you need to develop some tools that give you **an objective assessment**. Doing that about all the parts of the projects helps you decide where your efforts should be applied on.

More articles on Project Management

The SPIE (the international society for optics and photonics), American Association of Physicists in Medicine (AAPM), and the National Cancer Institute (NCI) developed a Grand Challenge for the quantitative task of determining cancer cellularity from pathology hematoxylin and eosin (H&E) slide patches of breast cancer tumors. Cancer cellularity is the percent area (0-100%) of the tumor bed that is comprised of invasive or in situ tumor cells and is one part of the MD Anderson residual cancer burden assessment approach with applicability, for example, in assessing neoadjuvant treatment of breast cancer. This intro and almost all the article are courtesy of Nicholas Petrick, Deputy Director Division of Imaging, Diagnostics and Software Reliability at the Center for Devices and Radiological Health, U.S. Food and Drug Administration. Nicholas led this challenge effort together with Kenny Cha (also from the FDA), Shazia Akbar and Anne Martel (both from Sunnybrook Research Institute) among others.

Challenge synopsis

The challenge included a training (2579 512x512 patches from 46 patients), validation (185 512x512 patches from 4 patients) and test phase (1121 512x512 patches from 18 patients) using digitized H&E stained breast cancer slides. The data was collected at the **Sunnybrook Research Institute, Toronto** as part of a research project funded by **Canadian Cancer Society** and IRB approval was obtained to allow the anonymized data to be shared. The training and validation phases included feedback to the participants on their algorithm's performance and had visible leaderboards allowing participants to compare their performance with other algorithms. The test phase was blinded without any performance feedback available to participants until after the challenge closed.

The performance metric selected was predication probability (p_k), which assesses the ordering of the patches between the method and the reference but not the absolute cellularity scores for each patch and ranges between 0.0 and 1.0. The reference standard ("truth") was a cellularity scores from one pathologist for the training/validation patches and cellularity scores from two pathologists for the test patches.

A total of **100 qualified algorithms were submitted from 37 different groups.** The winning algorithm achieved an p_k=0.94 with algorithm performance ranging from p_k=[0.21, 0.94].

Nicholas Petrick

What did we learn?

• We learned that participant engagement was critical. The BreastPathQ challenge include a forum section allowing participants and organizers to ask questions, answer questions, form teams and provide feedback to all participants at once. One example of how the forum was used was when a participant identified a problem with the original challenge performance metric. Once this was identified on the forum, we, as the challenge organizers, where able to assess the problem, determine it was a concern and implement a revised performance metric to address the issue. This is quite a nice example of how participant engagement directly lead to an improved challenge.

• We learned that paying attention to the performance criteria from comparing algorithms is crucial. We found that a rank-based performance metric, predication probability, would work best in our challenge because of high variability in human pathologist absolute cellularity scoring of individual patches. A rank-based performance metric assesses how the algorithm orders patches from lowest cellularity to highest cellularity. The human pathologist "truthers" had smaller variability in ranking patches compared with determining absolute percent cellularity scores. By taking this into account, we believe we implemented a more relevant approach for comparing the ability of algorithms to estimate patch cellularity as part of our challenge.

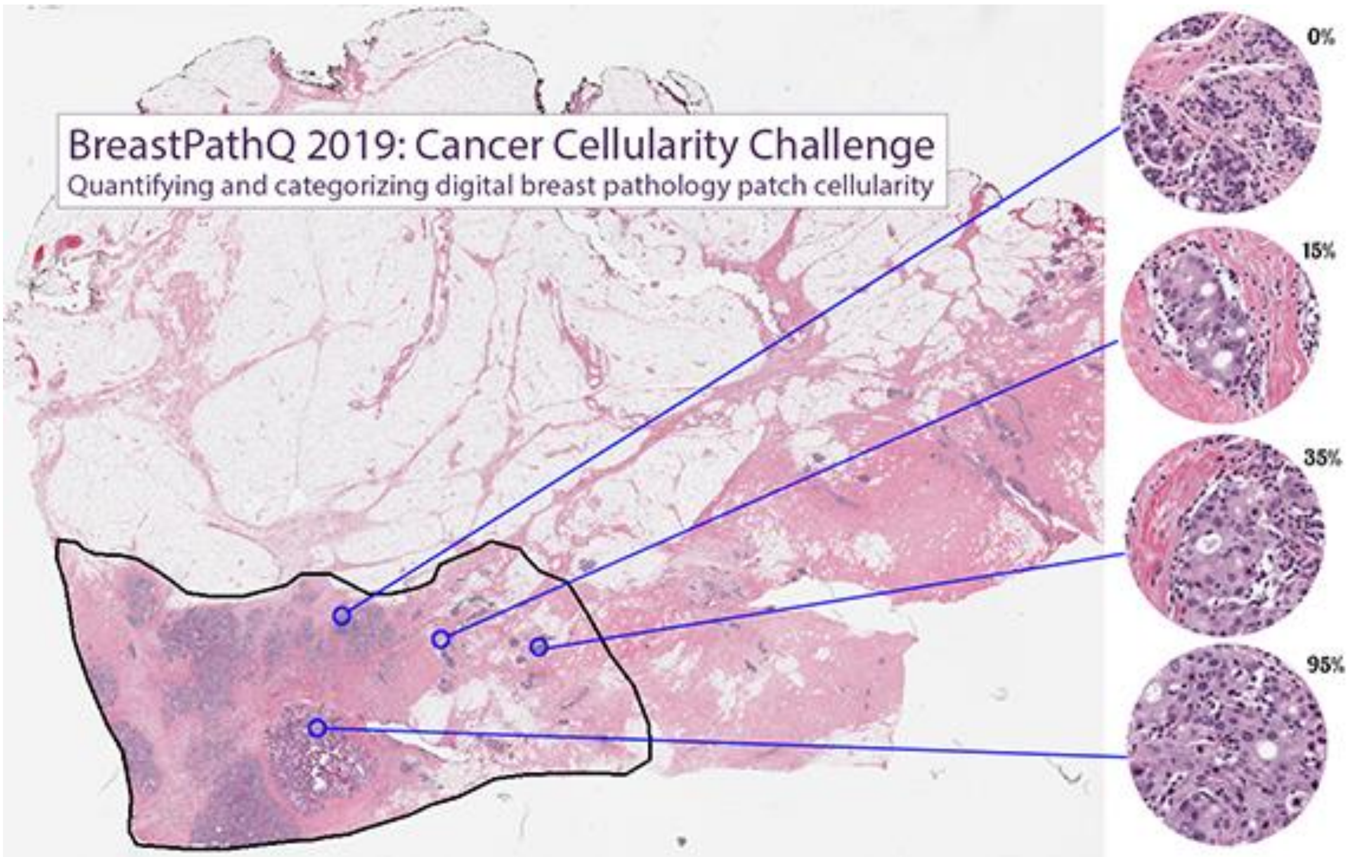


Figure above shows different patches within the pathology slide with various level of cellularity scoring by a pathologist. Courtesy of Anne Martel, Sunnybrook Research Institute, Toronto (and used as part of the BreastPathQ Challenge logo).

• We learned that the submitted algorithms generally performed quite well in assessing cancer cellularity for H&E breast cancer tumor patches with the majority of submitted algorithms having a predication probability value (p_k) greater than 0.90 on a scale of 0.0 to 1.0. This is somewhat comparable to the human pathologist readers who achieved predication probability values of 0.96 and 0.93 on our test dataset. While algorithm performance still needs to be confirmed on a much larger and more diverse patient's dataset to verify these results, the challenge suggests that automated cancer cellularity scoring may be a reasonable approach to consider in order to reduce variability among pathologists and potentially streamlining the assessment of residual cancer burden in breast and other cancers.

• We learned that a well-organized challenge in the medical field requires people of many expertise working together. For this challenge, we had experts in the clinical task (to define the task to study and its clinical relevance), statistical analysis (to determine the evaluation of the results), logistics support (to ensure that the challenge runs smoothly), as well as the support of organizations for this challenge (to advertise the challenge, and have a venue for reporting the results). Without these experts working together, the challenge would not have been successful.

We have collected comments by the two winning teams. **David Chambers**, Senior Research Engineer at the **Southwest Research Institute (SwRI)**:

"The cellularity is a measure of the area of the patch that is occupied by malignant cells, so the problem can be thought of as a segmentation problem underneath. The Southwest Research Institute and Univ. of Texas Health Sciences Center at San Antonio team approached the problem as a weakly-labeled segmentation problem, and iteratively refined our algorithm by providing strong labels for hard examples and retraining. Pathologist expertise played a large role in our success.

Because global context is important in this problem, we used a network with "Squeeze-and-Excitation" architectural units, a recent development in neural network architectures that allows for reweighting features according to global content."

David Chambers

Mamada Naoya

Co-winner **Mamada Naoya**, a master's course student at the **Tokyo Institute of Technology**, concludes:

"I study deep learning applications for material science and I know next to nothing about pathology and cancer.

So my winning shows the versatility of deep learning, I believe.

And I was surprised to hear that top models' performances are as good as expert pathologists'.

With larger amount and more variety of data (annotation by many pathologists, different optical devices, rare clinical cases, etc.), super-human models will be possible."

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

Delivery robots put sidewalks in the spotlight:

How ready robotics technology is to navigate the surprisingly perilous world of the neighborhood sidewalk? Now that **Amazon** has announced a pilot of **autonomous delivery robots**, what are the main technological hurdles that need to be solved? Read More...

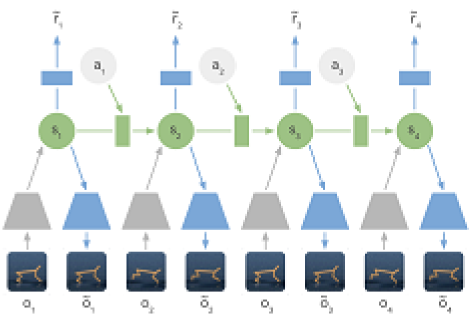
'Fauxtography' is now a fact of life:

Unedited photos are today no more than a romantic ideal. Theverge.com tells us what **image processing edits** are silently done by our smartphones even before we see the pic; after that, **manual photo retouching** has become more a habit than an option. Whitening and flattening the skin, enlarging the eyes, shrinking or straightening the nose... the article is well written and fun to read. Enjoy!

Don't Let Artificial Intelligence Pick Your Employees:

Are algorithms sophisticated enough to make strategic decisions like **hiring**? Even though a **Stanford** organizational behavior teacher thinks they are not, the **use of machines in hiring** has become widespread long ago and now between 65% and 70% of all job applications are touched first by a machine. Read More...

Exploring PlaNet, Google-DeepMind's Solution for...



...for Long Term Planning in Reinforcement Learning Agents. **Google and DeepMind AI researchers** work on a **Deep Planning Network (PlaNet)**, a new model that can learn about the world using images and utilize that knowledge for long-term planning, with two new concepts: a **Recurrent State Space Model** and a **Latent Overshooting Objective**. Read About It Here...

Artificial Intelligence Should Start With Artificial Joints:

Artificial intelligence is converging with medicine and one of the areas involved is joint replacement: large high quality patient data may feed a sophisticated algorithm that predicts the risks of a joint replacement in the future as well as expected costs and recovery info. Read More **and** What RSIP Vision Offers in This Field

What's Behind JPMorgan Chase's Big Bet on Artificial Intelligence?

America's biggest bank bets on **AI** to shape its future strategies. Read More Listen

3D Stereo Vision Calibration

by Aliza Minkov

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 computer vision project by **RSIP Vision**: **3D Stereo Vision Calibration and Triangulation**.

Our client's main goal in this project was to find the **3D location** of certain points of interest. They had automated operations where they need to know the exact location of objects in space. The original system was not accurate enough for the client's ambitions.

For this purpose, we used two cameras to find the exact 3D location out of 2D image points. At the **Proof Of Concept** phase, the client defined some of the camera specifics, like the distance and the height of the objects of interest; the specs were of course expected to be different in the real production run. The two cameras system was needed to solve the depth part, **transforming the 2D image into 3D points**. Before starting the project, we did a simulation to verify whether additional cameras would enhance the performance in that environment: this was not the case, so for simplicity we decided to keep working with two cameras, optimizing the choice of angles, lenses and focus to be used in order to see the object with both cameras.

There were two main challenges to obtaining accurate results in this project, that we solved using computer vision techniques. The first one was the stereo vision part and the triangulation part between the two 2D points: theoretically, this sounds like something that simple linear algebra equations can solve. In practice, these two points need to correspond: it has to be the same point in 3D that you see in both 2D cameras. In real life this is not always the case, due to **distortions of the lenses**, which we solved doing an appropriate **calibration** at the beginning of the project. This is a one-time process that fixes the distortion. We also needed to find the camera geometry, to do a correct triangulation. That means the camera's intrinsic parameters as well as its rotation and translation to get to the coordinate system that we want and the relation between the cameras. Of course, in this process of finding the parameters you also have some errors and you need to be very organized and punctual in all the steps of the calibration, which involves repeating some of the parts to get better results. In this case, we used an object with a calibration pattern in the form of a known geometry: a chessboard was chosen in this and in many other cases since it has a grid with boxes of the same size and a repeated pattern, with straight lines, so that you can fix the distortion (by finding the distortion coefficients, so that all the corners appearing on straight lines in reality, will appear on straight lines on the image as well). By taking many images of the chessboard in different locations and using optimization algorithms, we were able to fine-tune the parameters and find the best ones to satisfy these conditions, solving the calibration problem. The need to cover most of the field of view with a single chessboard was challenging, since the field of view was pretty large and the chessboard limited in size. We solved this challenge by moving around and taking many images; in this way, we gave many 2D 3D correspondences to the algorithm, as if you were taking images of a big chessboard. It's actually good to use more than one image in any case for the optimization algorithm, even if most of the field of view is covered by one image. We even had to replace the chessboard since the first one was printed over a slightly bending material, bending the straight lines with it, which contributes to the error. In addition, for the extrinsic calibration part (the calibration between the two cameras), you want the camera to see exactly the same 3D points to correspond and then in some of the images we saw that there was a slight movement due to humans holding physically the chessboard in their hands. This slight difference can have a huge influence on the error, because it is actually not the same point that you are showing. The conditions for stereo calibration require that you use a grid with straight lines, with no movement between the two pictures. This solution per se is not very difficult: it requires organization and precision in order to make it work well at the first time, localizing **exact 3D point images** to give as input to the **navigation system**.

The second challenge consisted in automatically extracting the same feature points from both images. To that purpose, we used deep learning neural network, which we trained to accurately localize these points. Of course, this requires a lot of data, which must be collected without delaying the Proof Of Concept's schedule. Trying to triangulate image points that do not correspond exactly will cause an error. That's why a great accuracy is required, even after the training of the network, hence some algorithms were used to enhance this accuracy of points correspondence. For instance, if the point of interest is on some plane seen by both cameras, you may want to get the **homography matrix** – a matrix that relates the transformation between two planes. Using the homography matrix, you can transform one image point to its exact corresponding point in the second image. How do you get the Homography? Again, there are several traditional computer vision algorithms. One way is using **SIFT** – matching similar feature points between the two images, after the neural network has identified the area of interest. It must be noted that SIFT might not work very well when the orientation of the cameras is too different; in addition, it may fail for pattern-like repeating structures, because the algorithm can mistakenly match similar wrong feature points. In that case another approach can be taken, which is called a **dense stereo matching**. In this method we find the depth map of our area of interest. When you have a depth map, you can better match each point with its correspondent point on the second image. With enough corresponding points, you can use algorithms like **RANSAC** to get the best fitting plane equation for the inliers point correspondences, and the plane equation will give you the homography.

There are many challenges in computer vision projects like this and it is key that you work with experts like RSIP Vision's engineers to solve them in the optimal way and reach your goals.

Project

by Assaf Spanier

**What happened?
Is this a legitimate transaction, an unusual purchase, a change in behaviour?
Or something is wrong...
What is the reason for this outlier?**

Outlier detection, also known as anomaly detection, refers to identifying rare occurrences, observations, or in the most general sense -- data points -- that show a distinct variance from the general population. Ground truth labeling is scarce, especially for outlier detection tasks. Outlier detection is especially valuable in fields processing big data. Industry applications include financial fraud detection (Ahmed, et al), mechanical failure detection (Shin, et al), network infiltration (Garcia-Teodoro, et al) and pathology detection in medical imaging (Baur, et al).

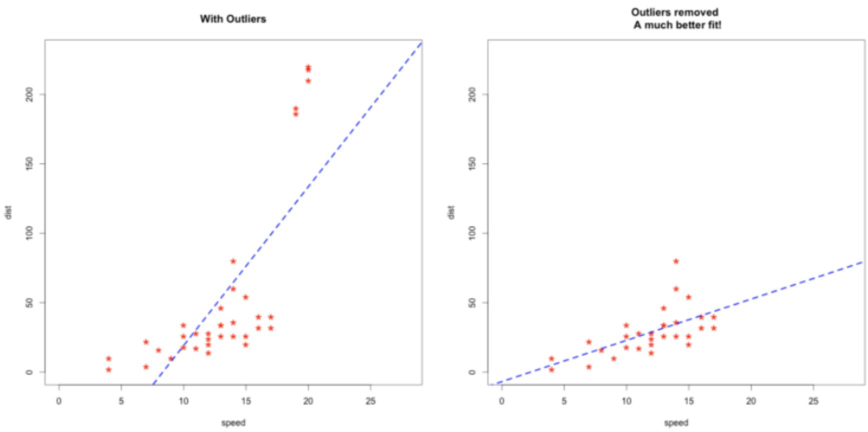
An outlier is any datapoint that is very different from the rest of the observations in a set of data. Some examples: When an 8th-grade class includes one student who is 1.85 m, when all the other students are between 1.55 m and 1.70 m. When a client's purchasing patterns are analyzed and it turns out that while most of his purchases are under $100, a single purchase of over $20,000 shows up out of nowhere. What happened? Is this a legitimate transaction, an unusual purchase, a change in behavior -- or is there anything wrong? What is the reason for this outlier?

There are many reasons for the occurrence of outliers. Perhaps there was an error in data entry, or a measuring error, incorrect data can even be given purposefully -- individuals who don't want to reveal the real data about themselves, may feed made up data as input (for instance into online forms). Of course, we must remember, outliers may also represent real unusual occurrences.

Why do we need to discover outliers?

Outliers can dramatically affect the results of our analysis and our statistical models. Let's look at the following example to realize what happens to our model when outliers are included in our dataset versus when they have been discovered and removed from it.

Let's assume our model is linear regression, and we are trying to fit a line to a given dataset. Look at the difference between the line on the left (fitted to the dataset including the outliers) and the data on the right (fitted after the outliers were removed).



PyOD provides you with a wide variety of outlier detection algorithms, including established outlier ensembles and more modern neural-network-based approaches. All available through a single, well documented, API made with both industry users and researchers in mind. PyOD was implemented with emphasis on unit testing, continuous integration, code coverage, maintainability checks, interactive examples and parallelization.

Here is the link to download PyOD (it is compatible with both Python 2 and 3)

PyOD offers a number of advantages over previously available libraries:

1. PyOD includes over 20 algorithms, covering both classic techniques such as Local Outlier Factor and the latest neural network architectures like autoencoders or adversarial models.

2. Implements a number of methods for merging / combining the results of numerous outlier detectors.

3. Uses a unified API, includes detailed documentation with interactive examples of every method and algorithm for clarity and ease of use.

4. All functions and methods include code testing and continuous integration, parallel processing and just-in-time (JIT) compilation.

5. Can be run on both Python 2 and Python 3, and on all the major operating systems (Linux, Windows, and MacOS).

6. PyOD includes the following popular detection algorithms:

Method	Category	JIT Enabled	Multi-core
LOF (Breunig et al., 2000)	Proximity	No	Yes
kNN (Ramaswamy et al., 2000)	Proximity	No	Yes
AvgkNN (Angiulli and Pizzuti, 2002)	Proximity	No	Yes
CBLOF (He et al., 2003)	Proximity	Yes	No
OCSVM (Ma and Perkins, 2003)	Linear Model	No	No
LOCI (Papadimitriou et al., 2003)	Proximity	Yes	No
PCA (Shyu et al., 2003)	Linear Model	No	No
MCD (Hardin and Rocke, 2004)	Linear Model	No	No
Feature Bagging (Lazarevic and Kumar, 2005)	Ensembling	No	Yes
ABOD (Kriegel et al., 2008)	Proximity	Yes	No
Isolation Forest (Liu et al., 2008)	Ensembling	No	Yes
HBOS (Goldstein and Dengel, 2012)	Proximity	Yes	No
SOS (Janssens et al., 2012)	Proximity	Yes	No
AutoEncoder (Sakurada and Yairi, 2014)	Neural Net	Yes	No
AOM (Aggarwal and Sathe, 2015)	Ensembling	No	No
MOA (Aggarwal and Sathe, 2015)	Ensembling	No	No
SO-GAAL (Liu et al., 2018)	Neural Net	No	No
MO-GAAL (Liu et al., 2018)	Neural Net	No	No
XGBOD (Zhao and Hryniewicki, 2018b)	Ensembling	No	Yes
LSCP (Zhao et al., 2019)	Ensembling	No	No

Let's review some of the algorithms included:

Isolation Forest - A set of trees is used to partition the data and outliers are determined by looking at the partitioning and seeing how isolated a leaf is in the overall structure. Isolation Forest handles multidimensional data well.

Histogram-based Outlier Detection - This is an effective system for handling unsupervised data, it assumes feature independence. The metric used for outlier detection is the construction of histograms and measuring distance from the histogram. It's much faster than multivariate approaches, but at the cost of lower accuracy.

Angle-Based Outlier Detection (ABOD) - The method measures the distance of every data point from its neighbors, taking into account the distance between those neighbors -- the variance of the cosine scores is the metric used for outlier detection. ABOD handles multidimensional data well. PyOD includes 2 versions: 1) Fast -- only using the k nearest neighbors, and 2) Taking into account all data points.

k Nearest Neighbors Detector - For each data point, the distances from its k nearest neighbors are looked at for outlier detection. PyOD supports 3 versions of kNN: 1) using the distance from the k-th nearest neighbor as the metric for outlier detection, 2) using the average of the k nearest neighbor distances as the metric, and 3) using the median of the k nearest neighbor distances as the metric.

Local Correlation Integral (LOCI) - LOCI is very effective at detecting both individual outliers and clusters of outliers. The method produces a LOCI plot for every data point, which summarizes the information about the data points in the area surrounding that point. It determines clusters, micro-clusters, the diameter of clusters and distances between clusters. And from these measurements determines the degree of anomaly of the data point.

Feature Bagging - Feature bagging fuses a number of base classifiers to improve prediction accuracy, fusion can use simple methods like averaging or median, or more sophisticated ones. Local Outlier Factor is used as the default Outlier Detector, but any other outlier detection algorithm, such as kNN or ABOD, may be substituted.

As stated, PyOD can be run on either Python 2 or 3, using package six. It's based on NumPy, SciPy and SciKit-Learn, and uses Keras for advanced neural network methods such as autoencoders and SO_GAAL. To improve scalability, all algorithms were optimized with JIT and Numba, and the library supports parallel processing on multi-processor computers.

The PyOD API, inspired by the SciKit-Learn API, is basically a replica of that well-known interface. In particular, it includes the following:

1) the fit function, which trains the model and collects the appropriate statistics.
2) decision_function, which ranks outliers for every new data-point, once the model is trained.
3) predict, which returns a binary label for each data point.
4) predict_prob, which delivers the result as a probability measure.
5) fit_predict, which corresponds to calling the predict function, after performing fit.

The package is available as a Python open source toolbox and it can be easily extended to implement new methods. New models are very easy to train within this framework (using the unified and well-known API), and you can use polymorphism to easily inherit any function and implement it for your own needs.

How to install the package - as simple as:

```
pip install pyod
```

Example:

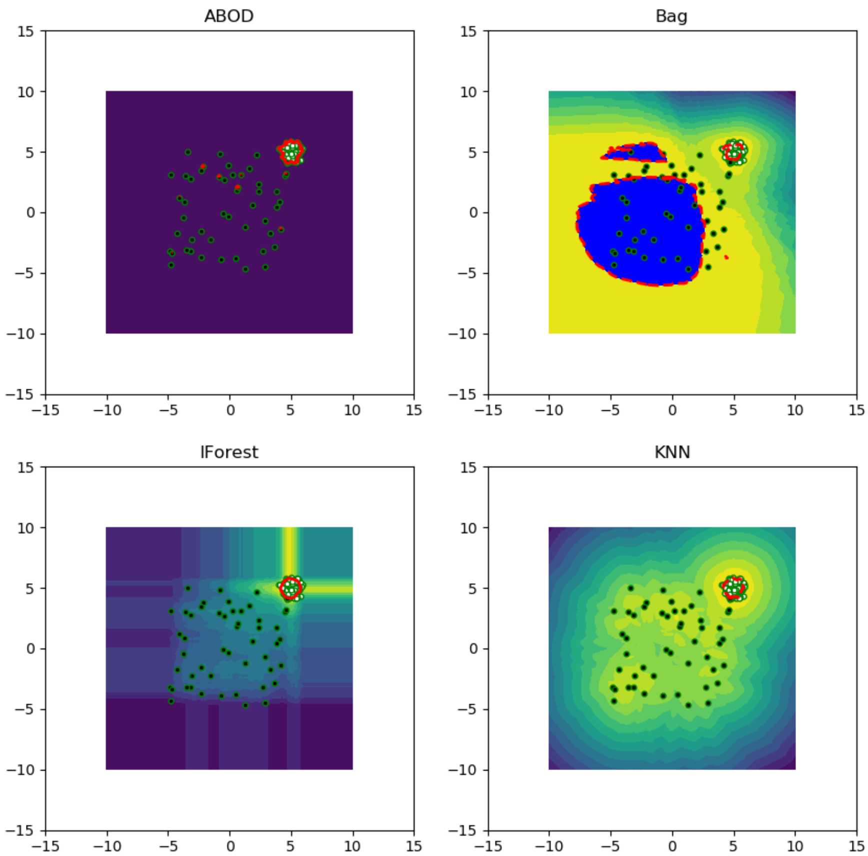
Testing 4 methods on random generated dataset

```
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt

from pyod.models.abod import ABOD
from pyod.models.knn import KNN
from pyod.models.loci import LOCI
from pyod.models.iforest  import IForest
from pyod.models.feature_bagging import FeatureBagging

from pyod.utils.data import generate_data, get_outliers_inliers

# random data
X_train, Y_train = generate_data(n_train=500,train_only=True, n_features=2)

# store outliers and inliers in different numpy arrays
n_outliers, n_inliers =
list(map(len,get_outliers_inliers(X_train,Y_train)))
x1_outliers, x1_inliers = get_outliers_inliers(X_train,Y_train)
xx , yy = np.meshgrid(np.linspace(-10, 10, 200), np.linspace(-10, 10, 200))
plt.figure(figsize=(10, 10))

# Test 4 different methods
contamination_ = 0.2
classifiers = {
     'ABOD'   : ABOD(contamination=contamination_),
     'KNN'    : KNN(contamination=contamination_),
     'Bag'  : FeatureBagging(contamination=contamination_),
     'IForest'    : IForest(contamination=contamination_)
}

for i, (clf_name,clf) in enumerate(classifiers.items()) :
    clf.fit(X_train)

    # predict outlier score
    scores_pred = clf.decision_function(X_train)*-1

    # predict outlier
    y_pred = clf.predict(X_train)
    n_errors = (y_pred != Y_train).sum()

    # threshold value outlier
    threshold = stats.scoreatpercentile(scores_pred,75 *contamination_)

 # calculates the raw anomaly
    t = clf.decision_function(np.c_[xx.ravel(), yy.ravel()]) * -1
    ZZ = t.reshape(xx.shape)
```

```
    # plot outliers and contour
    subplot = plt.subplot(2, 2, i + 1)
    subplot.contourf(xx, yy, ZZ, levels = np.linspace(ZZ.min(), threshold, 15))
    subplot.contour(xx, yy, ZZ, levels=[threshold],linewidths=2, colors='red')

    # fill orange contour lines where range of anomaly score is from 
threshold to maximum anomaly score
    subplot.contourf(xx, yy, ZZ, levels=[threshold, ZZ.max()],colors='blue')

    # scatter plot of inliers with white dots
    subplot.scatter(X_train[:-n_outliers, 0], X_train[:-n_outliers, 1], 
c='white',s=12, edgecolor='g')
    # scatter plot of outliers with black dots
    subplot.scatter(X_train[-n_outliers:, 0], X_train[-n_outliers:, 1], 
c='black',s=12, edgecolor='g')
    subplot.axis('tight')

    subplot.set_title(clf_name)
    subplot.set_xlim((-15, 15))
    subplot.set_ylim((-15, 15))
plt.show()
```

Where the output is:

We clearly see that ABOD, Isolation Forest and KNN found the main groups (red contour) and left the outlier out, while the BAG failed to find the outliers.



Chip Huyen

Chip Huyen works as a deep learning engineer with the Artificial Intelligence Applications team at NVIDIA, where she develops new tools to make it easier for companies to bring the latest deep learning research into production. Our attentive readers might remember her as the author of SOTAWHAT. She has recently released lazynlp, a library that allows easy scraping, cleaning, and de-duplicating webpages to create massive monolingual text datasets. Originally from Vietnam, Chip graduated from Stanford University with a BS and MS in Computer Science. More interviews with women scientists

Chip, I think that you have many things going on in your life already even at such a young age. You grew up in Vietnam, traveled all over the world for three years, and studied at Stanford in California. Which one do you want to talk about first?

I don't know! I think all of them are in the past.

So tell me about your current work. What is it?

I'm working with language models mostly. My team is building a toolkit to help companies bring AI research into

"Physical distance is not that big of a deal!"

"… the language we speak helps form who we are!"

production). My team focuses more on speech recognition, but I am focused more on the research side: the language modeling.

How did you get there?

I have always been interested in languages. My background is in writing, so I find languages fascinating. And also, in my time traveling, I realized that language is a big part of any culture, and the language we speak help form who we are. When I came to Stanford, I got super interested in computer science, programming, and NLP. I could combine my love for languages and computer science.

Once Amy Bearman told me about the duck syndrome at Stanford, where students are like ducks: on the surface very placid and self-confident, but under the surface they are paddling like mad. Do you know about it?

Yeah, it's pretty common to hear about it.

Is it particular for Stanford?

I think there's this certain perception of California that makes it worse. I feel like people here are very outgoing, healthy, and have a very positive look on life, which is great, but it also creates a pressure to do well and be happy all of the time. I live in this amazing place with amazing weather, like 300 days of sunshine. When somebody asks me, "*How are you doing?*", the only correct answer is to say, "*Oh, I'm doing great!*" because, how can you not? So I just feel like if I'm not happy, what's wrong with me? Why am I not happy in such a perfect place? And you have to pretend to be happy, and I think that leads to the duck syndrome.

So tell me something about California. What do you like there which is better than Vietnam?

Woah [laughs]. I think the weather is nice. Have you been to Vietnam?

No, but I'm going there in a few months. You will have to give me tips.

Well, it is very humid.

I expect to be there in October.

You are likely going to sweat a lot. In California, it's never really too hot or really too cold. Even in the winter, you can still have sun, go out, and play sports. The air quality is really nice, and there's a blue sky. When I went back to Vietnam the last time the sky seemed to be foggy because of the pollution.

Tell me something nice about Vietnam that we don't know.

There are definitely more human interactions, personal interactions. I think it's great that everyone here is really motivated, and I am also motivated talking to them. I also feel like I have to live up to all of these

expectations, and sometimes I feel like people judge me for my potential. They become my friends because they think I'm going to do great in the future or do something for them. It's not because of my personality. So there are some things that are a bit transactional about human interactions instead of companions. I miss that a lot.

A friend of mine told me a few months ago that Vietnam is heaven on earth because of the kindness of people. Would you agree with that?

I think that each place has its own characteristics, and some are pros. Some are cons. The thing is there are people living in both places. That means that no place can be that bad if people still live there. I'm actually writing a book on Vietnam. I'm trying to figure out what makes Vietnam, Vietnam.

That's very nice. This one is in English?

I hope so. Yes, it is in English

Tell me about AI in Vietnam. I know that you are working on something.

I'm a member of this organization called VietAI. It's a nonprofit organization that hopes to help create a new generation of AI engineers and researchers in Vietnam. We have courses that only cost 250 dollars for ten weeks of learning. We charge enough for operations, and we have events. Last December we had this AI Summit. I think it was the very first AI-centric conference in Vietnam.

You do a lot of stuff! I tried to see how many things you do and I lost count. You have your blog. You have VietAI. You have the new Stanford in Real Life posts. You have your Learn365Project…

The 365Project shut down to make room for Stanford in Real Life.

What happened to it?

Nothing, I'm just changing. I've been working on the Learn365Project for a while, and it's just like something to scribble on every day. I tried, but I didn't do it every day. Sometimes I would do it and not publish it. Stanford IRL is more focused. It's solely about people: what they are doing, the struggles that they are facing after college, and the career paths they are choosing. I'm working a lot on my full-time job at NVIDIA. I really like my job. I'm working on my book, and the Stanford in Real Life series.

You do a lot of things! What drives you?

What drives you, Ralph?

I try to live according to my values. I try to make sure that I do things so I will be happy with myself a few years ahead of now. My formula is a mix of three things. One is that I do things that I like. The second is to do things that I think are useful. The third is to do things for which I don't have to compromise my standards, my inner values.

What are your inner values?

“When you contribute something, you value yourself and feel useful.”

Mainly truth, honesty, kindness... this kind of stuff. The most important one is truth. What about you?

I think for me it’s just a desire to contribute, to create content.

For other people to benefit from?

Yes, some sort of value in myself. When you contribute something, you value yourself and feel useful. You’re making a difference, even if it is small. I write a lot in Vietnamese. I contribute to a couple of columns on Vietnamese newspapers. I feel super, super privileged. I had a chance to travel and see different countries first hand. There are a lot of things that I learned that I would like to share with other people in Vietnam, so I write about it.

What should you get out of it? It sounds perfect.

I just really feel like maybe I’m just obsessed with this desire to share!

That includes the free hugs campaign…

Yeah!

Tell our readers.

It started in high school. I lived away from home. It was a great chance for me to experience life. I don’t think many people have the chance to stand on their feet at 15. It gave me a taste of adulthood. But sometimes, when I got sick, it got really lonely. I realized, in Vietnam, parents don’t really tell you that they love you. There’s no hugging. It’s a different way to express feelings and love to each other, but sometimes I feel like if somebody hugged me, I would feel so much better, but I didn’t know how. Then I went online, and I saw this free hugs campaign. I think it was by an Australian guy. I was like, “*Wow! This is great!*” Strangers can also hug each other! A hug is not just a hug. It’s a symbol of connection, of both emotional and physical connection. So I wrote this very emotional blog post about how everyone should hug each other. It got very popular. We decided to do a campaign in Vietnam. It was one of the biggest youth-organized events in the country. A thousand young people went in the street all across the country, and we just went around and hugged people. I remember there was this lady who was selling things on the street. She was really old. We went to hug her, and she just started crying. She had never felt that before. It made me feel really sad, but also happy at the same time. It’s weird. It was very emotional. After that, we already had a

“… maybe I’m just obsessed with this desire to share!”

“We’re having less and less unknowns every day, but how far are we from being completely unignorant?”

huge network of volunteers. We started having this organization. We called it Free Hugs Vietnam. At first, we tried to help foreigners like foreign NGOs. When they come to Vietnam, we help them liaison with the local organizations. We give them volunteers. We have a lot of events for young people like English-speaking clubs and planned competitions. We raised money. We bought meals. We went around and gave meals to homeless people during the holidays.

How nice!

It’s really nice. I’m really happy. I’m not actually involved with it anymore. There are other young people in charge.

You have traveled a lot and met a lot of people, so you have plenty of friends almost everywhere. And of course, your family is very far away. By being in different places all of the time, you certainly miss people. How do you handle this?

I feel like the fact that I was able to travel makes me realize that physical distance is not that big of a deal. For me, when I don’t see my friends regularly, I know that if we see each other again, we’re going to be friends again. I’m not traveling that much anymore. I’m mostly in California now.

Are you planning a PhD?

I’m not doing a PhD.

Can you tell me why?

I think PhDs are great. I just don’t think that it’s for me.

That’s fair enough. Tell me something about NLP. What do you like about it?

NLP still has a lot of challenges. I feel like the way that we approach NLP nowadays is not going to solve everything. I see a lot of big, important papers coming out, but I don’t see a fundamental approach to it yet. Actually, I’m very interested in transfer learning for NLP.

I’m also very interested in something that I see as a curriculum learning. I feel like children who learn languages have a very small vocabulary and they gradually build it up over time. I’m trying to see whether it’s possible with a model. You start with a dataset of simplified English, and we learn more complex English. So I created a set of children’s books with very simple English, and I’m not sure it’s going to work. It’s an open research problem.

Do you believe that human knowledge is going upwards?

We’re having less and less unknowns every day, but how far are we from being completely unignorant? I don’t know!

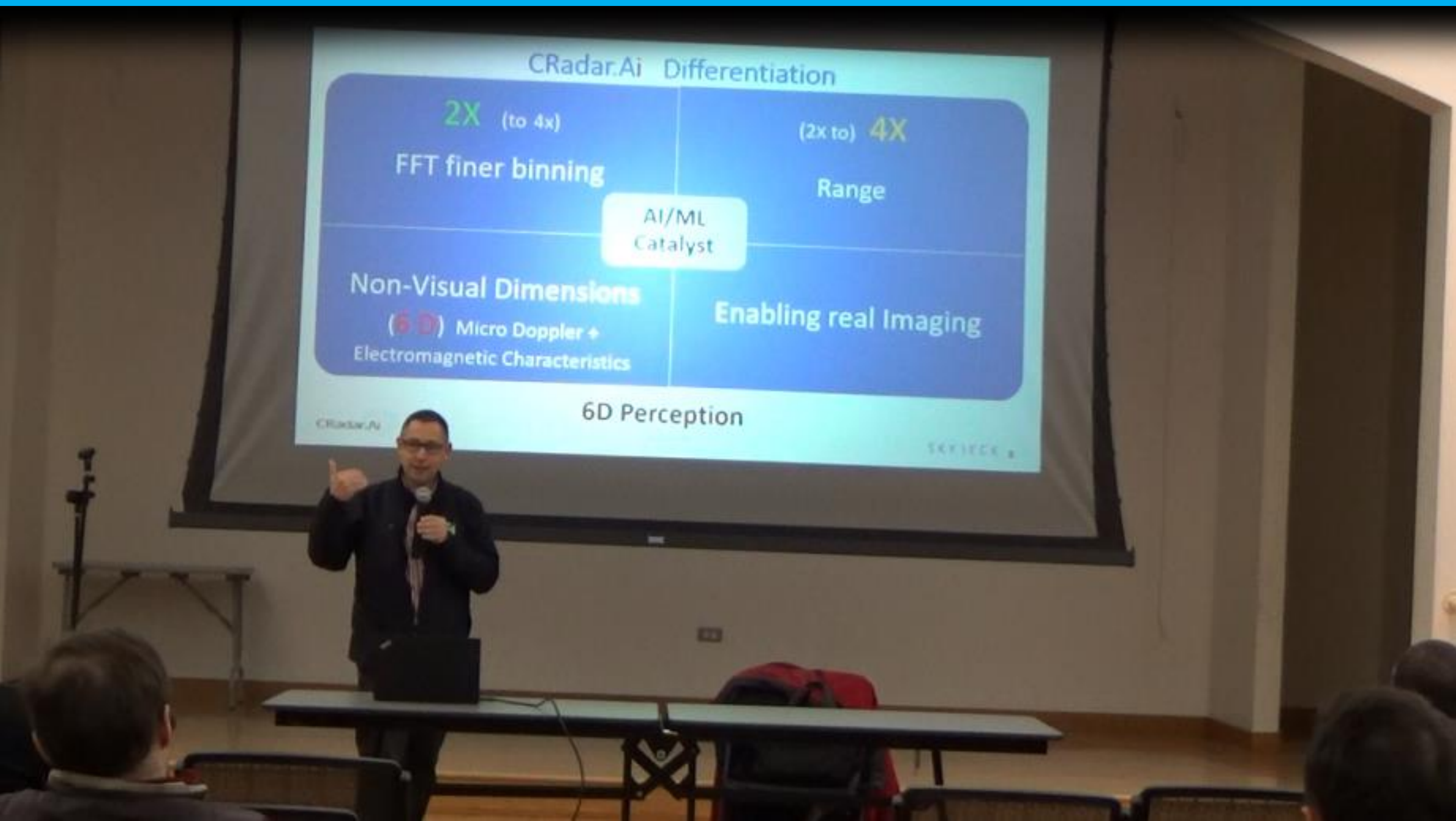
Read more interviews like this

"I had a chance to travel and see different countries first hand!"




Among many other initiatives, RSIP Vision sponsors the Bay Vision Meetup. We invite great speakers in the Silicon Valley. The February 27 meeting held in Cupertino focussed on the topic of Developments in Artificial Intelligence for Autonomous Vehicles. In these two images, Shmulik Shpiro (RSIP Vision) says a few words of welcome and introduces the two speakers.

On top, the first speaker: Mohammad Musa, Founder and CEO at [Deepen.ai](), speaking about "3D Point level segmentation". Below, the second speaker: Eran Dor, Co-Founder & CEO at CRadar.Ai, discussing how to utilize signal to improve and change the way Radar sensors detect, classify and perceive the car's surroundings. Thank you both for your great talks! The full video is [here]()!



RSIP VISION

Global Leader in Computer Vision & Deep Learning

Women in Computer Vision

by Ralph Anzarouth

Women in Computer Vision (also called Women in Science) is a series of interviews conducted by Ralph Anzarouth. New interviews are regularly published on all RSIP Vision's publications: Computer Vision News and the Daily magazines (CVPR Daily, MICCAI Daily and many more).

Find now on the project page the direct links to almost 100 interviews... **at the click of a button**

Leadership

Mentoring

Competence

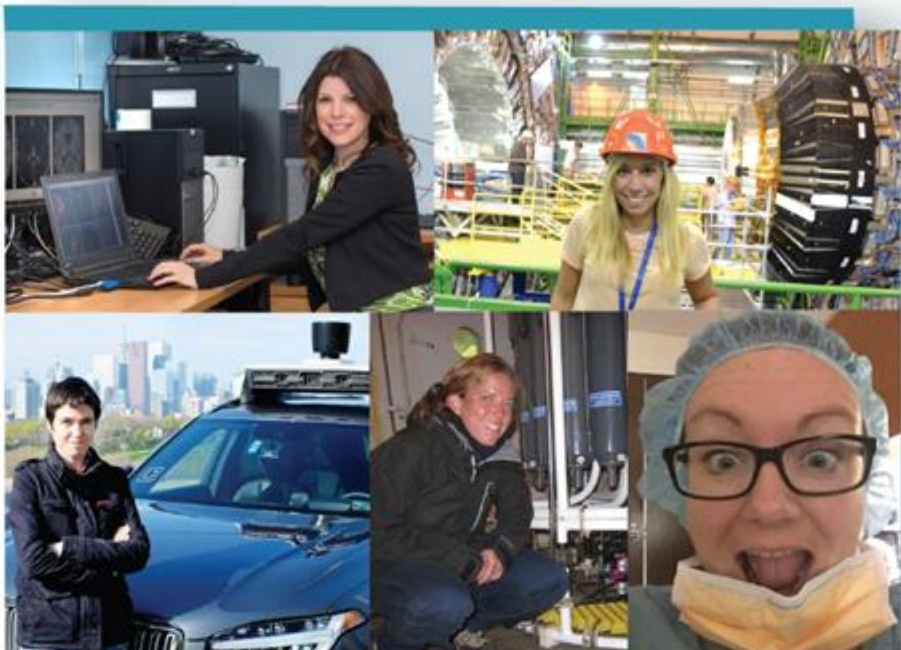
Confidence

Community

"The only way to succeed is to really start believing in yourself!"
Michela Paganini

"Most of all, you have to believe that you can do it!"
Laura Leal-Taixé

"It may look like a long list of names, but behind each name there is a fascinating world in which we were let in."
Ralph Anzarouth



Did you miss an interview? No worries, you can find them all in the **Women Scientist section of RSIP Vision's website**

WELCOME TO Fabulous COMPUTER VISION EVENTS

TRI-CON - International Molecular Medicine Tri-Conference
S.Francisco, CA — Mar 10-15 — Website and Registration — MEET US!

AAOS Annual Meeting American Academy of Orthopaedic Surgeons
Detroit, MI — Mar 12-16 — Website and Registration — MEET US!

ACC - American College of Cardiology's 68th Annual Session
New Orleans, LA — Mar 16-18 — Website and Registration — MEET US!

ADAS Sensors 2019
Detroit, MI — Mar 20-21 — Website and Registration — MEET US!

IAPR Computational Color Imaging Workshop
Chiba, Japan — Mar 27-29 — 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

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.

Did you read the Feedback of the Month? It's on page 29!

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!