

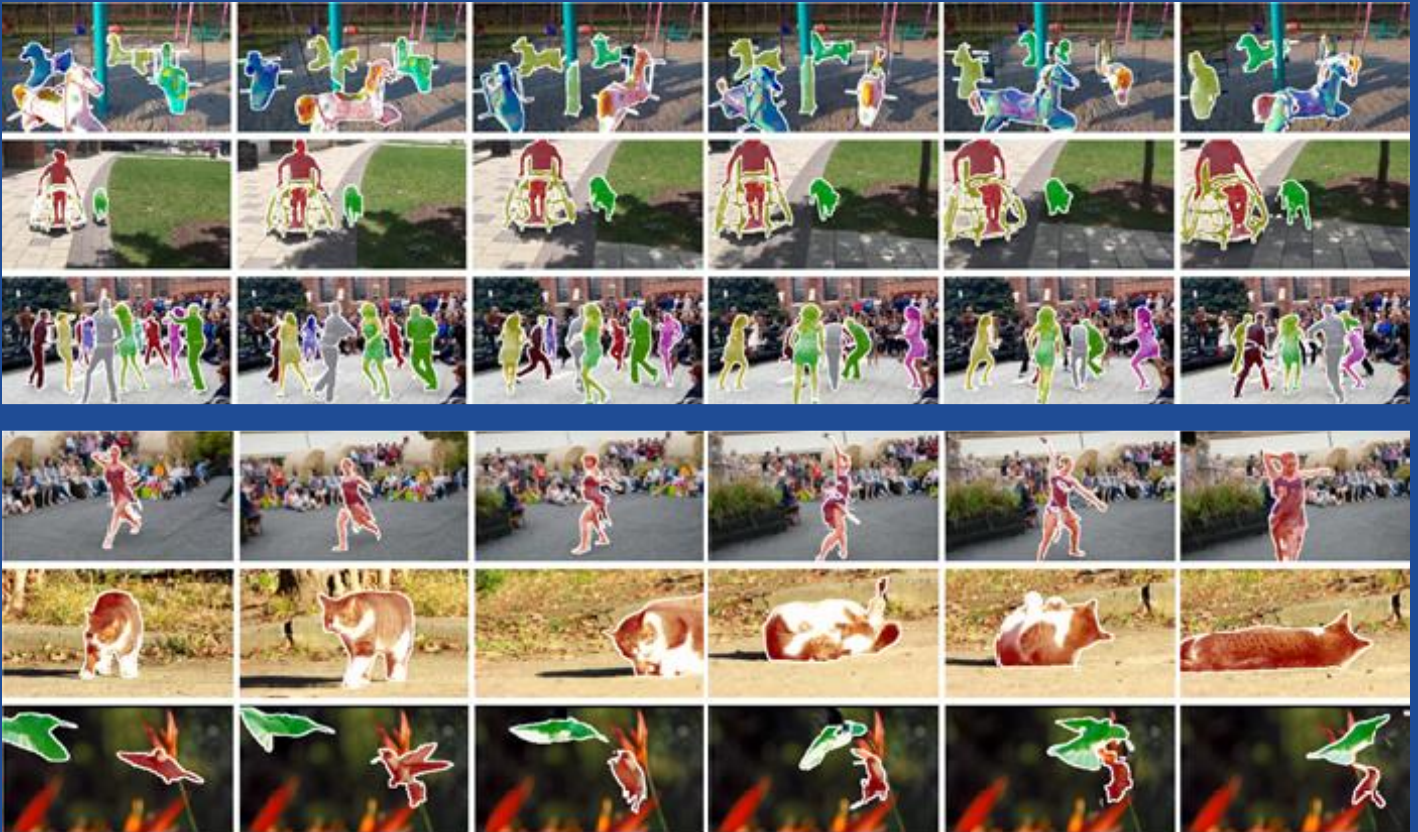
Computer Vision News

The magazine of the algorithm community

A publication by



December 2018



Research:

CNN in MRF: Video Object Segmentation via Inference in a CNN-Based Higher-Order Spatio-Temporal MRF

Spotlight News

**Women in Computer Vision:
Lydia Chilton**

**Focus on:
BOHB**

Challenge:

CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation

Upcoming Events

Application:

Inky - Anti-phishing Protection

Project Management:

Collecting Data for Medical Projects

Computer Vision Project:

Left Atria Reconstruction with Deep Learning



04

Research Paper Review CNN in MRF

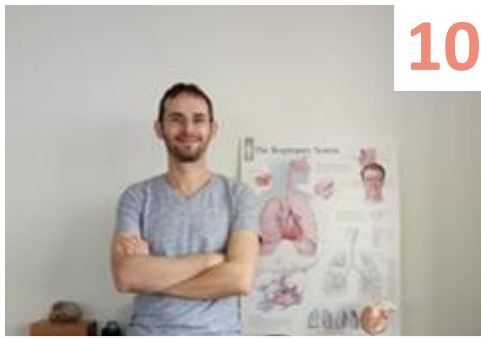


15

Spotlight News

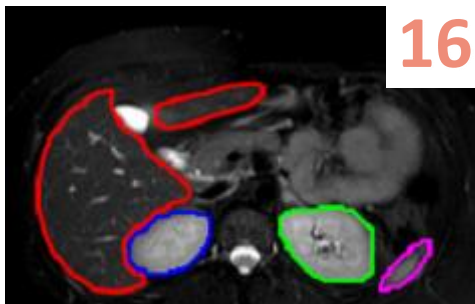
Parameter Name	Parameter type	
Learning rate	float	24
Optimizer	categorical	{adam, s
Momentum	float	[0, 0.99]
Number of conv layers	integer	[1,3]
Number of filters in the first conv layer	integer	[4, 64]
Number of filters in the second conv layer	integer	[4, 64]
Number of filters in the third conv layer	integer	[4, 64]
Dropout rate	float	[0, 0.9]
Number of hidden units in fully connected layer	integer	[8,256]

Focus on: BOHB hyperparameter optimization



10

Project Management Tip Deep Learning in Cardiology



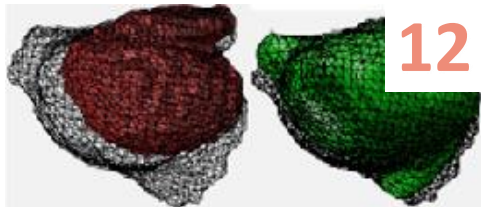
16

Medical Challenge CHAOS



32

Women in Computer Vision Lydia Chilton



12

Project by RSIP Vision Deep Learning in Cardiology



18

Application: Inky Anti-phishing Protection



37

Upcoming Events

03 Editorial by Ralph Anzarouth

04 Research Paper Review
CNN in MRF *by Assaf Spanier*

10 Project Management in Comp. Vis.
Collecting and Selecting Data

12 Project by RSIP Vision
Left Atria Reconstruction with Deep L.

15 Spotlight News
From Elsewhere on the Web

16 Medical Challenge Review
CHAOS *by M. Alper Selver*

18 Application - INKY
Anti-phishing Protection w. comp. vision

24 Focus on:
BOHB *by Assaf Spanier*

32 Women in Computer Vision
Lydia Chilton - Columbia University

37 Computer Vision Events
Upcoming events Dec2018 - Feb2019

38 Subscribe for Free



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](#)

Copyright: **RSIP Vision**

All rights reserved

Unauthorized reproduction
is strictly forbidden.

Follow us:



Dear reader,

As **Computer Vision News** gets closer to its **3rd birthday**, most of our readers notice how much energy and enthusiasm is devoted to bring about its publication every month. It is probably judicious to remind our readers that this magazine is a gift offered by **RSIP Vision** to the **Artificial Intelligence community**.

RSIP Vision is a pioneering software house offering **R&D outsourcing and consulting services** to the industry. Enterprises love to work with us and starting this month we will display their feedback.

The first feedback is on page 30: it's the very kind **message of appreciation** that Christiane Michaelsen recently sent us. Christiane, we also loved working with you!

Enjoy reading this new issue of Computer Vision News and, as always, **take us along for your next Deep Learning project!**

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

Did you read
our exclusive
interview
with
Yann LeCun?



by Assaf Spanier



Every month, Computer Vision News reviews a research paper from our field. This month we have chosen **CNN in MRF: Video Object Segmentation via Inference in a CNN-Based Higher-Order Spatio-Temporal MRF**. We are indebted to the authors (**Linchao Bao**, **Baoyuan Wu** and **Wei Liu**), for allowing us to use their images to illustrate our review. Their article is [here](#).

An attempt to mine the higher-order potentials of combining MRF/CRF with CNNs, by embedding a feed-forward pass of a CNN inside the inference of an MRF model

Background, motivation and novelty:

1. Aim & Motivation / Challenge:

This paper deals with segmentation of video, where the input includes (in addition to the sequence of images) a preliminary object mask -- this is the object whose segmentation we will want to refine. The authors propose a **spatio-temporal algorithm**, which uses a **Markov Random Field (MRF)** to the pixel-space of each-image (the spatial part) and between pixels in consecutive frames of the video (the temporal part). Its innovation on regular MRF models is the modeling of the spatial dependence between pixels for MRF using **convolutional neural networks (CNN)**, the temporal dependence between pixels is modelled by optical flow. The resulting MRF model combines spatial and temporal clues for video object segmentation. Performing inference directly from this MRF model is a very difficult computation problem due to the model's high interdependence between pixels. Therefore, the authors propose performing **approximate inference using embedded CNN** -- this algorithm alternates between a temporal fusion step and a feed-forward CNN step.

Initialized with one-shot video object segmentation CNN, the proposed algorithm achieved the top performance in the [DAVIS 2017](#) public benchmark.

2. The main ideas:

An innovative spatio-temporal MRF model for video object segmentation. The authors' algorithm performs approximate inference from the MRF model by alternating between a temporal fusion operation and a mask refinement feed-forward CNN, incrementally inferring video object segmentation.

3. The main contributions are:

(a) The Authors' proposed spatio-temporal MRF model for video object segmentation encodes spatial dependency by CNNs trained for objects of

interest, so higher-order dependencies among pixels can be modeled to enforce the holistic segmentation of object instances.

(b) The iterative algorithm performs video object segmentation efficiently. The algorithm alternates between a temporal step (between video frames) and a spatial step using a CNN along the image-space to refine segmentation results.

4. Background:

The idea of applying CNNs in combination with MRF/CRF models is not new: DeepLab semantic segmentation framework attempted to improve the semantic labelling results produced by CNN, by using fully-connected CRF post-processing. The Jang and Kim video object segmentation method combined the outputs of a triple-branch CNN using MRF optimization. These loosely-combined algorithms did not, however, take full advantage of the MRF/CRF models' strengths. **Schwing and Urtasun** jointly trained CNN and MRF by back-propagating gradient obtained during the MRF inference to CNN, however, no distinct improvement over separate training was achieved. The CRF-RNN model used a mean-field to approximate the CRF inference, within an RNN, arriving at an end-to-end trainable deep network, which greatly boosted performance. **Deep Parsing Network (DPN)** is an attempt to use a mean-field to approximate MRF inference in one training pass. The authors' paper is trying to model higher-order potentials in MRFs with CNNs.

What is **CRF**? What is **MRF**? And what are the differences between them?

- A CRF can be thought of as an extension of logistic regression, CRF models conditional probability $P(Y|X)$. For example the code for implementing SVM as a special case of CRF, can be found [at this link](#):
- MRF models the joint probability of both Y and X together. It models $P(Y,X)$, and also can be used to compute $P(Y|X=x)$ for a given input x.

The advantage of CRFs is their focus on the "standard" inference problem $P(Y|X=x)$, making them often more precise. On the other hand, that is the only problem they are capable of solving. The advantage of MRFs is that they are completely general, and thus able to model arbitrary inference problems. For instance, let's assume for some reason some of input x is missing, an MRF can fill-in the missing values because it produces the entire probability distribution.

What is **Mean-Field Approximation (MFA)**?

Approximating the inference and learning problem, using independence assumptions and decomposition into several products, leads to the idea of "mean-field" approximation. In other words, mean-field approximation is a way to simplify the Bayes procedure. MFA can be computed using coordinate ascent. See more [at this link](#).

5. Method:

Model Structures & Energies

Let's set out the paper's model explicitly. The total energy in the model is defined as follows:

$$E(x) = \sum_{i \in V} E_u(x_i) + \sum_{(i,j) \in N_t} E_t(x_i, x_j) + \sum_{c \in S} E_s(x_c)$$

- X_i is a discrete random variable over all pixels in the video sequence
- $E_u(x_i)$ is the **unary energy**, equal to the negative log likelihood of the labels for each random variable X_i
 - V is the set of pixels in the video
- $E_t(x_i, x_j) = \theta_t w_{ij} (x_i - x_j)^2$ is the **temporal energy**,
 - w_{ij} - is the weights of the temporal connections
 - N_t - is the set of temporal connections pixels, N_t defines a semi-dense optical flow by specifying a set of neighboring pixels which are connected to each other.
- $E_s(x_c) = \theta_s ||x_c - gCNN(x_c)||$ is the **spatial energy**
 - x_c - is the set of variables in the c clique.
 - S - is the set of spatial cliques, here all pixels in the frame are defined in the spatial clique.
- θ_u , θ_t and θ_s are the balance energy terms.

The exact MAP inference is NP-hard in general. The higher-order energy in this model makes the inference problem even harder and intractable even with efficient approximate algorithms like mean-field. Intuitively, this is because the algorithm needs to evaluate the total energy in the MRF for every frame in the video, requiring a CNN pass for each.

Inference

In order to make the problem tractable, the authors decoupled the temporal energy $E_t(x_i, x_j)$ and spatial energy $E_s(y_c)$ by introducing an auxiliary variable y , and minimize the following approximation of Eq. (3) instead.

$$\hat{E}(x, y) = \sum_{i \in V} E_u(x_i) + \sum_{(i,j) \in N_t} E_t(x_i, x_j) + \frac{\beta}{2} ||x - y||_2^2 + \sum_{c \in S} E_s(y_c)$$

β is a penalty parameter; y is a close approximation of x .

The above can be minimized by alternating steps updating either x or y iteratively. The two steps are temporal fusion and mask refinement. In addition **Iterated Conditional Modes (ICM)** is used to find an approximate solution of $E_t(x_i, x_j)$. At each iteration, ICM updated one random variable x_i while fixing the rest of the random variables x (see TF section in Algorithm 1 below). Moreover, $gCNN(x_c)$ is used as an approximation of y_c since solving it directly would be computationally difficult.

Algorithm 1 summarizes the inference model developed by the authors together with all the approximation assumptions describe above:

Algorithm 1 Our Inference Algorithm

Parameters: number of outer iterations K , number of inner iterations L , number of pixels N , and number of frames C .

Initialization: initial labeling $x^{(0)} = y^{(0)}$.

for k from 1 to K **do**

– *Temporal Fusion Step (TF)* –

$x^{(k,0)} \leftarrow x^{(k-1)}$

for l from 1 to L **do**

for i from 1 to N **do**

$$x_i^{(k,l)} \leftarrow \arg \min_{x_i} \left\{ \frac{\beta}{2} (x_i - y_i^{(k-1)})^2 + E_u(x_i) + \sum_{(i,j) \in \mathcal{N}_T} E_t(x_i, x_j^{(k,l-1)}) \right\}$$

end for

end for

$x^{(k)} \leftarrow x^{(k,L)}$

– *Mask Refinement Step (MR)* –

for c from 1 to C **do**

$y_c^{(k)} \leftarrow gCNN(x_c^{(k)})$

end for

end for

Output: Binarize $y^{(K)}$ as the final segmentation masks.

Implementation Details

- The $gCNN()$ takes a 4-channel input (RGB image + preliminary mask), and outputs a refined mask. $gCNN()$ is trained in two stages: (1) an offline model is trained using object segmentation data available, and (2) the offline model is fine-tuned using the ground truth mask in the first frame of a given video.

- A DeepLab framework was used for the mask refinement gCNN(). The backbone net is a VGG-Net. An additional skip was added, connecting intermediate pooling layers to a final output convolutional layer to enable multi-level feature fusion.
- In case of Multiple Objects -- each object is handled individually in each iteration before starting the next iteration. Overlapped regions are divided into connected pixel blobs and each blob is assigned to a label that minimizes for that blob.
- FlowNet2 used to compute optical flow.

6. Results:

Method	Global		Region \mathcal{J}		Contour \mathcal{F}	
	Mean	Boost	Mean	Recall	Mean	Recall
OSVOS [13]	0.574	–	0.546	0.598	0.601	0.675
Our baseline	0.596	–	0.558	0.617	0.633	0.715
+TF×1	0.589	-0.007	0.556	0.607	0.623	0.723
+TF×2	0.590	-0.006	0.556	0.609	0.623	0.722
+TF×3	0.590	-0.006	0.556	0.610	0.623	0.722
+TF×4	0.590	-0.006	0.556	0.611	0.623	0.722
+TF×5	0.590	-0.006	0.557	0.611	0.623	0.722
+MR×1	0.640	0.044	0.600	0.675	0.680	0.749
+MR×2	0.647	0.051	0.608	0.683	0.686	0.752
+MR×3	0.648	0.052	0.609	0.684	0.687	0.753
+MR×4	0.648	0.052	0.610	0.681	0.687	0.756
+MR×5	0.649	0.053	0.610	0.679	0.688	0.754
+TF&MR×1	0.692	0.096	0.652	0.728	0.732	0.822
+TF&MR×2	0.704	0.108	0.668	0.740	0.740	0.824
+TF&MR×3	0.706	0.110	0.671	0.742	0.741	0.816
+TF&MR×4	0.707	0.111	0.672	0.744	0.742	0.820
+TF&MR×5	0.707	0.111	0.672	0.744	0.742	0.820

In the above table: TF stand for temporal fusion, MR for mask refinement. **TF&MR \times n** means that the algorithm is performed for n iterations with both TF and MR. The table shows the results of an ablation study on the [DAVIS 2017](#) validation set. The baseline is OSVOS. The “Boost” column calculates the performance gain for each algorithm variant.



(a) Baseline

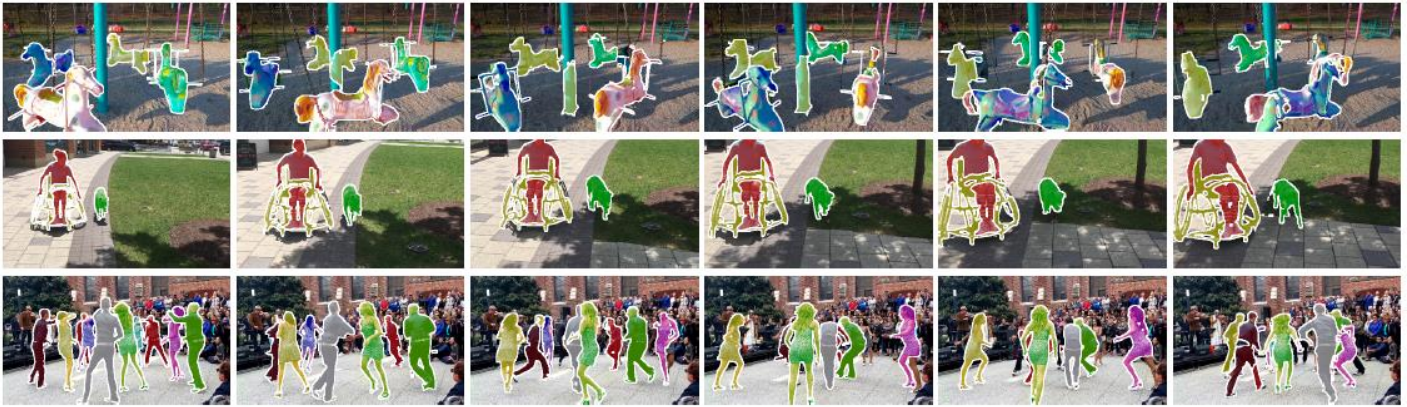
(b) Baseline+TF

(c) Baseline+MR

(d) Baseline+TF&MR

(e) Ground-Truth

The images at the bottom of the previous page are an example from the ablation study: use of TF (temporal fusion) or MR (mask refinement) separately achieved limited improvement, as seen in (b) and (c). However, combining TF and MR achieved a significantly improved performance (d).



Above, sample results from the **DAVIS 2017** test-dev set. The left-most column shows the ground-truth mask input for the first frame. Consecutive columns are segmentation results for subsequent frames. Different colors are used to highlight different objects. These samples show the highly challenging nature of the dataset. The authors didn't employ specific object detectors.



Above, sample results from the **DAVIS 2016** dataset.

Conclusion:

This paper is an attempt to mine the higher-order potentials of **combining MRF/CRF with CNNs**, by embedding a feed-forward pass of a CNN inside the inference of an MRF model. The authors implement an innovative spatio-temporal MRF model for video object segmentation. Their algorithm performs inference in the MRF model, and alternates between temporal fusion and a mask refinement feed-forward CNN, to incrementally infer video object segmentation. **The algorithm achieved state-of-the-art results on the DAVIS 2017 public benchmark.**

Collecting and Selecting Data for Medical Projects



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 **Arik Rond** tells us about **Collecting and Selecting Data for Medical Projects**. It's another tip by **RSIP Vision** for **Project Management in Computer Vision**.

If we understand what experts look for in the data, we can use this knowledge to generate input.

Availability of relevant data

It is a well-known issue for the project manager in computer vision. This problem is even more serious in medical projects, for many different reasons. Sparseness and limited number of existing cases in the real world; privacy constraints; scarce expert resources for annotation; reticence of those who own data to share it with others. These and other reasons explain project managers' struggle in finding medical data.

It is true that open sources do exist and medical challenges help making more data available for all. Of course, besides checking ownerships and legitimacy of any data, it must be verified if it was already labeled (when needed) and (if yes) whether annotation was done correctly. In many cases there are no labels and we shall explain in a follow-up article how to add them correctly and rapidly.

At this point, it is key that the project managers have a clear idea of the data they need and what their software is going to do. A good hint for a direction

is to think at what a doctor in the real world would care to look for in a patient. When an expert sees a medical scan and declares that it contains important information, it is generally possible to train a neural network or other system to do the same and reach the same conclusion. If the content is significant, the algorithm may see additional important information. Indeed, we have already seen computerized systems outperform physicians.

Of course, it is certainly better when this observer is a medical expert, like a doctor or a radiologist. However, expertise is rare and most times we have to make do with other experts, whose field of expertise overlaps certain aspects of the radiologist's work. They are biotechnologists, biologists or other professionals who have already done this kind of work in the past. They can be expected to know enough of the terminology and the process. In this way we are able to validate what we are looking for and where it can be found in the scan.

When we can't get a doctor or a biologist

Training data should be as diverse as the real world data which the algorithm will be fed with in due time.

or a medical student, we can "train" a human. This intuition is better obtained by having that person do manually the same procedure which the automated system does - looking at the data and the labels. After doing this for a few hours, the human observer will be well trained to look in the right place for the right thing.

Once data is collected it must be carefully validated, in order to be sure that the software will work on a valid input. All the research and development work would be made much more difficult (and in many cases impossible) by the use of wrong data. In any case, software results correctness must be validated too. Software should pass through many

controls and this is particularly true when input is limited in size.

Another key point in this phase is to care for the diversity of data, according to what the software is going to look for. Is it developed to analyse scans coming from different devices or from a single scanner? Will it be used only on adult scans or also on children? Bones of patients looks differently depending on age and even on ethnicity.

Training data should be as diverse as the real world data which the algorithm will be fed with in due time. Finally, the test set should somehow differ from the training set, i.e. including different patients. Otherwise the test would not be a real one.



Retina of patient with Diabetic Retinopathy. Fundus photo reveals scattered hemorrhages with cotton wool spots and some edema in the macula. If we understand what experts look for in the data, we can feed the algorithm this data as well.

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 **RSIP Vision's 3D Reconstruction and Deep Learning** work in the field of cardiology: **Left Atria Reconstruction from a Series of Sparse Catheter Paths Using Parametric Model and Neural Networks**. This research is the result of a cooperation between RSIP Vision, a major industrial concern and the **University of Tel Aviv**.

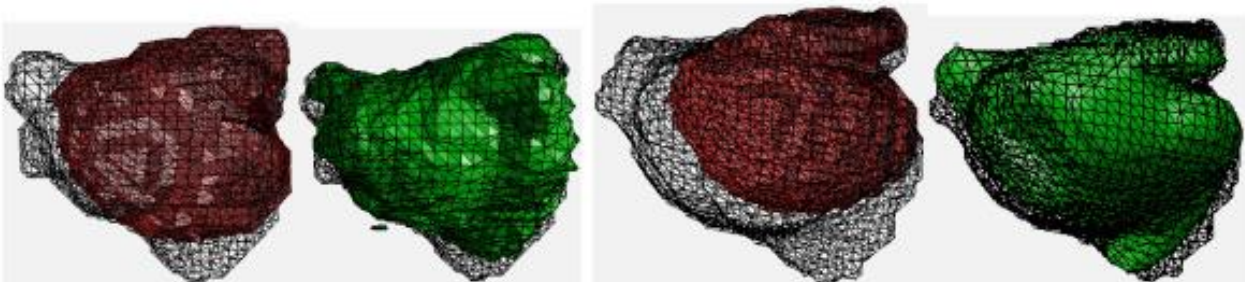
Here is where RSIP Vision's expertise is needed: our algorithm creates a map of the heart chamber using sparse data

Modeling and reconstructing the shape of a heart chamber from partial or noisy data is useful in many minimally invasive heart procedures. Current solutions involve navigating a catheter to the heart chamber, where a map of the chamber itself is needed, with its whole three-dimensional shape. A magnetic field with special properties is created in the area and the catheter is equipped with number of electrodes measuring the magnetic fields, by which it is possible to know exactly where the electrode is and as a result where the catheter is or moves. Then you can acquire a point cloud and eventually construct a 3D map of the chamber. But this process is time consuming and also prone to errors, as the catheter itself can deform the walls of the chamber, which will make it difficult to understand the real shape. Solutions marketed until now obtain only a noisy map which is very difficult

to understand and use.

Here is where RSIP Vision's expertise is needed: our algorithm is able to create a map of the heart chamber using sparse data. Even when we have only a sparse set of points, we can map the whole chamber. And even when the data is noisy, we use mathematical methods to reconstruct the shape of the chamber. As a consequence, we do not really need each and every point in the chamber: we can acquire a small set of key points and then construct the most probable shape, given this subset of points.

The way we did it is by creating a parametric model which describes the shape using only a small set of properties. In this way, the shape is calculated and displayed in real time to the physician, including information about the position of the pulmonary veins and how curved they are. The

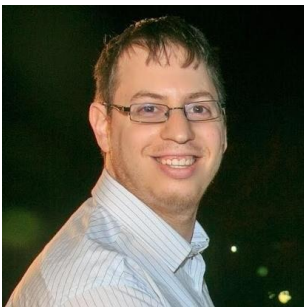


Side view of reconstruction - Red: input volume in ground-truth wire frame; Green: reconstruction result.

Take us along for your next Deep Learning project!

central part of the chamber is described by another mathematical formula, with different parameters. After all these components are described mathematically, they can be blended using mathematical functions to create a shape. By comparing these shapes to database from CTs, we show that these mathematical methods can describe almost any left atrium shape from the database of real patient data.

Being this a deformable model, it can be manipulated using a small number of parameters instead of many thousands of points or mesh vertices. We only need some tens of numbers to describe a wide variety of shapes. This model can be used to generate many artificial atria that look plausible, each with its own score: this can be used to augment the dataset of real patient atria to train a neural network.



Alon Baram
Tel Aviv University



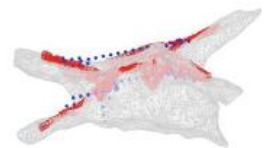
Moshe Safran
RSIP Vision

The model is based on a denoising autoencoder attempting to reconstruct the input given to it. The architecture is fully connected, but when training a denoising autoencoder one should set some percentage of the input to zero. The result obtained is a noisy input. Then, instead of just using noise, we deleted some area of the data, either by using an intersection of the sphere of the atria or by simulating the path of

a catheter introduced to perform an ablation, as would be done by a physician during a real procedure. This part takes only one minute or so and is followed by the attempt to reconstruct the shape.

Alon Baram of Tel Aviv University says: *“What is challenging here is that the network needs to imagine how the 3D shape will look like using the statistics that it has learned during training. For this, we introduced a new regularization that helps the network learn smooth shapes. This is done by using a spatial gradient of the weights, as low as possible. This has not been tried on any network: unlike a convolutional neural network where weights are shared along the space, this network is fully connected but constrained to learn only smooth shapes.”*

We can say that this collaboration combines two technologies in a creative manner: one is a parametric model, describing the shape using a limited number of parameters to get a geometrical and mathematical representation of it; the second is a neural network. The first technology is used to generate artificial models to provide data augmentation needed for training the neural networks to do the 3D reconstruction task.



Left: rigid phantoms;
center: CAD to path registration;
right: phantom registration - red is recorded path and blue is template




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



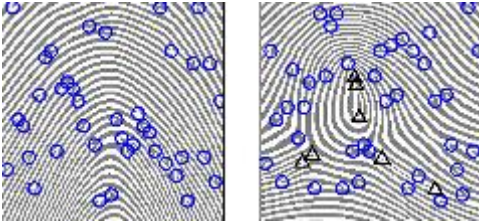
[Why Big Tech pays poor Kenyans to teach self-driving cars](#)

Or what I call “the other side of training”: whatever you think of it, this is the cheap way often used to provide **correctly labelled data** to all kind of **data-hungry Artificial Intelligence networks**. They also talk about the great impact of our friends at [Samasource](#)! [Read More...](#)



[How to Encode a Secret Message in a Fingerprint:](#)

Fingerprints are no longer just a way to identify people: they can also be used to send private notes and it's easier than you think. Researchers construct synthetic fingerprints in which ridges and bifurcation patterns hide a secret meaning, while mimicking legitimate fingerprints. [Read More...](#)



[Create animated GIFs with OpenCV:](#)

Adrian Rosebrock has posted on **his excellent blog** a tutorial teaching how to create animated GIFs with **OpenCV**, **Python**, and ImageMagick. You can even download the code and do something fun while learning to use important computer vision tools. [Enjoy!](#)



[Machine Learning resources that are legally accessible online for free:](#)

Harvard-MIT PhD candidate Sam Finlayson have collected and posted on his github a very nice list of **Machine Learning resources**: courses, textbooks, tutorials, notes, cheatsheets, you name it. When you thank him, tell him that he forgot to mention **Computer Vision News!** But tell him kudos anyway because he's promoting [a very noble cause](#). [Go to his list of ML Resources...](#)

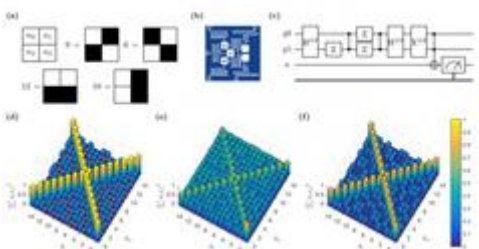
[What is adversarial AI and why does it matter:](#)

No, this is not about GAN. It says how criminals can use great technology to harm governments, businesses and citizens. It also offers valid solution. [Worth Reading!](#)



[Intelligent Machines - Machine learning / quantum computing](#)

While we work on deep and machine learning, another information processing revolution is in its infancy: quantum computing. Researchers are starting to test the potential of merging these two technologies. [A Fascinating Read!](#)



[How Much Do You Really Want AI Running Your Life?](#) Are there tasks which should always remain under human control? [Good Question! Here Is a Tough Answer...](#)

by M. Alper Selver

The CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge has two aims: segmentation of liver from CT and segmentation of four abdominal organs. We have asked main organizer M. Alper Selver to give our readers an exclusive overview of this challenge and we are grateful for this contribution.

Clinical Motivation:

The segmentation of abdominal organs has critical importance for several clinical procedures such as pre-evaluation of liver for living donor based transplantation surgery or detailed analysis of abdominal organs and vascular tree for correct positioning of a graft onto an aortic aneurysm. This motivates an ongoing research to achieve better segmentation results and requires overcoming countless challenges originating from both highly flexible anatomical properties of abdomen and limitations of imaging modalities.

History and Background:

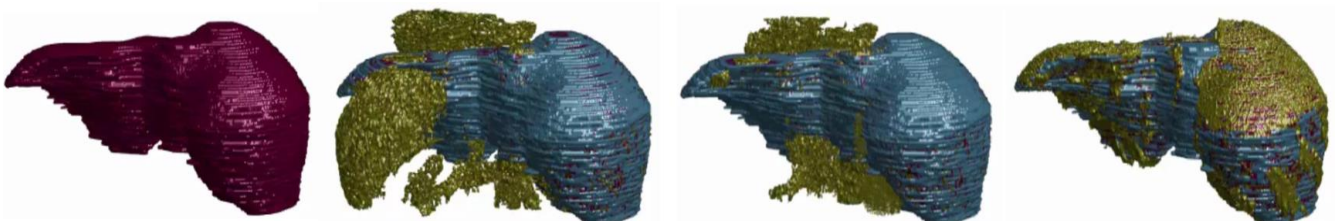
In 2007, **SLIVER challenge** provided a comparative study of a range of algorithms for liver segmentation from CT and reported a snapshot of the methods that were popular for medical image analysis. In 2015, the **VISCERAL Anatomy challenge** has



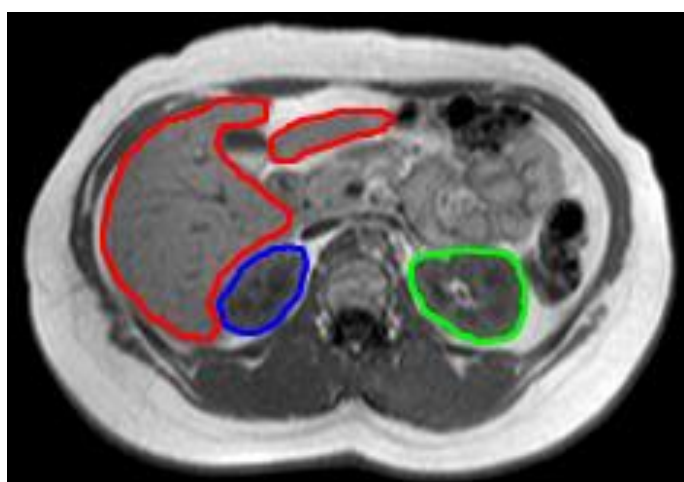
A. Emre Kavur (holding a 3D printed liver) and M. Alper Selver at the 11th Turkish Medical Informatics Congress (November 2018, Ankara Turkey).

brought significant contributions to the field. Since then, **machine learning** based automatic strategies, especially **deep learning through convolutional neural networks**, introduced significant novelties and improvements to medical image segmentation. We want to observe the contributions of these latest developments through CHAOS.

Therefore, while developing and training the algorithms, it is allowed to include other datasets from other sources such as SLIVER or VISCERAL. Besides using other data sets, utilization of new approaches such as



3D visualization of the ground truth and outcomes of different segmentation methods



Example annotations for T1-DUAL images

(CT+MRI), 5. Segmentation of abdominal organs (MRI only).

Teams can participate in a single category or can submit for multiple categories using different systems. However, categories 1 and 4 are especially important for us since there already exist very successful models individually for CT and MRI, but not working effectively in both at the same time.

Follow Ups of CHAOS:

CHAOS is not only about finding the best method. Besides providing a comparative study of participating algorithms, the competition will be used to gain insight about complementarity and diversity of different methods. Recent developments show that classifier ensembles can provide higher performance than its components if certain conditions are met. **EMMA**, which won the [BRATS challenge](#) in 2018, is one of the most recent examples. We believe that the outcomes of CHAOS will spark many ideas to improve existing ensemble strategies.

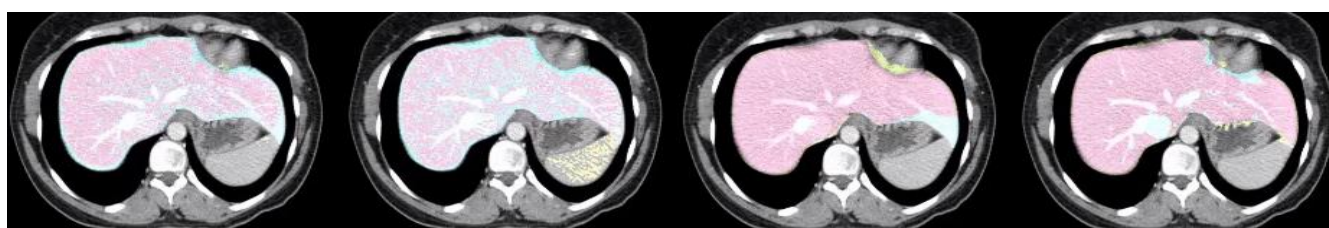
Furthermore, it is already being planned to extend the CHAOS to cover vessel tree extraction inside the liver and Coinaud classification to determine the volumes of its sub-segments.

transfer learning to fine tune a trained model to abdominal organ segmentation or other strategies such as data augmentation are also encouraged.

Competition categories and aim:

CHAOS has two aims: segmentation of liver **from CT**; and segmentation of four abdominal organs (i.e. liver, spleen, right and left kidneys) **from MRI** acquired with two sequences (T1-DUAL and T2-SPiR).

There will be five competition categories in which the participating teams can take place and submit their result(s): 1. Liver Segmentation (CT-MRI), 2. Liver Segmentation (CT only), 3. Liver Segmentation (MRI only), 4. Segmentation of abdominal organs



2D illustration of the outcomes of different segmentation methods
Red: True Positives, Blue: False Negatives, Yellow: False Positives

Inky is a cybersecurity company which claims to have developed the first email protection software to detect phishing attacks using computer vision, artificial intelligence, and machine learning. Dave Baggett, its Founder and CEO, elaborates on the different types of phishing attacks and the many challenges in developing this technology.

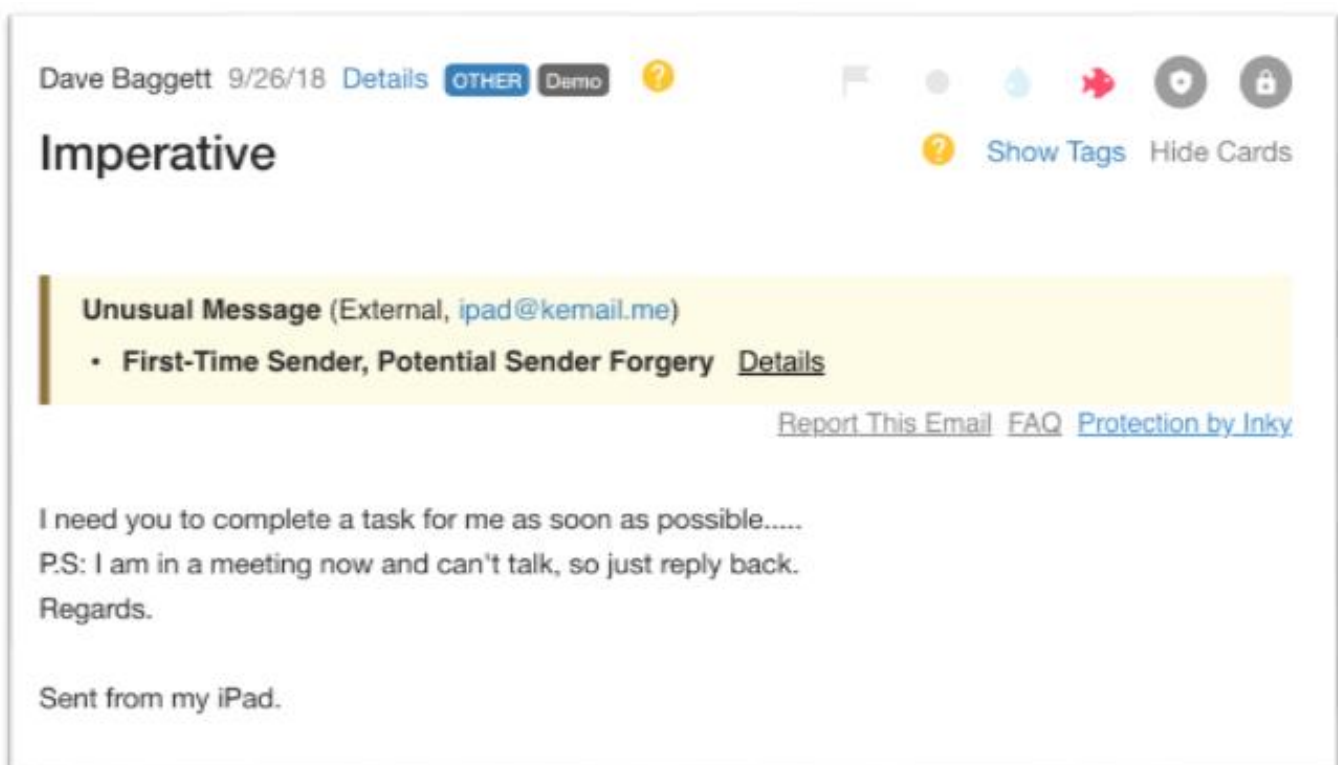
Phishing emails come in two forms. The first, **spear phishing**, resembles an email from someone important, either inside or sometimes outside of a company. For example, an employee named Lisa receives an email that looks like it came from her coworker a few weeks ago, when in fact, it actually came from an impersonator. This is also referred to as **CEO impersonation**

or business email compromise.

The second type of phishing involves **forgery of a brand** such as a fake Microsoft message that alerts the user that they have exceeded their mail quota and directs them to log in to fix it. When the person clicks on the link, they do not open an actual Microsoft site. They enter their password anyway, giving the attacker their Microsoft Office password.

On one hand, Inky has algorithms that block the spear phishing or individual person impersonation cases. On the other hand, they have computer vision algorithms that help **identify and block brand forgeries**. Baggett explains:

“There are a few reasons why it is a hard problem. One is that most of the work in computer vision has been done on three-dimensional photographic imagery. You know, take a picture and



Example of CEO Impersonation or Business Email Compromise

put it into a model, and the model tells whether the picture is a cat, right? And that's not really useful for this problem because most of the imagery that is relevant for identifying brands is more two-dimensional logo graphics kind of imagery. There is a fair bit of work to just apply what's obviously very well treated in the literature on standard computer vision to this kind of use case."

Their technology requires a very specific domain application of a general set of techniques that have been perfected in the academic literature over the last 10 years. Transferring that sort of technique to this domain has been quite challenging, and computer vision, although a powerful tool, does not solve this problem sufficiently on its own. In general, it needs to corroborate with a whole host of other signals extracted from the mail. The software uses a lot of domain-specific email techniques, many of which do not involve machine learning such as

approximate string matching or heuristics. Developing their technology requires experts in both computer vision and email. Because not many people have this intersecting kind of work experience, **it has taken a long time for computer vision to bleed over into mail protection.** "Imagine taking just the standard, off the shelf TensorFlow computer vision model like any of your practitioners would do. Now you give it an email. Now what? That whole process of applying the off the shelf stuff to the mail involves a couple of interesting problems" shares Baggett.

A fraudulent email may or may not even contain any imagery at all. Baggett says that they often come across an email containing what looks like a logo but is actually just HTML. For instance, an email that impersonates the retailer, Target, will have a big white word, Target, on a red background, which looks quite convincing. That means Inky must deal with the process of rendering HTML.

From: Dropbøx <aalert@auth.dropbox-bus.net>

Date: Tuesday, February 27, 2018 at 1:17 PM

Subject: Payroll Report in dropbox file

The image shows a blue rectangular button with the word "Dropbox" written in white, sans-serif font.

Your payroll report has been added and uploaded in your drop file. You need your e-signature to encrypt the secured doc.

[Encrypt Secure Doc.](#)

You're subscribed to get updates like this one from Dropbox premier partner. To stop receiving them, [unsubscribe.](#)

Your Online Banking Information needs to be verified.



Bank of America
Wed 5/23, 11:35 AM
Phish



Reply all | v

Suspicious Message (External, noreply-support@comcast.net)

- Google Safe Browsing URL, Spam Content, and more... [Details](#)

[Report This Email](#) [Powered by Inky](#)

Bank of America 

Online Banking



Online Banking Alert

Irregular Check Card Activity

We detected irregular activity on your Bank of America Check Card. For your protection, you must verify this activity before you can continue using your card.

Please visit Online Banking

at www.bankofamerica.com/protectcard.cgi to review your account activity. We will review the activity on your account and upon verification, we will remove any restrictions placed on your account.

This alert relates to your Online Banking profile, rather than a particular account. This is for verification purposes only.

Want to confirm this email is from Bank of America? Sign in to Online Banking and select Alerts History to verify this alert.

Because email is not a secure form of communication, please do not reply to this email.

Baggett elaborates, *“Think about of all the complexities embedded in the Chrome browser. HTML is sort of its own monstrously complicated area which vastly complicates the problems simply using computer vision models for this task.”*

Often, when looking at a logo in a marketing email or a transactional email, like a Bank of America logo, your brain assumes that the logo appears in isolation in the email. This is almost never the case, and emails usually have a giant section of imagery which combines notionally separate elements. Baggett and his team cannot take each of the images that appear in the mail and run them through a model. Instead, they must analyze and semantically separate the components of one big image like the logo, text, and so on. This presents a problem of image segmentation and even some OCR-related issues.

Not every email containing a logo indicates a phishing attack either. If an email has a Facebook logo, for instance, the image might just guide their readers to like them on Facebook. Inky must take this into account as well so as not to generate tons of false positives.

“They are trying to fool the machines, and they are also trying to fool the humans!”

Combating phishing is *“just an arms race”*. As Inky improves the models, attackers will improve their tactics. Baggett points out:

“Imagine an email might come in that claims to be from American Express.

One of the things the attackers are doing now is they are simultaneously trying to fool the machines into believing the mail is legitimate or not even transactional. They are trying to fool the machines, and they are also trying to fool the humans. They will make a mail that looks really convincing like, let’s say, an American Express transactional mail. Maybe it tells you your card was used in China. Then they will make subtle modifications to that mail to try to pass through the mail protection software.”

He continues: *“One of the things they will do is try to cloak all of the strings in the mail that are brand indicative. For example, they might take the A in American Express and replace it with a different or similar looking Unicode character, or they might modify American Express slightly.”*

Inky must take on the task of applying approximate string matching across the entire text of the mail to indicate that, although similar, it’s not really American Express. They continue to do active work in this area to go from Generation 4 to Generation 5, particularly around accelerating the performance of those. Performing heavy weight computation on every single mail is costly: they are working on reducing costs by making these algorithms faster.

Facebook needs to have a lot of labeled examples in order to train a face recognition model to have a billion face images. Similarly, Inky has a lot of example emails, but they are not labeled. One of their scientists, who did his PhD work in a semi-supervised learning environment with a very large



Dave Baggett

“...it becomes totally obvious that the approach they were using would not work.”

set of unlabeled examples, continues to explore how much leverage they can get out of totally unlabeled examples of emails.

Throughout the process of making their technology work, they have come across funny stories every step of the way. At certain points, it becomes totally obvious that the approach they were using would not work. Baggett recalls:

“I mentioned the logo graphic, HTML. We had a user who reported a fake Target email, and Inky didn’t catch it. As usual when someone reports a mail, we look at it and at how would we have ever caught this. We are looking, and there is no image at all. That motivates us to think more broadly about these kinds of things in a way that is very hard to anticipate. Until you see the data, it’s like anything else

in science, you can’t really imagine all the pitfalls. So there has been a lot of things like that.”

Working in the industry, Baggett has come across many terrifying and entertaining examples of real-world phishing emails, from Bitcoin messages and even phony Thanksgiving emails. They recently published a blog post on spear phishing attacks that impersonate a family member or friend to wish a Happy Thanksgiving. Baggett suspects that this particular scam comes from a single phishing kit. It requires an impressive amount of work to create these various greetings. In his own words, *“It’s not surprising, but it is entertaining to see the length that some of these people will go to try to scam people.”*

Attackers now simply take a real American Express mail and simply change the URL, making the email visually identical. To combat this, they have to figure out a way to tell if the mail really comes from American Express. After all, they do not have a clean data source indicating all of the mail domains of American Express.

In some cases, when the forgery looks so bad, they might assume that users would not fall for it. For example, a fake Microsoft mail that doesn’t have a Microsoft logo, but it just says log into your Office 365 account. Baggett and his team might not think to warn their users because the forgery looks so fake. However, they still need to accommodate the psychology of the end user and deal with these types of obvious forgeries even when you expect they will not fool anyone.

Baggett elaborates: *“I think that is a really interesting area because it’s not*

“That is when you really have something powerful!”


just about solving a computer science problem, it’s about modeling the psychology of your end user.”

Drawing from that thought process, Inky developed a feature that no one else has. Their software puts a warning banner right in the mail telling the users specifically what is wrong with the mail. In the case of a Microsoft

banner impersonation, the email will have a big red banner indicating the fraudulent component at the top. Their users love this feature because they can learn to recognize the phishing attacks and differentiate between a real email and a fake.

He concludes: *“There is a feedback effect there and that is really interesting. You can release something to the market, which not only changes the marketplace, but also changes the end user’s psychology. **That is when you really have something powerful.”***


Alert: Your American Express was used to signed in from a different IP address.

 American Express Membership Support 🔗 📧 ↻ Reply all | ▾
 Wed 5/23, 11:37 AM
 Phish ▾


Suspicious Message (External, id169@aexp-ib.com)
 • **Brand Impersonation** [Details](#)


[Report This Email](#) [Powered by Inky](#)

Please verify a recent activity



Hello Membership Card,

Account Ending: 31*** 

 **Fraud Protection**

For your security, we regularly monitor accounts for possible fraudulent activity. Below are the details :

Login Date:	17/08/17
Country :	China
City:	杭州市
IP Address:	180.092.060.274

For your security, new charges on the accounts listed above may be declined. if applicable, you should advise any Additional Card Member(s) on your account that their new charges may also be declined. To safeguard your account, please access your account

Thank you for helping us to protect the security of your account.

American Express Account Protection Services

Your Cardmember information is included in the upper portion of this message to help you recognise this as a customer service email from American Express. We are unable to answer replies to this email. You may contact us securely using the customer service link below.

[Contact Us](#) | [Privacy Statement](#)

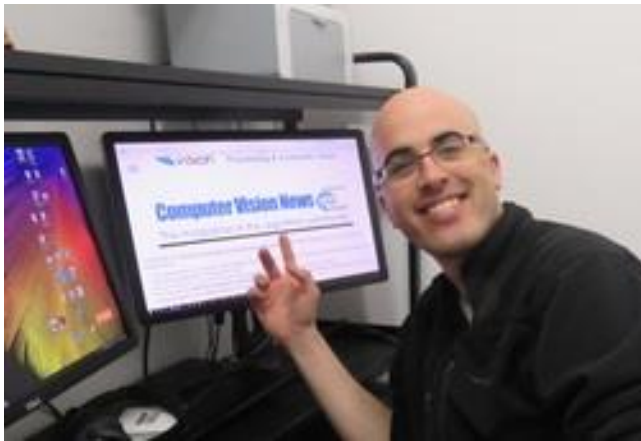
American Express Limited 12 Shelley Street, NSW Sydney 2000 (ABN 92 108 952 085). Australian Credit License No. 291313. © Registered trademark of American Express Company.

© 2017 American Express Company. All rights reserved.

AUSENALEFRD0004

Application

by Assaf Spanier



A new tool for dealing with the hyperparameter optimization problem

This month we shall review **BOHB**, a new tool for dealing with the **hyperparameter optimization problem**, presented at the **ICML2018** conference. BOHB offers a significant speedup by combining the advantages from both of the optimization approaches currently in use: **Hyperband** and **Bayesian optimization**. BOHB converges much more quickly than either approach, moreover, finding several optimal hyperparameter value combinations.

The last decade has seen an explosion in the successful application of Machine Learning, and Deep Learning in particular, to a very wide range of fields, these achievements depend to a large degree on **correctly selecting hyperparameter values during the training stage**. Choosing the wrong value for the most fundamental hyperparameter -- the learning rate -- will result in a failed training of the network. Of course, there are plenty of other hyperparameters that need to be set during network training, such as: network architecture, the regularization module, and many more. A common solution in current use is to simply employ a random search within a certain range of values considered relevant, train with each setting and pick the most precise result, but this is obviously extremely inefficient.

Instead, the performance of deep learning algorithms can be defined as an objective-function of their hyperparameters. We will define the **Hyperparameter Optimization (HPO)** problem as finding the value for each hyperparameter such that it minimizes this objective-function's value.

Two methods for tackling the problem of minimization of the objective-function are currently popular:

- (1) **Bayesian optimization** (BO; Shahriari et al., 2016)
- (2) **Hyperband** (HB; Li et al., 2016)

The Hyperband methods employ a Bandit-Based algorithm, which uses a random search, and therefore sometimes don't quickly converge to the best configuration. Bayesian approaches, on the other hand, usually require immense computation power, nearly impossible to achieve. BOHB is an approach that combines **the advantages of Bayesian optimization with the efficiency of using Bandit-Based methods**, to achieve the best of both worlds: strong performance

and rapid convergence to optimal configuration. BOHB is a practical, innovative method for optimization, that for a wide range of different fields outperforms both optimization methods currently in use separately, by combining them.

1. Bayesian Optimization -- models the hyperparameter objective-function using a probabilistic model based on the set of observed data points. Using this model, BO proposes a new hyperparameter configuration setting. BO iterates over three steps: (a) select the point that optimizes for the current data point selection method; (b) evaluate the objective-function at this point; and (c) add the new observation to the data and refit the model. The selected data points determine a trade-off between exploration and exploitation. BO requires long computation times to build-up its model so as to find better configurations. With sufficient computational power the model acquires more data points and achieves high optimization.

2. Hyperband Optimization (Hyperband) is a method that uses low budgets of quick-and-fast approximations of the objective function. HB calls **SuccessiveHalving** (SH; Jamieson et al., 2016) to identify the best out of n randomly-sampled configurations. SH evaluates these n configurations with a small budget, keeps the best half and doubles their budget.

Hyperband Optimization balances between quick training with a low budget and long training using a higher budget. This helps you conserve your resources, by using a low budget when only a short training time is needed. The literature reports Hyperband Optimization performs very well with a low to medium budget, better than either random search or Bayesian Optimization, however, its convergence is limited by dependence on randomly-drawn configurations: with higher budgets its advantage over random search decreases radically.

Bayesian Optimization, on the other hand, exhibits results very similar to random search in early training iterations, which are equivalent to low to medium budgets, and does not achieve results as good as Hyperband. For higher budgets, however, Bayesian Optimization usually manages to scan much wider search areas, thus outperforming Hyperband.

BOHB is a simple yet efficient approach to hyperparameter optimization, which offers the following advantages: it is **strong, flexible and expandable** (both in terms of handling more dimensions and parallel processing); and it achieves **strong performance results under all conditions**. BOHB showed impressive results in the series of experiments conducted that met the requirements and expectations set for it. But you don't need to trust published papers blindly: we recommend you try to implement BOHB to see for yourself and we shall demonstrate a simple code for doing so.

Code:

In the following example we will show how to set up a small CNN network on Keras and train it on the MNIST dataset. The table below shows the common hyperparameters for which we will be seeking optimal values, including dependencies between hyperparameters (for instance, momentum only receives a value if SGD optimizer is used). The code is short, simple and pretty self-explanatory.

Parameter Name	Parameter type	Range/Choices
Learning rate	float	[1e-6, 1e-2]
Optimizer	categorical	{Adam, SGD }
SGD momentum	float	[0, 0.99]
Number of conv layers	integer	[1,3]
Number of filters in the first conv layer	integer	[4, 64]
Number of filters in the second conv layer	integer	[4, 64]
Number of filters in the third conv layer	integer	[4, 64]
Dropout rate	float	[0, 0.9]
Number of hidden units in fully connected layer	integer	[8,256]

The following are the three basic ingredients needed to apply HpBandSter to a new optimization problem:

- (1) Implementing a Worker -- the worker is responsible for evaluating a given model with a single configuration on a single budget at a time.
- (2) Defining the Search Space -- defining the parameters to be optimized.
- (2) Picking the Budgets and the Number of Iterations -- besides the number of iterations you also need to specify a meaningful budget (as will be discussed later, this is one of the drawbacks of HpBandSter).

The `__init__` function

As its name implies, this is the initialization function -- it's responsible for reading the data, uploading data and normalizing data -- to prepare it for training. In this case we are reading the well known MNIST dataset.


```
class KerasWorker(Worker):
    def __init__(self, N_train=8192, N_valid=1024, **kwargs):
        super().__init__(**kwargs)

        self.batch_size = 64

        img_rows = 28
        img_cols = 28
        self.num_classes = 10

        # the data, split between train and test sets
        (x_train, y_train), (x_test, y_test) = mnist.load_data()

        self.x_train, self.y_train =
            x_train[:N_train], y_train[:N_train]
        self.x_validation, self.y_validation =
            x_train[-N_valid:], y_train[-N_valid:]
        self.x_test, self.y_test = x_test, y_test

        self.input_shape = (img_rows, img_cols, 1)
```

Don't miss Women in Computer Vision at page 32



Who is this unsuspecting fan photobombed by William Shatner at a Star Trek convention?

Next up is the compute function, which sets up the network. As you can see, throughout the process of setting up the network there are points at which instead of setting values explicitly it calls config. The config file comprises the hyperparameters we will be selecting at optimal values for using BOHB.

```
def compute(self, config, budget, working_directory, *args, **kwargs):
    model = Sequential()
    model.add(Conv2D(config['num_filters_1'], kernel_size=(3,3),
                    activation='relu',
                    input_shape=self.input_shape))
    model.add(MaxPooling2D(pool_size=(2, 2)))

    model.add(Dropout(config['dropout_rate']))
    model.add(Flatten())
    model.add(Dense(config['num_fc_units'], activation='relu'))
    model.add(Dropout(config['dropout_rate']))
    model.add(Dense(self.num_classes, activation='softmax'))

    if config['optimizer'] == 'Adam':
        optimizer = keras.optimizers.Adam(lr=config['lr'])
    else:
        optimizer = keras.optimizers.SGD(lr=config['lr'],
        momentum=config['sgd_momentum'])

    model.compile(loss=keras.losses.categorical_crossentropy,
                 optimizer=optimizer,
                 metrics=['accuracy'])

    model.fit(self.x_train, self.y_train,
             batch_size=self.batch_size,
             epochs=int(budget),
             verbose=0,
             validation_data=(self.x_test, self.y_test))

    train_score = model.evaluate(self.x_train, self.y_train, verbose=0)
    val_score = model.evaluate(self.x_validation, self.y_validation )
    test_score = model.evaluate(self.x_test, self.y_test, verbose=0)

    return ({
        'loss': 1-val_score[1], # HpBandSter always minimizes!
        'info': { 'test accuracy': test_score[1],
                  'train accuracy': train_score[1],
                  'validation accuracy': val_score[1],
```

Focus on

Lastly, we have the `get_config` function, which defines the range of values to be considered and searched by BOHB for each parameter.

```
@staticmethod
def get_configspace():
    cs = CS.ConfigurationSpace()

    lr = CSH.UniformFloatHyperparameter('lr', lower=1e-6, upper=1e-1,
                                         default_value='1e-2', log=True)

    optimizer = CSH.CategoricalHyperparameter('optimizer', ['Adam', 'SGD'])

    sgd_momentum = CSH.UniformFloatHyperparameter('sgd_momentum',
                                                    lower=0.0, upper=0.99, default_value=0.9, log=False)

    cs.add_hyperparameters([lr, optimizer, sgd_momentum])

    num_conv_layers = CSH.UniformIntegerHyperparameter('num_conv_layers',
                                                        lower=1, upper=3, default_value=2)

    num_filters_1 = CSH.UniformIntegerHyperparameter('num_filters_1',
                                                      lower=4, upper=64, default_value=16, log=True)

    cs.add_hyperparameters([num_conv_layers, num_filters_1,
                            num_filters_2, num_filters_3])

    dropout_rate = CSH.UniformFloatHyperparameter('dropout_rate',
                                                    lower=0.0, upper=0.9, default_value=0.5, log=False)
    num_fc_units = CSH.UniformIntegerHyperparameter('num_fc_units',
                                                     lower=8, upper=256, default_value=32, log=True)

    cs.add_hyperparameters([dropout_rate, num_fc_units])

    return cs
```

To run the code, call `worker.compute`, the static function containing the configuration; you will also need to set the budget for each training.

Please note: The budget should be as low as possible, while still being informative. By informative we mean that performance on the limited budget will be indicative / sufficiently correlated with performance on a high budget. Correctly selecting the budget is completely dependent on the problem at hand: it requires knowledge of the field and data being studied and for which you are training your network.


```
if __name__ == "__main__":
    worker = KerasWorker(run_id='0')
    cs = worker.get_configspace()

    config = cs.sample_configuration().get_dictionary()
    print(config)
    res = worker.compute(config=config, budget=1, working_directory='.')
    print(res)
```

For more explanations and examples, see the following websites:

<https://automl.github.io/HpBandSter> and https://www.automl.org/blog_bohb/

Drawbacks:

To use BOHB (also Hyperband) you must be able to set (and allocate) a meaningful budget: The low budget used for each training needs to (relatively) cheaply give a good indication of the function's performance when the full budget will be used to run it on the entire dataset with a longer run-time. That is, the relative ranking of different hyperparameter configurations needs to be correlated with the relative ranking of the same configurations for the full budget. If the evaluations arrived at for low budgets are biased in some way, or simply too noisy, to be a good indication of the configurations that should be used for optimal performance on the full budget, then the Hyperband element of BOHB is a waste. In these cases, BOHB will just be k times slower than regular BO, where k is the number of Hyperband iterations implemented by that BOHB (this is the reason Hyperband can actually be worse than random search in the worst case scenario.)



Feedback of the Month



*We got in touch with **RSIP Vision** with a challenging project with many adversities and imponderables. The **good and successful cooperation with RSIP Vision** allowed us to take a big step forward.*

*We really much appreciated their **project management** and the **clear and transparent way of communication!** Thank you very much!*

Christiane Michaelsen
Managing Director, IDENTT SWISS GmbH

Focus on

Can you see what these visual metaphors mean?
Do you know what creative message they convey?
Read about their connection with Computer Vision
on the next pages...



Lydia Chilton


Lydia Chilton is an assistant professor at Columbia University in New York where she works in the Computer Science Department. Her research is in human-computer interaction (HCI), crowdsourcing, and computational design. [More interviews with women scientists](#)

Lydia, can I ask you about your work?

I do crowdsourcing. One of my main goals is to decompose hard problems so that individuals can do them better or they're easier for groups to break up. Many of those problems are about visual communication, how visual images can convey a meaning. This is used a lot in print media. Often, if you have a news article, you want an image

to go at the top that will attract attention. The cover of the Economist does it a lot. They have a magazine with lots and lots of words printed inside it. If you don't see the cover and get interested in the main story, you may not get to all those words in the middle. Visual communication tries to send a message like a headline or a public service announcement. It requires some creativity, but also there's some grounding in what it has to do so there's an actual purpose to it that we can evaluate in its meaning.

The first time I heard about your work was at CVPR2018, at [Adriana Kovashka's](#) workshop about understanding visual advertisements, and I was impressed by your presentation. What is the connection between your work and computer vision?



“When you are interpreting symbols, first associations are better than logic”



“Most symbols are highly ambiguous. They need a little bit of context and a little bit of background knowledge to ground them at all.”

Advertisements are one way to convey a message, often a very simple one. For example, they want to associate that *Tabasco sauce is hot* or *Red Bull gives you energy*. They do it through these things called visual metaphors. It's not just a technique in advertising. It's also used in journalism. It's also used in public service announcements. When I give a talk, the first slide I have is a visual metaphor. What's nice about visual metaphors is, not only do they capture attention and convey a meaning, but they are a little bit of a puzzle because you usually see two

objects blended together. So like a Red Bull can and a battery packet. You recognize both individually, but your brain is like: *“Huh, something is wrong there”*. It takes you half a second to figure out what the message is. That is valuable for people remembering and recalling this information. You don't need to think about it twice.

Does it really have to be wrong?

It's something that your brain can't automatically classify. If you see just a Tabasco bottle, you don't notice it. Because you've seen it before, there's nothing interesting or challenging about it. Metaphors are useful across language and vision in getting your product picked. It helps you explain something new in terms of something else that you already know. There is a cultural dependency here. The overall rule is you need to use symbols that your audience will understand.

You have a very strong sentence that says: “When you are interpreting symbols, first associations are better than logic.” Can you explain that?

One of these visual metaphors is trying to convey that Tabasco is hot. That's a message that most people have already heard so you have some background knowledge of that information. If I were trying to tell you that Tabasco is for curing cancer, that would be very weird because you never heard that before, and it would probably need more information than a Tabasco bottle put in a fire extinguisher. Instead of the fire extinguisher cylinder, they put Tabasco. Because it has the handle, the strap, the little hose, and the gauge to tell you how much pressure it has, all those things identify it as a fire extinguisher.



***“Your brain is like:
‘Huh, something is
wrong there!’”***

Almost everyone gets it immediately that Tabasco is hot. If you think about it, really, really carefully, you could say, *“Well, fire extinguishers are really for putting out heat so doesn't this say that Tabasco is not hot?”* No, it doesn't mean that.

So it could miss the point completely!

This is true for all metaphors. You have to decide what property is being transferred. The textbook example of a literary metaphor is *“the classroom was a zoo”*. So what property of a zoo is being transferred? It means that the kids were out of control and wild like

animals, but it could equally well mean that you had to pay \$12.50 to get in or that the children were in cages being abused. All sorts of things that could apply, don't. It is unusual with the Tabasco case that the fire extinguisher is the opposite of being hot, but still, it is sort of a symbol of being hot. Most symbols are highly ambiguous. They need a little bit of context and a little bit of background knowledge to ground them at all. This is an amazing part of the human language's ability to communicate. Whereas, nothing we say is perfectly 100% clear and a lot of the time we can fill in the gaps with our background knowledge. As long as 50% of people get it, and there's no other meaning that is too threatening to it, it's pretty good. If you look into any metaphor or anything really too hard, you'll find the flaws and the point where it cracks. No one really has time to think through all of this. It's about those first impressions.

You have spent the biggest part of this century in academia: Washington, Stanford, MIT and Columbia. What is that keeps you in academia?

I have done internships at Microsoft and at Google, sort of on the research side. I like to pick problems that will take five to ten years to do well. It's hard to see the immediate business value of it. Understanding visual metaphors and creating them is not something that I can just go to a company and say, *“Hey, I want to spend my time doing this. It's going to require careful study, and it's not going to be useful for quite a number of years. It may not work at all.”* I'm attracted to those kinds of problems. I like things that I don't know how to do,

and I don't have any tools from the beginning. I just want to spend a bunch of time figuring it out. I like the pleasure of finding things out. I found that I have the freedom to do that in the academia. It is very possible that if I looked harder, had the right connections, or just had a better first experience or a different first experience in the industry, I would have found that as well. You don't have time to search forever. If you find something that works, it's good enough. I enjoy teaching as well! I enjoy undergraduates and working with them. I use them as a great way to tell whether people think the research is cool or not.

What is the dream output of your work?

There are two possible things that would be nice. One is a fully automated system where you input the message that you want to convey and it gives you an image. Here is a visual metaphor that we have already tested. It has "such and such" a click-through rate, and it's done. That would be impressive if we could totally understand those problems to the degree that it can be done without any input. That is the sort of traditional artificial intelligence, machine learning, and computer vision way of seeing things. I think that communication, particularly with humans, is very hard to fully decompose. Even all the images on Google don't take into account all the things that we see in everyday life. They are very biased towards things that are exciting enough to take a picture of. Another fantastic outcome for me would be a decomposition of the problem. We get the computer vision to do some of the parts that are hard for people like mocking up the results, creating that final image, doing a big

search for symbols like what represents fire, and hot, and energy, and all these things. Computers can help with those. Humans continue to drive the search. We'd like to have more feedback in the system, decomposing the problem and making it far easier for a person to do. A person could spend like 30 seconds guiding the search and then redirecting it when it's wrong.

"I like the pleasure of finding things out!"

Tell us about China, where you lived three times.

There are some parts about China that I really enjoyed. Excellent food, obviously. Microsoft Research Asia was amazing. It has a very international group. Obviously, there are a lot of Chinese people there, but one of my bosses was from England. There was a designer from Italy, and a graphic artist from New Zealand. There were a lot of people who ended up there. I always felt welcome. I speak Chinese. I don't read or write. That's the hard part, but I can order food.

I got a lot out of the multicultural nature of it. I met a ton of friends there. A lot of interesting people got together. It was where I first got exposed to design. They were mostly working on physical objects and augmented and virtual reality application, how you can use technology as a means of communication. I never saw it that way before, but that is very important.

This CVPR workshop was the first one. Where do you expect this work to go in the future?

This is a very applied problem. I hope more advertising companies will get

interested in engaging in the academic community. This is a way to meet us all at once. The few of us that were there have been keeping in contact, and that's been really useful. We're trying to figure out who is in this community and having a list of names. I don't think we all know each other. It was actually quite random that [Adriana Kovashka](#) and I got connected. She knew someone at Columbia who does a lot of video processing related stuff. He happened to know my work. It was really coincidental how we found each other at all. With these coincidences, you find a couple then you find a couple more. It's really a huge problem. There are computer vision problems. There are human communication problems. There are some ethical issues as well like which kind of images should we be showing people. There are business-related questions.

What ethical related questions?

In advertising, you associate your product with some image, usually what your customers want. There are many kinds of advertising. In the 1950s, advertising was only about the product. *"We made a microwave. Buy it or don't buy it!"* You advertised exactly what the product was. Now, we're in an age where you're selling a lifestyle. You're putting a message out there. You want people to believe it. If you have really powerful techniques for that, you have to consider how do we really want to choose to influence people? It's hard to say this for certain, but I've heard a lot of arguments that say that the emerging desktop computer industry in the 80s advertised to boys rather than girls. That had a huge effect. Then they advertised gaming. That really is the first application. That has a huge effect on society. In advertising, you must find

your market segment. You can't advertise to everyone at once. These things have effects. I'm hoping with computer vision, we don't have to segment markets anymore. We can find out why it's useful to everyone and find the right customers. That's the idea in advertising, to find the people that are right for it, and give them the information that they want.

What should our magazine Computer Vision News do in that respect? Should we also use visual metaphors?

Good question! The first thing you want to think about is who is your audience. You don't always need a visual metaphor. If you want to convey more of the details in the mechanisms of how an algorithm really works, you want more of a visual explanation. If you wanted to associate something concrete with something more abstract as a way to draw people into the message, maybe use the fact that there are interviews about computer science.

And computer scientists! The human aspect which is behind the technology!

I think it's that idea about getting to know someone. That's one thing about academia: we get to know people's work and we invite them for talks. I'm kind of surprised, in a good way, to see the publication that you made. I've never seen anything like it before in academia. I always enjoy getting to know the person and seeing what motivates them. You never put that in the paper. No one ever says: *"Well, I was walking around one day, and I saw this visual metaphor! Well, that's kind of cool! How'd they do that?"* We never do that!

One more message for our readers?

Yes, I'm always looking for interesting students!



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.

ACCV 2018 - Asian Conference on Computer Vision

Perth, Australia Dec 2-6 [Website and Registration](#)

AI World - Accelerating Innovation in the Enterprise

Boston, MA **MEET US!** Dec 3-5 [Website and Registration](#)

NIPS 2018 - Neural Information Processing Systems

Montreal, Canada Dec 3-8 [Website and Registration](#)

AI Summit (practical implications of AI for the enterprise)

New-York, NY Dec 5-6 [Website and Registration](#)

CloudCom 2018 - Cloud Computing Technology & Science

Nicosia, Cyprus Dec 10-13 [Website and Registration](#)

IEEE Big Data 2018 - IEEE International Conference on Big Data

Seattle, WA Dec 10-13 [Website and Registration](#)

WACV: IEEE Winter Conference on Applications of Computer Vision

Waikoloa Village, Hawaii Jan 7-11 [Website and Registration](#)

RE•WORK Deep Learning Summit

S.Franisco, CA **MEET US!** Jan 24-25 [Website and Registration](#)

SPIE Medical Imaging





San Diego, CA Feb 16-21 [Website and Registration](#)

BMVA: British Machine Vision Association - Deep Learning in 3D

London, UK Feb 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 30** 

FEEDBACK

Dear reader,

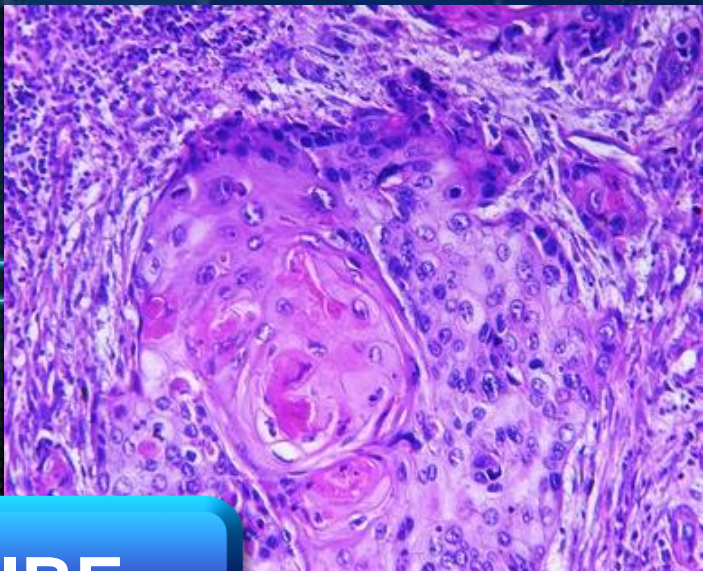
How do you like Computer Vision News? Did you enjoy reading it? Give us feedback here:

[Give us feedback, please \(click here\)](#)

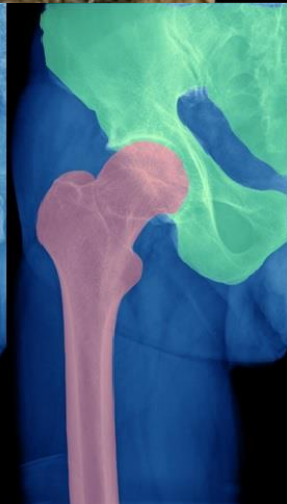
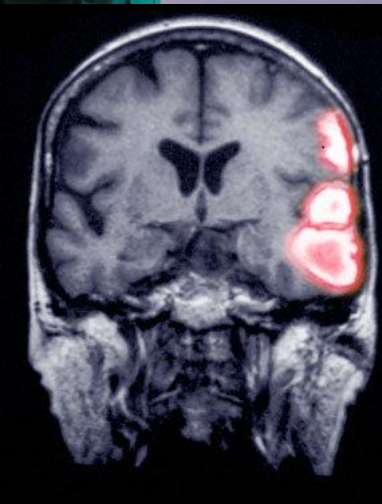
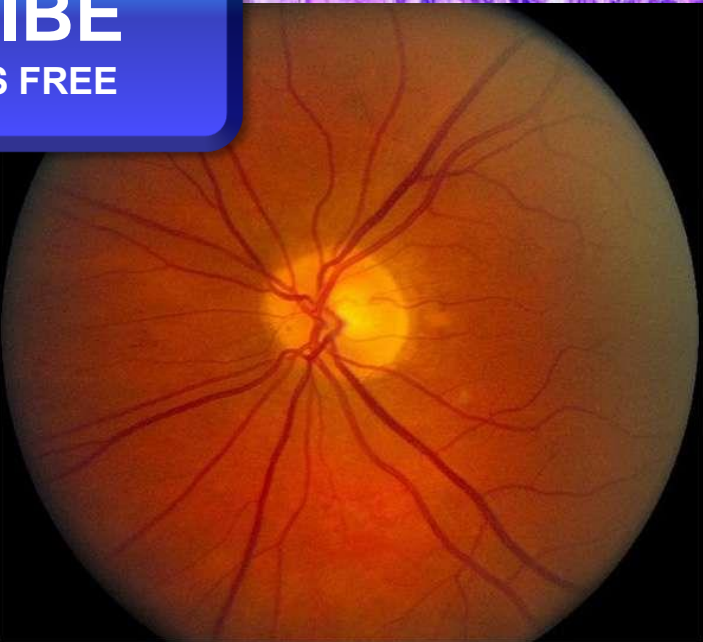
It will take you only 2 minutes to fill and it will help us give the computer vision community the great magazine it deserves!

IMPROVE YOUR VISION WITH Computer Vision News

The Magazine Of The Algorithm Community



SUBSCRIBE
CLICK HERE, IT'S FREE



A PUBLICATION BY



Global Leader in Computer
Vision and Deep Learning