# Computer Vision News

The magazine of the algorithm community

A publication by RSIP Vision

## June 2019

*We Tried for You:*
**t-SNE**

*Women in Science:*
**Lena Maier-Hein**

*Artificial Intelligence Spotlight News*


**Application: Shone - Autonomous shipping**

*Upcoming Events*

*Project Management:*
**Being Creative with Deep Learning**

*Challenge and Workshop at CVPR2019:*
**Long-Term Visual Localization under...**

*Computer Vision Project:*
**Lung Tumor Segmentation and Classification**

*2 Research Reviews:*
- **BA-Net: Dense Bundle Adjustment Network**
- **Multi-Resolution Networks for Semantic Segmentation in ...**

# Contents

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

**Computer Vision News**

Editor:
**Ralph Anzarouth**

Engineering Editor:
**Amnon Geifman**

Publisher:
**RSIP Vision**

Contact us

Give us feedback

Free subscription

Read previous magazines



**Follow us:**

Dear reader,

**CVPR 2019** is just around the corner! The entire community will meet in **Long Beach, CA** starting on **June 17**. Of course, we will be there to publish the **CVPR Daily**, the official magazine of CVPR, and to meet clients and friends. Not coming? **Receive the CVPR Daily every day in real time and feel at CVPR as if your were at CVPR**.

**The Bay Vision Meetup group**, which we sponsor, hosts a monthly meeting on the different fields covered by **Computer Vision, Deep Learning and Artificial Intelligence**. If you work in the Bay Area or if you happen to travel there, come over and join the group. **The pizza is on us!** The next meeting on **June 6** will focus on **AI in Medical Imaging**. **And here is the full recording (with all the slides)** of the webinar hosted just a few days ago by Moshe Safran and Miki Haimovich: **"In a Heartbeat - Implementing AI in Cardiology**".

Finally, a warm welcome goes out to the many new readers that joined us this month for the first time. Whether you represent one of the hundreds of **new subscribers** or you received Computer Vision News from a friend, this magazine is written for you by algorithm professionals like you. As a leading company in the AI world, **RSIP Vision** is proud to offer this free service to the community, every month for the fourth year.

**See you at CVPR and enjoy the reading!**

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

by Amnon Geifman

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen **BA-Net: Dense Bundle Adjustment Network**. We are indebted to the authors (**Chengzhou Tang, Ping Tan**) for allowing us to use their images. The paper was presented at **ICLR 2019** a few weeks ago and is found here.

## BA net- Bundle Adjustment Network

**Structure from motion** is one of the fundamental branches in computer vision. Surprisingly, **classical solutions** still outperform **deep learning solutions** in many structure from motion tasks. One and very important example of such task is **bundle adjustment (BA)**, which is a highly non-convex optimization. The challenges in BA are the sensitivity to initial correspondences, non-linearity of the objective, sensitivity to outliers and more. In the BA-net paper, which was presented at **ICLR 2019**, the authors suggest a novel deep learning architecture that aims to solve the bundle adjustment problem. It uses BA as a differentiable layer in the network. The main novelty in the paper is to **learn a feed-forward network** to predict the damping factor of the Levenberg-Marquardt (LM) algorithm which makes the whole pipeline differentiable.



## Bundle adjustment

To better understand the novelty of the paper, we first explain the conventional bundle adjustment. Generally, almost every structure from motion algorithm uses bundle adjustment to refine its final solution. Given a set of N images $I_i$, initial guesses for N camera matrices $T_i$ and an initial guess for a set of K 3-D points $p_j$. The geometric bundle adjustment minimizes the sum of reprojection errors of the form:

$$\sum_{i=1}^{N}\sum_{j} ||\pi(T_i, p_j) - q_{i,j}||$$

Where $q_{i,j}$ is the 2-D point in the i'th image, corresponding to the 3-D point $p_j$. $\pi(T,\cdot)$ denotes the projection operator of a 3-D point into the image plane, using camera matrix T. The optimization variables are the camera matrices $T_1,..,T_N$ and the 3-D points $p_1,..,p_K$. This optimization is usually called ***geometric BA***.

On the other hand, sensitivity to outliers and texture has motivated the emergence of **photometric bundle adjustment**. In contrast to geometric, the photometric BA relies on maximizing the photometric (pixel intensity value) consistency and estimates the correspondences implicitly. Given a set of camera matrices T and a set of K points (in $I_1$ pixel coordinates) q. The photometric error is then defined as:

$$e_{i,j}(T,d,q) = I_i\left(\pi\left(T_i, d_j q_j\right)\right) - I_1\left(q_j\right)$$

Where $d_j$ is the depth of the point $q_j$, hence $d_j q_j$ upgrades the pixel $q_j$ to its 3D coordinates. The photometric bundle adjustment minimizes the sum of the photometric reprojection errors. With this intuition in mind, the authors define a differential pipeline that enables to exploit the power of deep learning. We next explain the method.

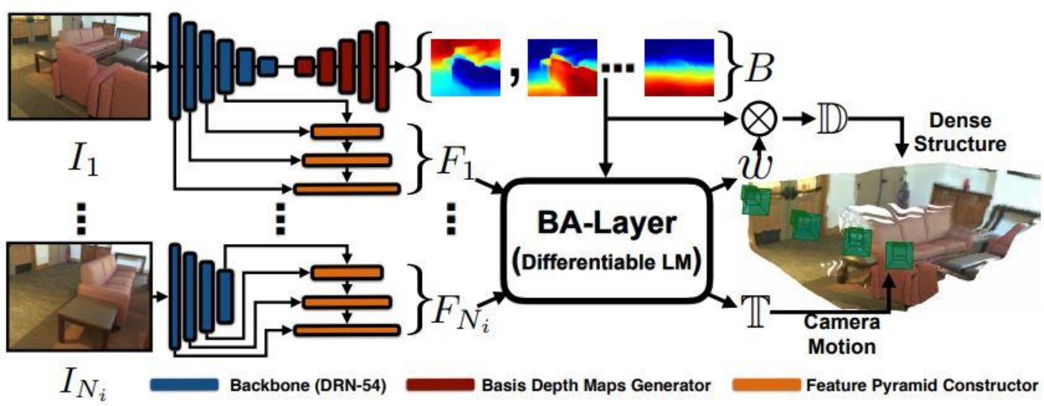
## Understanding the model

The authors suggest **a slight change to the photometric BA**. Instead of minimizing the photometric error, they offer to minimize the difference (error) of features related to a specific pixel, i.e. feature-metric difference of aligned pixel, that is:

$$e_{i,j}(T,d,q) = F_i\left(\pi\left(T_i, d_j q_j\right)\right) - F_1\left(q_j\right)$$

Where in this error term, everything stays the same as before, except that now $F_i(q_j)$ denotes a (learnable) feature pyramid i.e. a feature vector across multiple scales corresponds to the $q_j$ pixel in the i'th image. The minimization is over the camera parameters and depths in the sum $\sum e_{i,j}(T,d,q)$

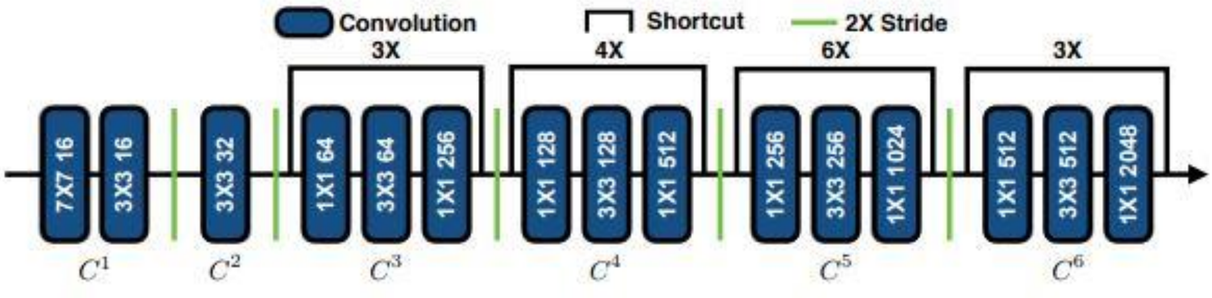





The above figure best explains the method. The input to the pipeline is a sequence of images. Then, every image is transferred through a DRN-54 network (see architecture in the next figure) to learn a feature pyramid. At the same time, a convolutional network generates multiple basis depth maps

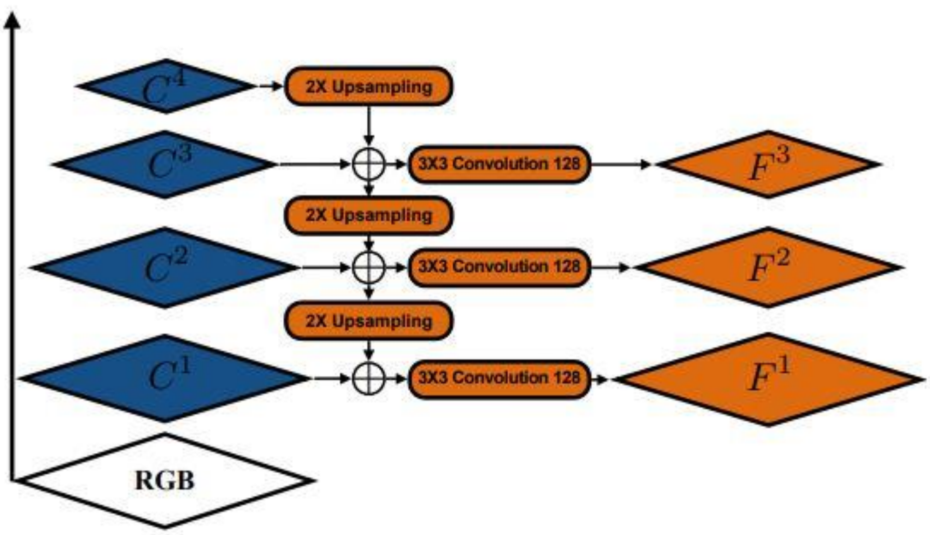
**A novel architecture that enables end-to-end differentiable bundle adjustment**

for the first image. Lastly, the pyramid and the depth maps are inserted into a BA-layer that outputs (via the LM algorithm) a dense structure and camera motion parameters via minimizing the feature metric error. We next review each of the steps separately.




(a) Architecture of the DRN-54 backbone

Feature pyramid: instead of using the pixel intensity value, it might be more robust to use a feature that represents the pixel. The paper exploits **the natural multi-scale of deep convolutional networks**, in order to construct a feature pyramid. It does that by using the intermediate layers from the backbone DRN-54 architecture. Using the residual block of convolution 1 up to 4, they up-sample a feature map from each layer to create a feature map for the next level. At each sampling, 3x3 convolution is used to concatenate features and reduce the dimension of each level to 128. The final feature pyramid is then $F_i = [F_i^1, F_i^2, F_i^3]$. It is fundamentally a function of the image. This process is best explained through the following figure:




(a) Feature pyramid construction

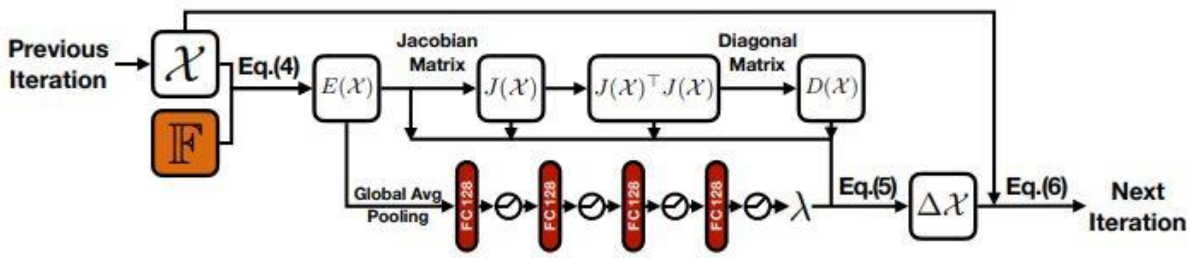
Basis of depth maps: At the same time of the feature extraction, the basis depth maps is generated. To cope with the infeasibility of parameterizing directly a dense depth map, the authors **use an encoder-decoder scheme to generate a monocular depth learning**. They use the DRN-54 network computed in the feature pyramid stage as the encoder. For the decoder, they modify the last convolutional layer of Laina et al. (2016) to generate the basis depth maps, denoted by B. The final depth map is generated as a linear combination of these basis depth maps. It gives the desired depth estimation as $d_j$=ReLU($w^T$B[j]) where w is learnable weights (learned at real-time) and B[j] is the j'th column of the feature maps generated from the encoder decoder scheme.

Bundle Adjustment Layer: the main advantage of the BA layer lies in the update of the Levenberg-Marquardt algorithm. The LM algorithm is used to iteratively optimize non-linear least squares problems of the form $\sum(y_i - f(x_i, \chi))^2$ where $\chi$ is the optimization variable. Using the Jacobian matrix J, LM defines the update rule by the equation:

$$\left(J^T J - \lambda\, diag(J^T J)\right)\Delta\chi = J^T\left(y - f(\chi)\right)$$

The **main challenge** here is that in the LM algorithm, the damping factor $\lambda$ is determined by a non-smooth thresholding. This makes the update non-differentiable.

To cope with the challenge of determining the damping factor, the author uses a network to predict the optimal value of $\lambda$. This makes the update rule differentiable and allows to minimize over this parameter. This network gets as input the feature pyramid F, camera parameters T and depth values d (computed by the depth maps) from the previous iteration and returns the damping factor (or equivalently the current update). It is trained in a supervised manner using the ground truth camera parameters. The following figure explains this process:




The three components described above give a differentiable pipeline that can be trained from end to end using **back-propagation**.

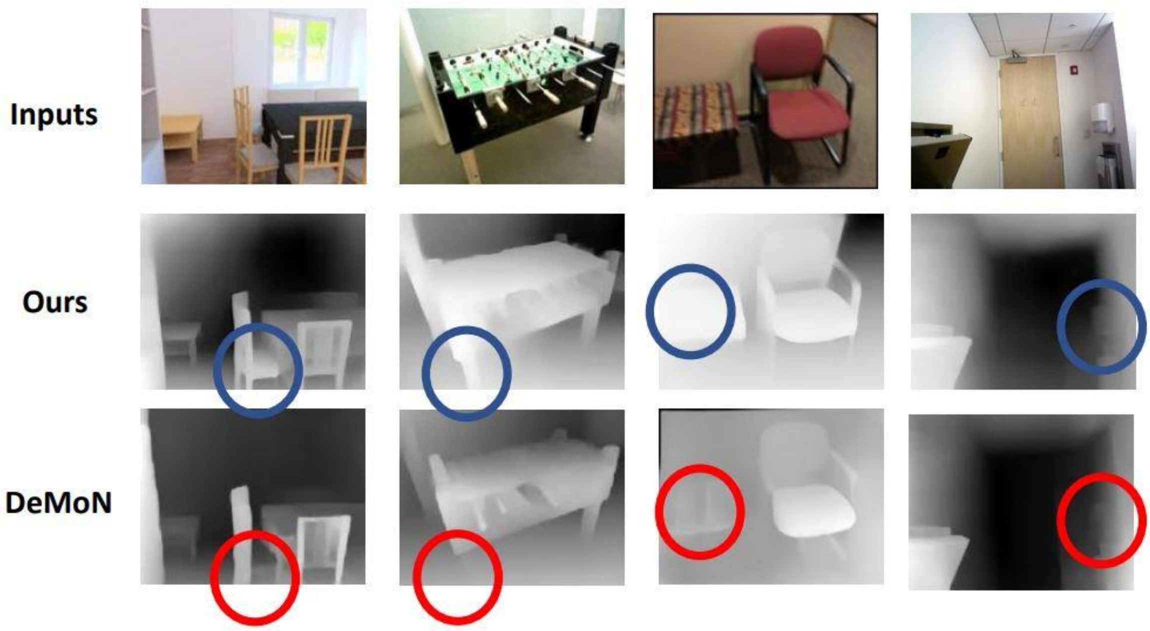
## Results

To demonstrate the effectiveness of their method, the authors compared their method to **DeMoN** which is a depth and motion network for learning monocular stereo, as well as to conventional BA. You can appreciate their results by looking at the following table:

| | Ours | Ours* | DeMoN* | Photometric BA | Geometric BA |
|---|---|---|---|---|---|
| Rotation (degree) | **1.018** | 1.587 | 3.791 | 4.409 | 8.56 |
| Translation (cm) | **3.39** | 10.81 | 15.5 | 21.40 | 36.995 |
| Translation (degree) | **20.577** | 31.005 | 31.626 | 34.36 | 39.392 |
| abs relative difference | **0.161** | 0.238 | 0.231 | 0.268 | 0.382 |
| sqr relative difference | **0.092** | 0.176 | 0.520 | 0.427 | 1.163 |
| RMSE (linear) | **0.346** | 0.488 | 0.761 | 0.788 | 0.876 |
| RMSE (log) | **0.214** | 0.279 | 0.289 | 0.330 | 0.366 |
| RMSE (log, scale inv.) | **0.184** | 0.276 | 0.284 | 0.323 | 0.357 |

As you can see, **BA-net outperforms all the other methods**. They show a significant advantage in camera rotation and translation estimation (first 3 rows) as well as in the root mean square error of the depth values.

You can also see below some **qualitative results** on the ScanNet dataset that show the recovered depth maps by the paper's method and the DeMoN method.



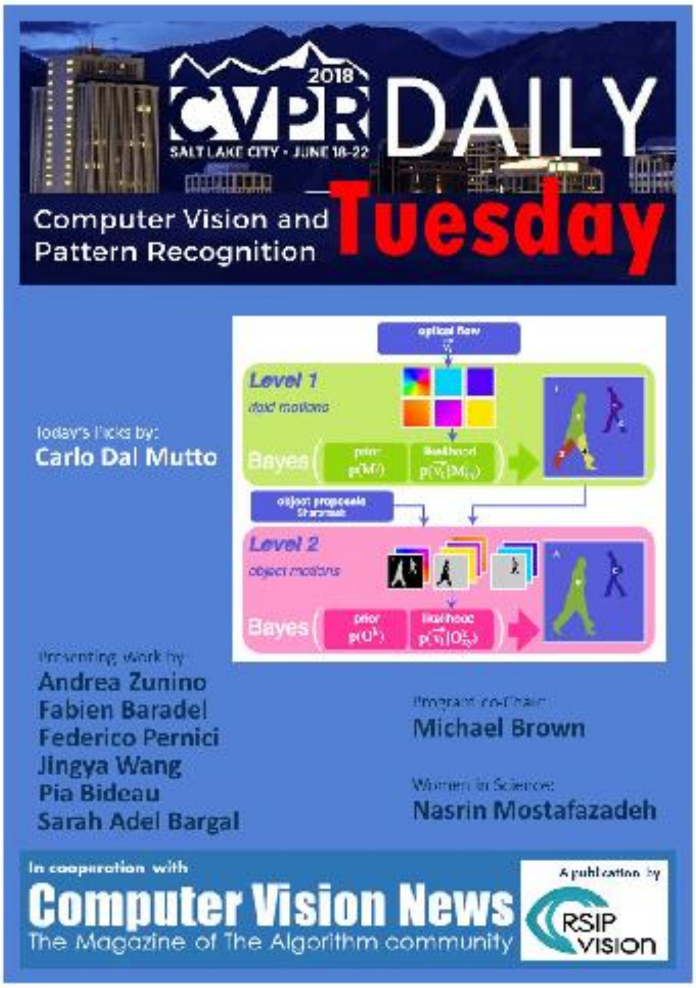

The circles demonstrate that the BA-net method tend to recover more shape details. Moreover, the results seem sharper and it looks like the BA net gives a better estimation for the depth map (at least in this specific example). Additional results are in the paper.

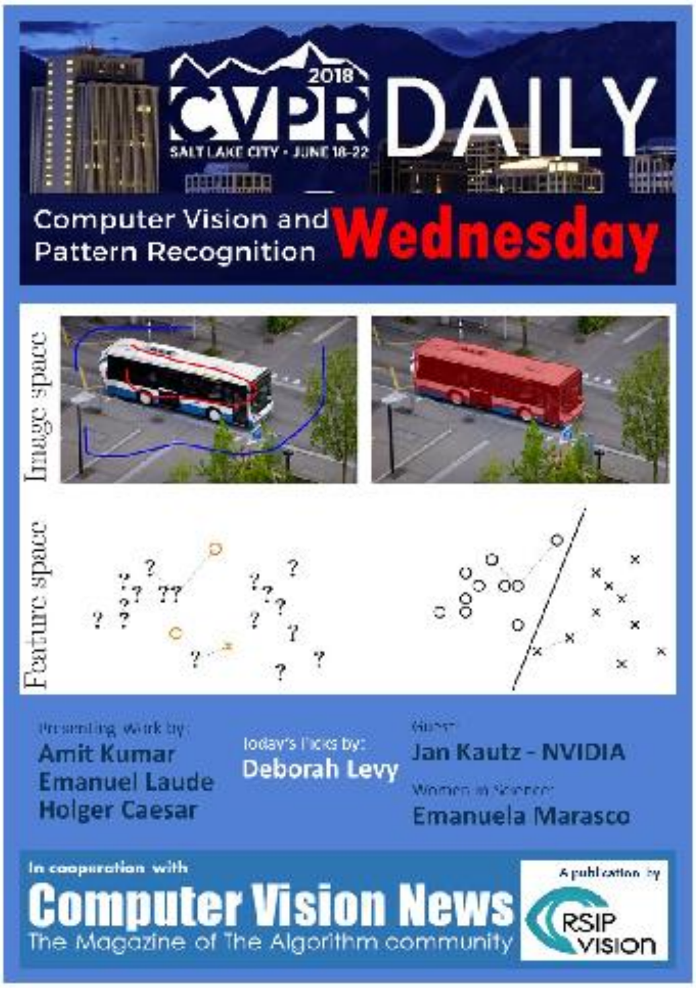
## Conclusion:

The paper presents a novel architecture that enables **end-to-end differentiable bundle adjustment**. It allows to deal with structure from motion tasks on images with exposure changes, moving objects, untextured images and more. The paper demonstrates state-of-the-art results on two data sets in both, camera motion estimation and depth estimation. There is no doubt that this paper is a promising step toward **solving a structure from motion problem using deep learning**. Although it shows some nice results and performance, let us remember that a robust, large scale and accurate deep learning solution is still waiting to be found. This is a hot area of research that will bring more news in the future.

# *Feel at CVPR as if you were at CVPR!*






**by Dorin Yael**

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen two. Here is **"Multi-Resolution Networks for Semantic Segmentation in Whole Slide Images"**. We are indebted to the authors (**Feng Gu, Nikolay Burlutskiy, Mats Andersson, and Lena Kajland Wilén at ContextVision**), for allowing us to use their images. Their paper is found here.

***"This approach yields superior results comparing to those gained using a single resolution by the classic U-Net"***

Analysis of **histological images** is one of the most commonly used tools in the diagnosis and research of a large variety of pathological conditions. Over the last few years, the field of digital pathology has undergone a significant progress thanks to the development of new scanning technologies alongside with the development in the field of computer vision. These newfound advances opened new powerful computational opportunities yielding higher performance levels in terms of accuracy and speed.

**Whole slide imaging (WSI)** is a technology allowing the acquisition of high-resolution digital images representing entire tissue slices scanned from glass slides. WSI contains multi-resolution information organized in a pyramid structure, which allows spatial navigation along multiple magnification levels. This **multi-resolution information** is used by pathologists and researchers to characterize tissues in levels ranging from sub-cell to multi cell complexes. In some cases, the ability to **examine the tissue in multiple resolutions** is crucial. For example, in cancer diagnosis, both local information such as regularity of cell shapes and cell density as well as contextual information such as global tissue structure, are highly important for achieving an accurate evaluation.

The digitation of histopathological images enables the development of automatic tools assisting in their analysis. Deep neural networks, and specifically **Convolutional Neural Networks (CNN)** became a gold standard in multiple image processing tasks. Due to their large size, WSIs are commonly cropped into small patches and in some cases also down sampled during the learning and prediction processes. These procedures lead to the loss of either contextual or local information and may damage the learning capability and prediction accuracy in turn. To address this issue, Feng Gu et al. have designed **a U-net based multi-resolution network (MRN)** allowing the use of multiple resolutions during the learning and prediction processes.

The classic U-Net architecture, widely used for the segmentation of histological images, consists of an "encoder" in which the feature maps are down-sampled, a "decoder" up-sampling them back and skip-connections concatenating corresponding layers' feature maps in the up-sampling pathway (Figure 1):

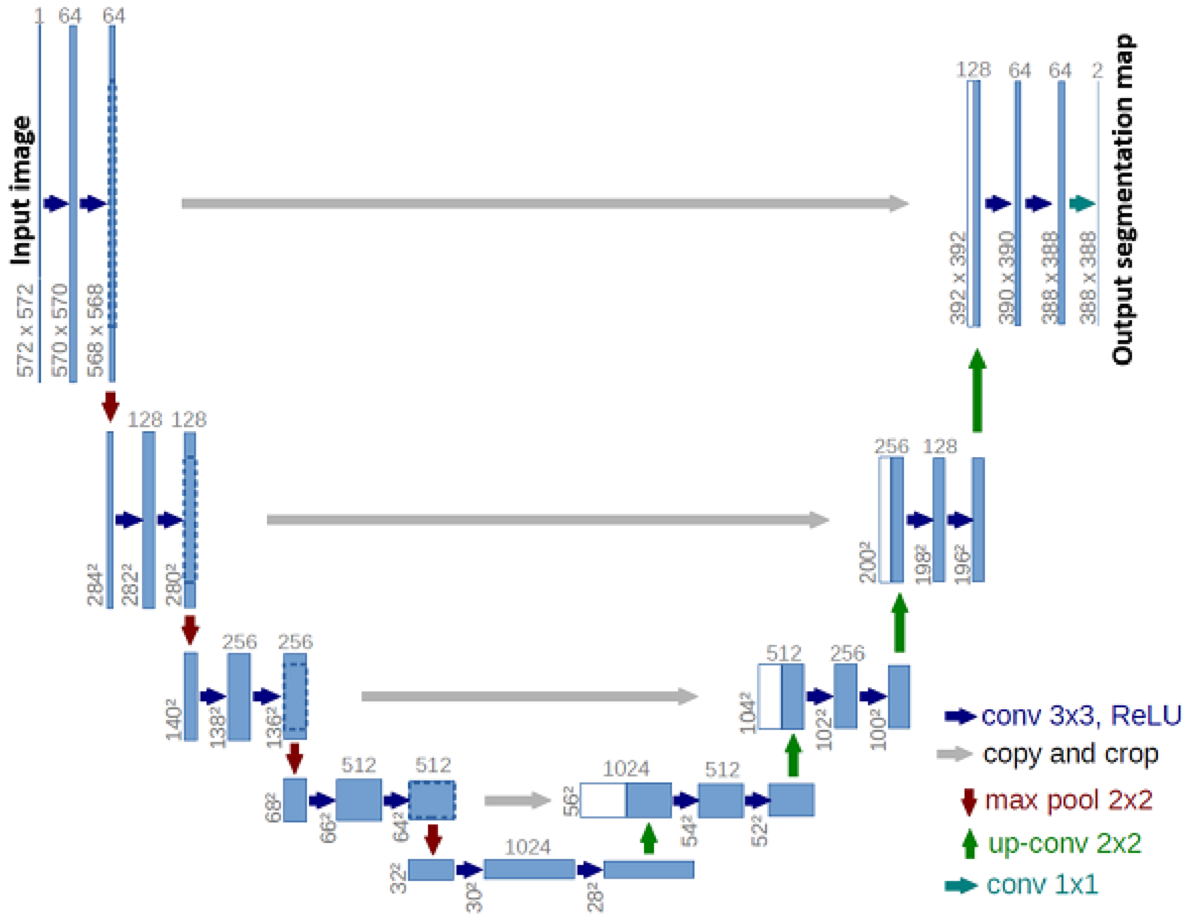



*Figure 1: U-net architecture (source: Ronneberger et. al, 2015). Right branch- encoder, left branch- decoder. Each blue box corresponds to a multi-channel feature map, white boxes represent copied feature maps.*

The proposed MRN is based on the classic U-Net architecture with additional encoders corresponding to the different resolutions (Figure 2 on the next page). The different input resolutions share central coordinates and cover tissue area in a pyramid-like structure (Figure 3 on the next page). The input shapes of all resolutions are identical and processed using identically structured encoders. To preserve a relevant information for the region of interest, the lower resolution is center cropped and upsampled to the original resolution. The outputs are concatenated with the high resolution convoluted feature maps and are passed through 1X1 convolution layer with an identity activation function. This process allows **a weighted summation of the multi resolution feature maps into a single feature map** which is concatenated in turn to the corresponding layer in the decoder. This architecture allows the use of peripheral low-resolution data during a pixel-wise segmentation of high-resolution areas. Similarly to the classic U-Net, this is done using a single network, relatively small number of parameters and a single loss function. **This approach yields superior results comparing to those gained using a single resolution by the classic U-Net.**

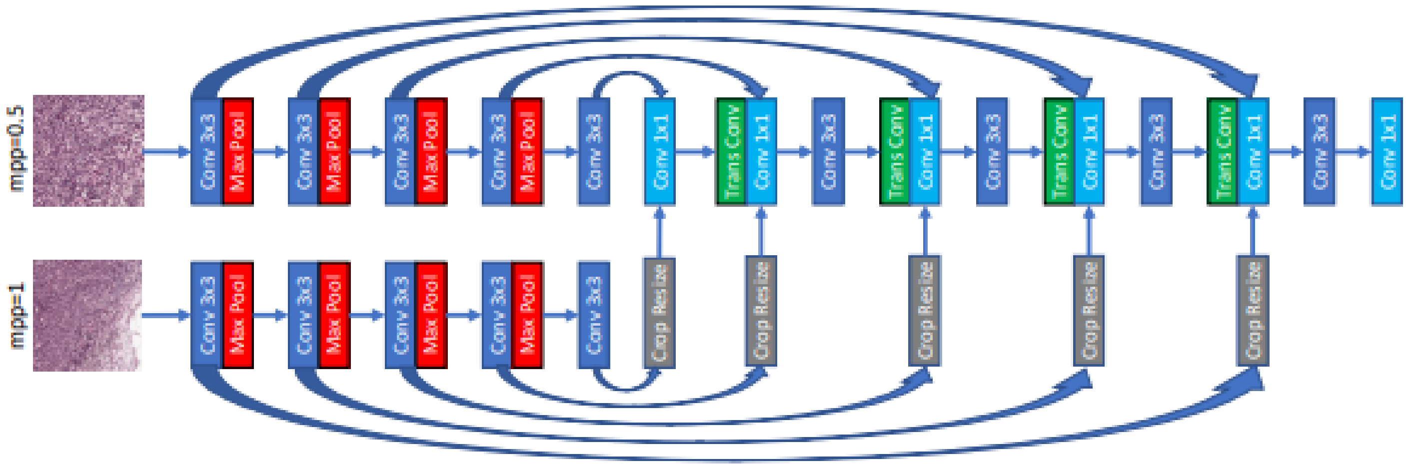



*Figure 2: An illustration of the proposed MRN for two resolutions. Dark blue boxes: stacks of two convolution layers with ReLU activations. Red boxes: max-pooling layers. Light blue boxes: convolution layers with identity activations. Green boxes: transposed convolution layers with ReLU activations.*

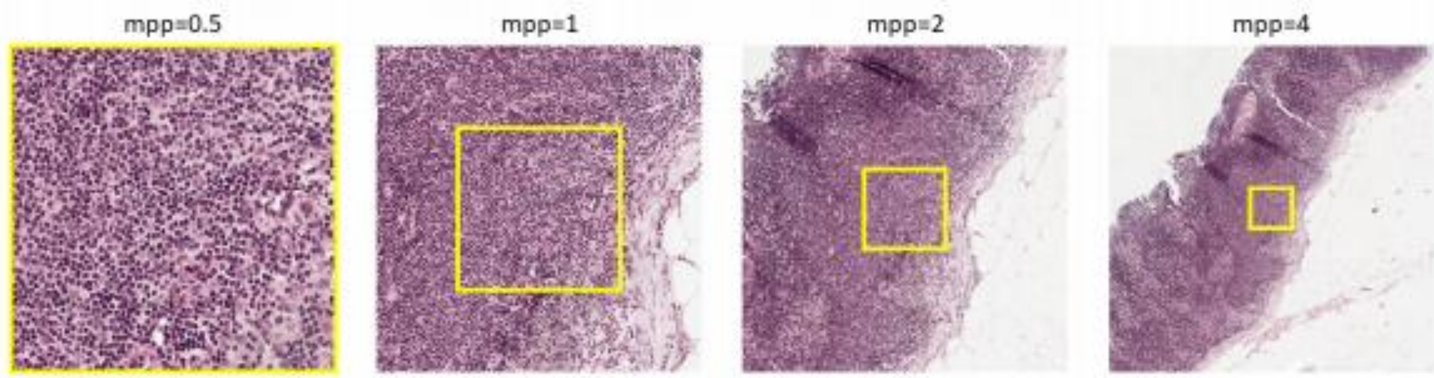



*Figure 3: An illustration of patches with the same central coordinates. Increased mpp values corresponds to zooming out action to enlarge the field of view. Yellow squares represent the effective tissue area at different magnifications.*



In a more general scope, the question regarding the balance between global and local information in visual perception is extensively studied by psychologists and neuroscientists since the rising of the gestalt psychology. In line with this, multiple deep learning researchers using networks which mimick the visual system have addressed this issue and designed **networks enabling the processing of images using multi-scale data**. The U-Net based MRN architecture allows the processing of local information using its global context while remaining the number of weights relatively compact. The small number of weights, which increases linearly with number of resolutions fed to the network, enables the use of a small dataset. In one of our histological image segmentation projects at **RSIP Vision**, we were inspired by this architecture and expanded the standard U-Net to process **multiple resolutions**. This approach was simple to implement and **increased the network's prediction accuracy**.

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**

Management

# Being Creative with Deep Learning

**RSIP Vision's CEO Ron Soferman** has launched a series of lectures to provide a robust yet simple overview of how to ensure that computer vision projects respect goals, budget and deadlines. This month **Yael Zak** tells us about another aspect of **Project Management with Deep Learning: Being Creative with Deep Learning**. Read more tips by RSIP Vision about Project Management in Computer Vision.

You can personally assist to **Ron Soferman's lectures**: he will talk about **Challenges in Project Management for Medical AI** at our next **Meetup** on **June 6** in **Cupertino, CA.** Register here**.**

*"... improve the network's architecture by exploiting the characteristics of the problem itself!"*

This month's **project management tip** regards the most attractive aspect of **deep learning**, since it involves the capacity of **being creative**. A creative approach might, in certain situations, become an asset. This happens when **the engineer can improve the network's architecture by exploiting the characteristics of the problem itself**.

At the beginning of a project, **choosing the network architecture is kind of straightforward**: given the specific problem which needs to be solved (segmentation, pose estimation, classification and so on), you just pick the most advanced or **the most commonly used architecture for that problem** and this would provide a reasonable solution.

However, it is not so rare that **a better and more satisfying solution can be found**. Maybe your project requires a lower error rate; maybe the network is too heavy and inference time should be improved; maybe the standard solution is costly in terms of resources (memory, time and/or GPU).

At this point, the **project engineer** has an opportunity to consider: do we know something about this specific project that narrows it down? For example, we need to identify objects: do we know in advance the expected number of objects, their approximate location or the slice on which they will be found?

Perhaps some **modifications of the architecture** could make use of previous knowledge we have on the problem. If we know in advance the number of organs that need to be identified by our network, for example, we could adjust the number of outputs; this way, our network is thinner comparing with a network aimed at finding an unlimited number of objects. This might reduce the cost of network inference, and allow the network to produce **more accurate results**.

Another example is in case that the network is supposed to, or could, answer **few questions at once**. It is possible to add suitable tracks to the architecture for the different questions, and sometimes even feed the network with its own answer for one question, as additional input to calculating output for another, harder question.

**This aspect in deep learning engineering is also fun**, because this is really the space where your perception of the problem, and translating it into concrete architecture characteristics, could have **a positive impact on the quality of your engineering**.

More articles on Project Management



**"This aspect in deep learning engineering is also fun…**

**by Michal Margalit**

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 challenges and solutions in a Medical Imaging project by **RSIP Vision**: **Challenges in Lung Tumor Segmentation and Classification**.

**Computerized Tomography (CT)** of the lungs is a common procedure to assess existence and character of **pulmonary lesions**, including nodules and masses: Although some cases are benign, it is crucial to detect the **cancerous** ones correctly in order to achieve clear and early diagnosis.

The order of nodule classification is the following:

first, according to their **transparency density** - some of them are solid and occlude all tissues behind them - others are almost transparent and their look similar to hazy glass is called **Ground-Glass Opacity (GGO)**. In between we find also semi-solid nodules which combine these two features: these have the highest chances to be **malignant tumors**.

Second, according to the smoothness of their borders - the most problematic nodules being those with **spiculated borders** that have protruding spikes like sun rays.

Nodules differ very much one from another, both in shape and density ; in some cases they might have non-solid/ ground-glass appearance. The classification of the nodules into solid, semi-solid or non solid appearance aids the physician in assessing the **probability of malignancy**.

It is not only the appearance of the nodule that counts, but also **the shape of its borders**. Smooth borders increase the chance of it being benign, as apposed to irregular, spiculated or lobulated borders, that increase the chance of malignancy. **Volume** also is key, as larger nodules and masses are more suspect than smaller ones.

Besides their large diversity, the **algorithm developer** encounters another challenge at the location of the lesion: they can be found anywhere within the lung and even touching the hilum or mediastinum, making it difficult even to **expert radiologists** to recognize where the boundaries are and whether the nodule belongs to the airway system or not.

Other findings may appear like pulmonary nodules while they are not. This might be tricky, making room for false positive findings. Possible misleading findings include blood vessels, atelectasis or fluid accumulation or scar tissue.

***"This was one of the main tricks that we used"***

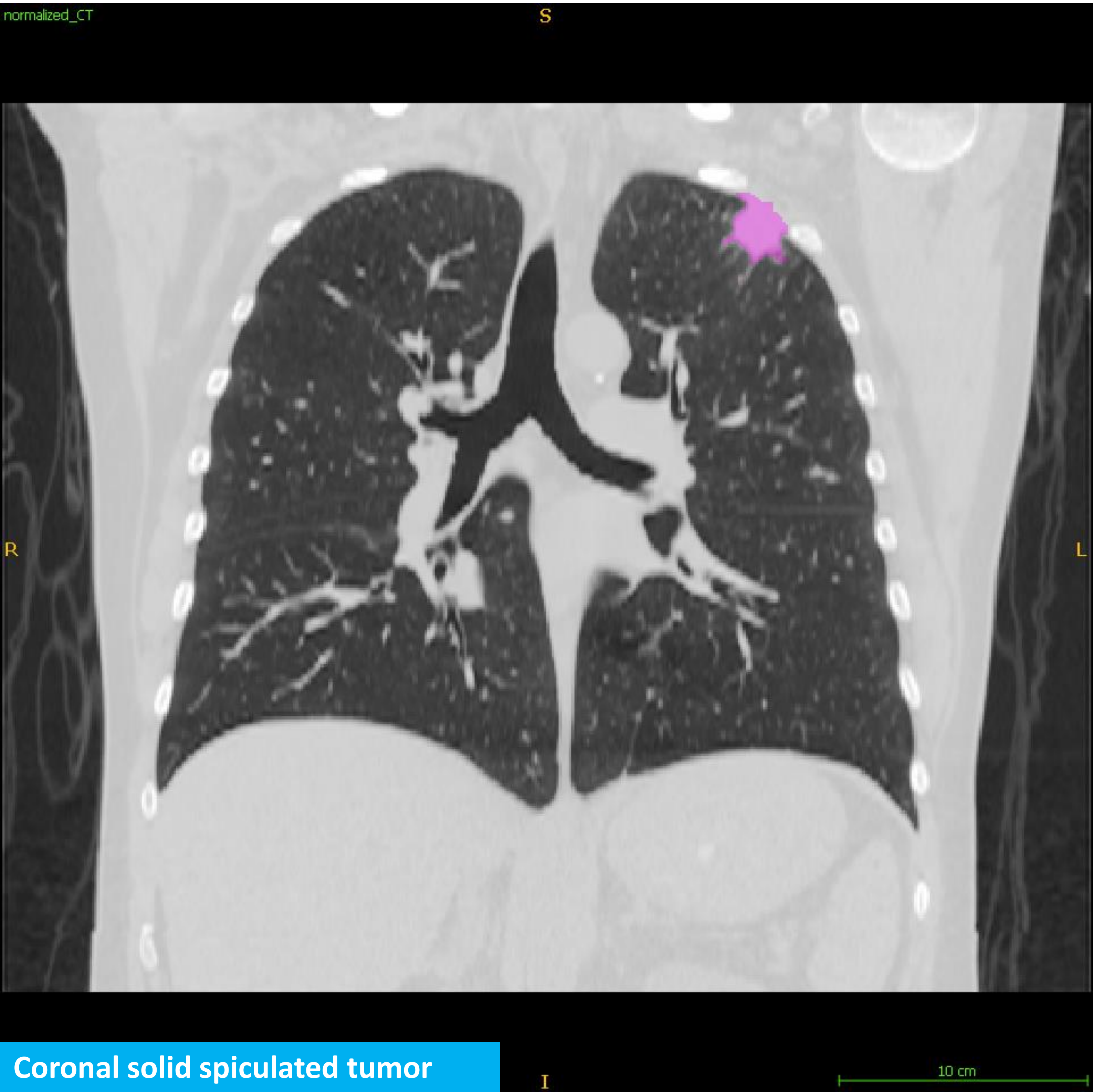
**We facilitated the algorithm** by annotating with a different color the accumulations of fluid inside the lungs. While there are problematic spots, they are not nodules and therefore they do not interest our study. We also classified the different appearances of lesions by color - red is solid, green is combined, blue is GGO. **This was one of the main tricks that we used to power the algorithm and minimize the false positives rate**.

**False positives** are problematic because they create an alert when there shouldn't be one. This and the **deep learning technique** enable us to provide our client the most accurate results. Read more about this software module on RSIP Vision's website.

Read about more projects in lung segmentation

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

Project




Coronal solid spiculated tumor


Ugo Vollmer

**Ugo Vollmer is the Co-founder & CEO at Shone, a company that retrofits manned cargo ships with autonomous technologies. He shares more about how their company uses artificial intelligence to improve the safety, security, and quality of life of the men and women who sail the seas.**



It's no secret that autonomy is happening everywhere, from autonomous cars to self-driving trucks. But the founders of Shone took their attention from the roads and set their sights on the sea.

**Ugo Vollmer** and his team recognize that autonomy will inevitably happen with **cargo ships**. Their question: how and when it will happen.

Others in the industry predicted that **autonomous shipping** would happen through new, unmanned electric cargo ships. At Shone, they didn't consider this a realistic way to tackle the problem. Instead, they took **another approach which brings autonomy progressively** on existing vessels through retrofitting without removing a single crew member on board.

A cargo ship has three main functions: **navigation, engine room maintenance, and cargo landing**. Shone primarily focuses on navigation because this can lead to accidents and inefficiencies. They have been working toward autonomy progressively primarily with **CMA CGM**, one of the biggest shipping lines in the world to develop technology on their operational vessels.

The seafarer has one of the most hazardous jobs on earth. Cargo shipping faces **dangerous storms with surging waves** and violent winds reaching 160 miles per hour. Political unrest and pirate attacks can also put cargo at risk.

*"Even in 2019, you still have **ships that are crashing right into an island when people are not standing their watch**. Their GPS is following a track that is going right into an island. This is obviously something that should not happen. There are a bunch of cases that nobody hears about because it's not happening in your national water. This happens pretty often, and the **consequences are disastrous** in many aspects, including on the environment"* Ugo shares.

Shone seeks to use autonomous technology to mitigate these dangers and increase the safety of cargo shipping.

Their first product in development is **the smart co-pilot** which helps the crew on board navigate safely and efficiently. One of the functions supports watch keeping and provides technology such as a camera with 360 views and 24/7 monitoring. This functionality proves **more reliable than humans**, who have limited visibility and may fall asleep or get sick while on watch.

The crew of a cargo ship can also have difficulties navigating in crowded waterways. Shone is developing **sensor fusion** between different data sources that improves consistency, helps prioritize dangerous targets, and provides a recommendation of what the crew should do in terms of safety.

Cargo ships face similar types of problems as autonomous driving for cars, but with a totally different set of constraints. On the technical side and in terms of retrofitting cargo ships, the existing technology is much different on cargo ships than in cars. Whereas a car only needs to see 15 meters (about 50 feet) ahead, **cargo ships need to see 500 meters (1,640 feet) away**. Seeing 15 meters away proves completely useless for a ship because, at that point, **the ship cannot avoid an accident**.

Long distances involve a lot of estimation and require the use of different types of cameras. On the algorithm side, they face the same challenges. As Ugo explains: *"I guess the difference is seeing not far away at a very high FpS (frames per second) versus seeing really far away at a lower Fps; because you have so much inertia, but you have also a lot of real time issues in the sense that* ***you need to detect from really far away****. Otherwise, you won't be able to prevent it."*

On the business side, autonomous shipping faces many different challenges than self-driving cars, primarily, because it doesn't involve the same type of vehicle. Meanwhile, on the control side, self-driving cars operate on roads, which do not move. On their side, autonomous shipping must navigate on moving waterways. **Autonomous shipping does share similarities with self-driving cars in terms of the algorithms**.

*"The algorithms are pretty much the same as self-driving cars. You need object detection, classification tracking, segmentation to know what's water and what's not water. We use neural nets which are state of the art to do that.*

*I guess it's not the same type of constraints of real time and FpS. You can use different models than ones that run in real-time that you could use for car applications. Again, there are some differences there. We usually use a mix of a **deep learning** type of approach also with a traditional **computer vision** type of approach. This is slightly less complex, as we primarily need to distinguish what's not water, but I guess that's pretty much the same for self-driving cars."*

Shone develops proprietary detection algorithms based on data gathered from the ship sensors **AIS, a communications system** which people are trying to implement in the autonomous space. They are also working on an **augmented reality** type of application to empower the crew to navigate more safely in bad weather conditions, which will help during situations in which the crew cannot see an obstacle with the human eye. The technology analyzes an augmented reality object on top of the camera feed projected from a radar target, AIS target, or an element of the map such as buoys or the ship's pathway. With this technology, the crew gains a better sense of their surroundings.

Currently, Shone has nine employees based in San Francisco and three in Europe. For over a year, several CMA CGM ships have been equipped with the technology **predicting terabytes of data from those vessels**. The system continues to improve with each day. They have already started to deploy to more vessels with the aim toward autonomy. Along the way, they need to develop a lot of technology, some of which they will package and sell as a product.

by Amnon Geifman

*"A very powerful visualization tool"*

*"...try the model yourself using our code..."*

## Understanding and implementing the t-SNE visualization

When receiving new data to perform a learning task, the first challenge of the learner is to understand the underlying structure of the data. For example, given a set of images and their corresponding labels, we would like to explore the relation between the feature vector and the target: **is there a pattern in the data? Is this pattern being complex? Simple?** A good answer to these questions will make the learning task much more straightforward.

The challenge lies in human perception. **Most of us cannot imagine a high dimensional vector space**. Furthermore, visualizing high dimensional feature vector becomes impossible when the dimension is larger than three. To this end, we will review and implement for you a very powerful visualization tool **called t-SNE**, which enables us to embed **high dimensional data** in a two- or three-dimensional space. We begin by explaining what exactly is t-SNE, and continue by implementing a code that performs such visualization on MNIST dataset.



## Understanding t-SNE

t-SNE -- t distributed Stochastic Neighbor Embedding -- is an unsupervised technique to visualize high dimensional data. It was first introduced by **Laurens van der Maaten** and **Goeffrey Hinton** in the paper **Visualizing Data using t-SNE** (2008). To understand the impact of the original paper, it has gained over 7900 citations to date and it keeps growing.

The t-SNE algorithm, given a set of n points in high dimension $x_1, x_2, .., x_n$ seeks to find a low dimension representation $y_1,y_2,..,y_n$ such that the local and global geometry of the point is preserved. The embedding can be divided into 3 simples steps: first, by defining a pairwise probability distribution (similarity) between a pair of points $x_i,x_j$. This distribution is denoted by $p_{ij}$. Then, by defining a pairwise probability for the low dimensional points $y_i$ , $y_j$ denoted by $q_{ij}$. Lastly, by optimizing over the low dimensional representation to minimize the distance between the distribution P to the distribution Q. We shall now review each step separately.

**1) Similarity measure in the high dimensional space:** imagine a set of data points, where for each data point, we center a gaussian around it. Given a point $x_i$ we look at the density of the point $x_j$ given $x_i$. The idea is that if $x_i$ and $x_j$ are close, their conditional probability will be high; while if they are far apart, their probability will be low. To mathematically formulate this idea, a gaussian kernel is used to estimate the density. Given n data points in high dimension we write:

$$p_{j|i} = \frac{\exp\left(-\frac{\left|\left|x_i - x_j\right|\right|^2}{2\sigma_i^2}\right)}{\sum \exp\left(-\frac{\left|\left|x_i - x_k\right|\right|^2}{2\sigma_i^2}\right)}$$

where the summation in the denominator is over all data point except the i'th. We arbitrarily set $p_{i|i}$ to zero. The width of the gaussian is manipulated by what is called perplexity, which influences the variance of the distribution

**2) Similarity measure between low dimensional points:** as described in the previous section, we now measure similarity between the low dimensional data point; however, instead of using gaussian kernel, we now use student-t distribution with one degree of freedom (also called Cauchy distribution). This distribution is similar to normal distribution but its heavier tails allows dissimilar object to be located far apart. The resulting conditional distribution is:

$$q_{j|i} = \frac{\exp\left(-\left|\left|y_i - y_j\right|\right|^2\right)}{\sum \exp\left(-\left|\left|y_i - y_k\right|\right|^2\right)}$$

where again we define $q_{i|i}=0$ .

**3) Distance optimization:** let us remind ourselves that the $y_i$ are the unknown representation in the low dimensional space. So, our objective now is to find the low dimensional representation that enables the pairwise similarity measures in both spaces to be as close as possible. Since we are dealing with probability distributions, the immediate solution is to use **Kullback-Leibler (KL) divergence**. KL divergence is a measure for similarity between two probability distributions P,Q. It is asymmetric in nature, which explains why we call it a measure and not a metric. Given our distribution P,Q defined in the two previous sections, we optimize KL(P||Q) using gradient descent where our optimization variables are $y_1, .., y_n$

Note that we do not have a guarantee that the above embedding converges; however, as we will see next, in well-behaved data sets t-SNE demonstrates very nice results. Now that we understand what t-SNE does, we are ready to use it and **explore different hyperparameters and different visualization styles** of the model.

## Implementation

We implemented for you a code that embeds the **MNIST dataset** into a

2-dimensional space. In our implementation, we use Sklearn library and matplotlib. We begin by importing the necessary libraries using:

```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import offsetbox
from sklearn import (manifold, datasets, decomposition)
```

Next, we write a simple function that gets as input a low dimensional representation and labels and as outputs a plot of them. Below, X is the embedding and y is the target/labels of X. This function will look like this:

```
def plotEmbedding(X,y):
    X = (X - np.min(X, 0)) / (np.max(X, 0) - np.min(X, 0))
    plt.figure()
    for i in range(X.shape[0]):
        plt.scatter(X[i, 0], X[i, 1],
                    color=plt.cm.Set1(y[i] / 10.),
                    label=str(y[i]))
    plt.xticks([]), plt.yticks([])
    plt.legend(['0','1','2','3','4','5','6','7','8','9’])
    plt.title("t-SNE embedding")
    plt.show()
```

Now we are ready to begin **manipulating our data**. The dataset object of Sklearn contains a few datasets that are ready to use. It contains datasets such as Boston house prices, iris dataset, diabetes dataset, and digits dataset. We use the digit dataset since it is simple (i.e. can almost be linearly separable), so we can expect t-SNE to give us good results. Moreover, we will be able to make some cool visualizations using the digits. To load the 10 classes of the data set and defining the feature and the target we use:

```
Dataset = datasets.load_digits(n_class=10)
X = Dataset.data
y = Dataset.target
```

Now, we are ready to use t-SNE and embed our feature vector X into a

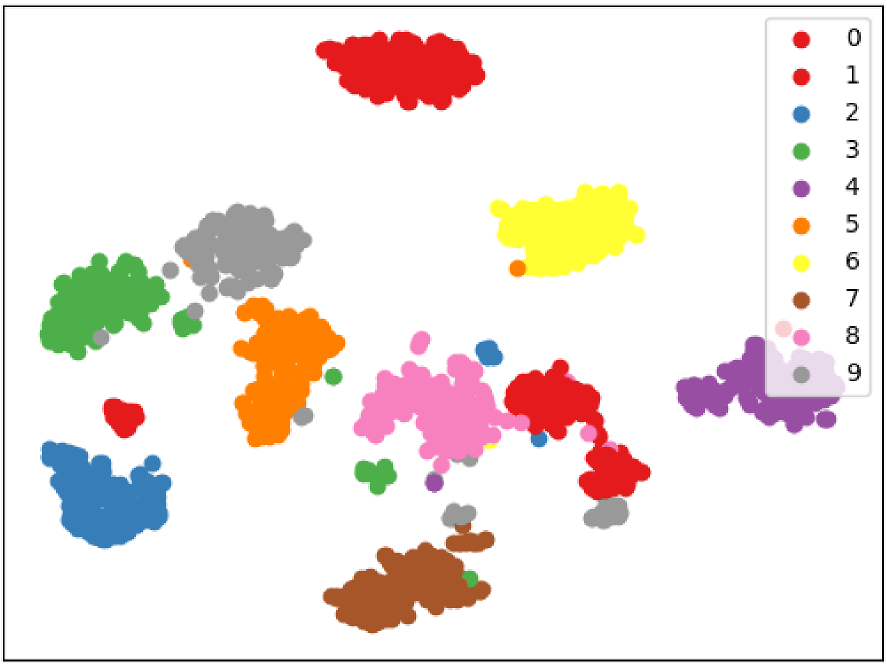
2-dimensional space. This will allow us to plot it using our plotting function. To use the Sklearn t-SNE function we first define the embedding object, then we fit the transformation using gradient descent on KL divergence as described above. When defining the object, we need to choose the model's hyperparameters which affect the final embedding. In our implementation, we first define the dimension of the embedding to be 2. To stabilize the optimization, we use PCA to initialize the embedding. Lastly, we choose the perplexity of the gaussian distribution to be 20. The perplexity in our case is the width of the gaussian and it is commonly set between 5 to 50. The above boils down into these two lines:

```
tsne = manifold.TSNE(n_components=2, init='pca', perplexity=40.0)
X_tsne = tsne.fit_transform(X)
```

There are additional hyperparameters that can be defined such as the number of iterations, stopping criterions learning rate and more. In our case, we set them as default parameters. These additional parameters can be used in the case that the embedding does not converge into the desired accuracy i.e. does not look good enough.

We are now ready to see some results. The above code generates the following graph:

t-SNE embedding



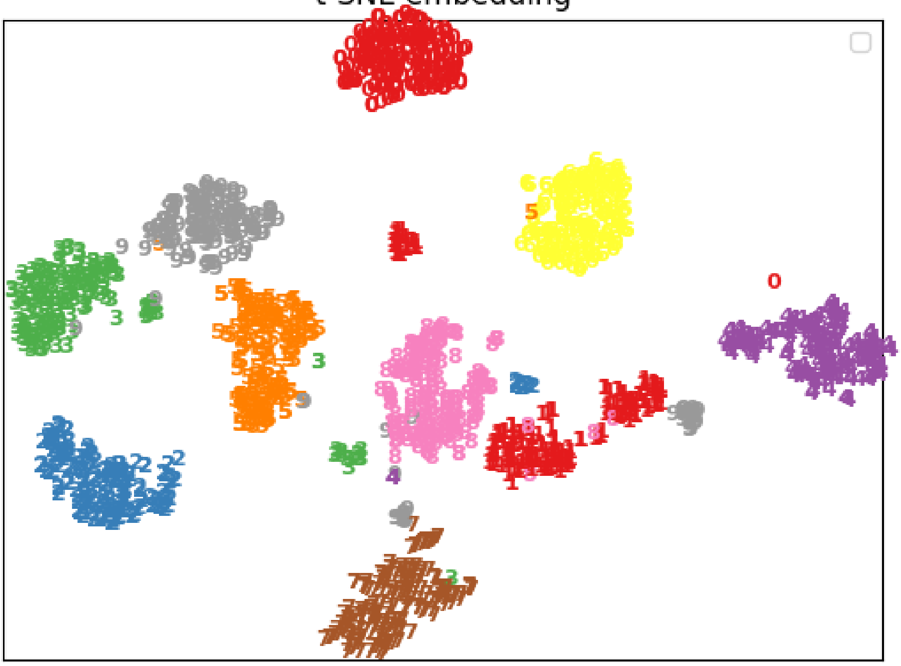

This figure is a typical clustering of t-SNE that shows the quality of the embedding. It shows that our data might be separable by using a linear separator. Moreover, each of the clusters is very concentrated around its means. In this case, the t-SNE gives us a very good intuition about the underlying data.

It is also possible to use some cooler visualizations using the same embedding. For example, we can write the digit of each cluster instead of each scatter. This is performed by replacing the *for* loop in our plotting function with this piece of code:

```
for i in range(X.shape[0]):
    plt.text(X[i, 0], X[i, 1], str(y[i]),
             color=plt.cm.Set1(y[i] / 10.),
             fontdict={'weight': 'bold', 'size': 9})
```

This new visualization gives us the following image:

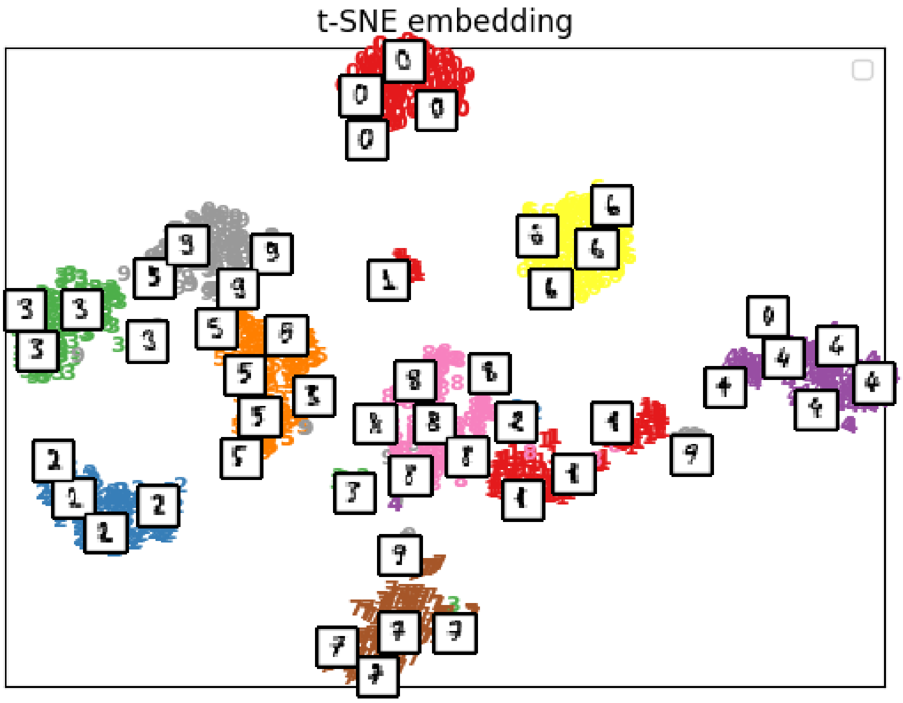
t-SNE embedding



Lastly, we can also attach an image to each of the clusters to get even more appealing results. This gives us our final results, displayed on the next page.

## Conclusion

As we have seen, t-SNE can be **a very useful tool to understand high dimensional data**. A widely used application of t-SNE is to visualize, in the way we did, the last layer of a neural network. In such visualization, one can see the level of separability of the last layer. Of course, it needs to be linearly separable since the last layer is only linear.




**You should take care of three caveats of this model!**

While t-SNE has great success and highly appealing results, you should take care of three caveats of this model: first, the clusters of the model might appear in non-cluster data due to incorrect parameters settings. Interactive exploration may thus be necessary to set the correct parameters. Second, t-SNE uses Euclidean metric, which makes it suffer from the curse of dimensionality when using high dimensional data. Lastly, the optimization process is not convex, and hence, **convergence is not guaranteed**.

We highly recommend that you try the model yourself using our code, to check whether it is useful for you.

# What Makes a Place a Place

**Among the organizers of the Long-Term Visual Localization under Changing Conditions workshop at CVPR 2019, Vassileios Balntas and Torsten Sattler share their insights on the challenges of long-term visual localization and the ways in which accurate long-term visual localization provides a vital component in many computer vision and robotics scenarios, including autonomous vehicles such as self-driving cars and other robots, augmented reality, structure-from-motion, and simultaneous localization and mapping (SLAM).**



With CVPR 2019 right around the corner, researchers in the field of computer vision and pattern recognition look ahead to the upcoming programs and presentations. The **Long-Term Visual Localization under Changing Conditions** workshop stands out as one of the most highly-anticipated events of the conference. Experts in the field from across industry and academia will attend the workshop to discuss practical challenges of developing visual localization under changing conditions.

Vassileios Balntas and Torsten Sattler, along with the other organizers, have been working on visual localization for quite some time. **Vassileios Balntas** leads the research team at **Scape Technologies**, which focuses on researching fundamental computer vision problems. Meanwhile, **Torsten Sattler** serves as an associate professor in the Electrical Engineering department at the **Chalmers University of Technology**.

Visual localization refers to the problem of accurately estimating position and orientation from which an image was taken with respect to a scene representation. Torsten shares: *"It's an interesting problem that we are*

## Changing conditions can dramatically affect the appearance of a scene

*addressing in the sense that it goes in the direction of what makes a place a place. It's a very common problem in computer vision, or in the general understanding of how humans work as well."*

***What makes a place a place****: that's what it boils down to, really. Humans can instinctively understand such cues. The workshop hopes to develop machine learning techniques to do the same. "We want to figure out how do I reliably recognize that I am at a certain position while the scene changes. To me, that is the essence of localization",* Torsten explains.

Current long-term visual localization algorithms depend on a scene representation constructed from images and must be robust enough to capture all potential viewing conditions including under different illumination conditions, seasonal conditions, and other changes over time. The dominant method involves recording images of a scene, which will remain valid forever. This works fine without much localization. However, any number of changing conditions can dramatically affect the appearance of a scene such as from day to night, from indoor to outdoor. When looking at outdoor scenes, vegetation grows. Buildings get renovated and demolished.

Torsten explains: *"If you want to augment reality, you need to know where you are in the world in order to be able to augment the view of the user with 3D objects that are in the scene. We would like to do this as accurately as possible, at pixel level or even more accurate, and we would like to do this robustly. If you can only do it on sunny days between 12 and 2 PM, who's going to use this, right?"*

*"Similarly, you need this for any type of intelligence system that needs to operate*

*in the same world that we operate in so self-driving cars or any type of autonomous robot. Again, the same thing: if you sell a car for a lot of money, but you can't drive it at night because our algorithms do not work at night, I would not pay for it. So you need this type of robustness!"*, he continues.

Existing localization algorithms only go so far and haven't yet solved the problem of long-term localization under a wide range of conditions. "*You try to match scenes with other random scenes, but you don't know what is inside the scenery. You don't have any representation of the semantics of the scene. Now, with neural networks and all of these recent ideas, we can build on top of that and actually understand what is going on in the scene*" says Vassileios.

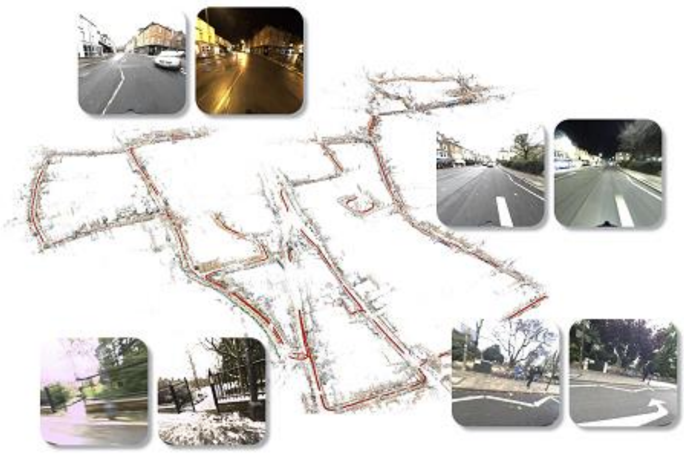
Long-term visual localization has many applications in industry, and the organizers of this challenge hope the workshop will motivate others who might have a better solution.

Vassileios shares: "*The obvious one is augmented reality. The basic rule is if you want to augment reality then you need to understand where you are. Self-driving cars is another very important aspect of it.*"

In addition, visual localization can allow extra-precise delivery: a drone system that can deliver a package directly to someone's balcony or other precise, private location. For example, the winning entry in Scape Technologies' first Hackathon (team Inition) demoed smartphones in which users draw a square with their finger inside the camera view. Then the drone lands exactly in this area.

"*Anything that has to do with extremely accurate position systems can benefit from this work. We use GPS systems now. We can use vision-related technology in the next years. Hopefully, we will have much, much better results!*" Vassileios elaborates.

The challenge aims to capture the impact of changes on localization algorithms, to create interesting, realistic datasets and apply state-of-the-art algorithms to the models. Looking at some of the immediate results, they still remain a long way from solving some of the datasets.



Current visual localization methods focus mostly on very accurate matching-based localization, which uses an image and then matches the image within the accuracy of centimeters. Another aspect involves deriving a very rough estimation, very

Challenge

quickly before moving on to a more robust, accurate localization.

Vassileios expands: *"I think this is also important: to make new methods that combine the two. It is also a very generic problem. You have to be robust to different things, but not to get lost in scenes that might look similar. Imagine you have a huge city. You have so many places that look exactly like each other. You have streets with repeating patterns and things like that. How do you localize in different levels of hierarchy of precision?"*

In recent work, they have found that it helps to integrate some sort of higher level scene understanding, for example, sample semantic segmentation, into the localization process, similar to what humans do. One problem that they ran into is that those semantic segmentation algorithms are not necessarily robust against all types of changes. In that sense, the datasets are a challenge.

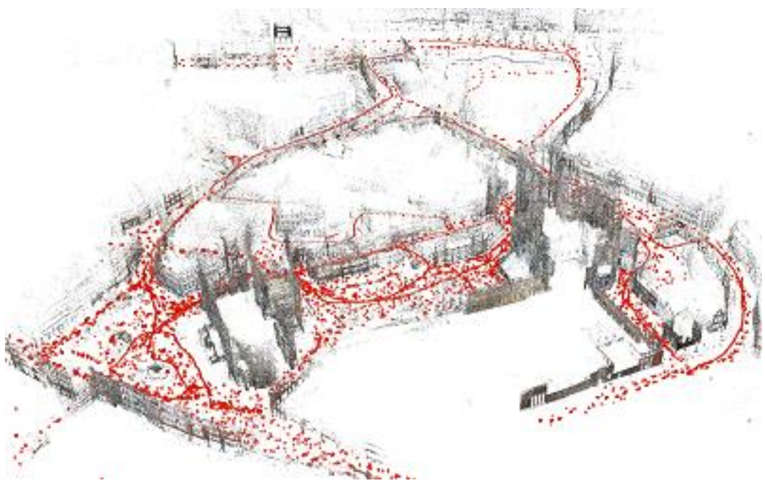
Organizers and speakers at the upcoming workshop are experts from both industry and academia. The blend of expertise at the workshop will give an interesting perspective from different communities which are often disconnected.

**Jan-Michael Frahm**, the head of the 3D computer vision group at the **University of North Carolina at Chapel Hill**, will share his experience researching a variety of topics on the intersection of computer vision, computer graphics, robotics.

**Srikumar Ramalingam**, associate professor in the school of computing at the **University of Utah**, will also join, sharing his research on computer vision, machine learning, robotics, and autonomous driving.

**Niko Sünderhauf**, senior lecturer at the **Queensland University of Technology (QUT) in Brisbane, Australia**, will share his insights on robotic vision, the intersection of robotics, computer vision, and machine learning.

Also set to speak: **Simon Lynen**, a tech lead/manager **at Google Zurich**, who works on large-scale localization from an industrial point of view. He can provide interesting feedback about the problems they run into that others may not have confronted.



**The Long-Term Visual Localization under Changing Conditions will take place on Monday, June 17th from 1:30 PM to 6 PM at CVPR 2019 in Long Beach, California. Computer Vision News will be there to publish for the 4th consecutive year the CVPR Daily magazine!**

The website of the workshop and the Impressive list of organizers are at https://sites.google.com/view/ltvl2019/

# Lena Maier-Hein

**Lena Maier-Hein is a computer science professor from the German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ) in Heidelberg. She received a prestigious European Research Council (ERC) starting grant for a multidisciplinary project that combines computer-assisted minimally invasive surgery with novel, gentle imaging technology based on sound and light. Machine learning-based image analysis is performed to convert high-dimensional data into clinically useful information.**

Read more interviews with women scientists

**Tell us about your work, Lena.**

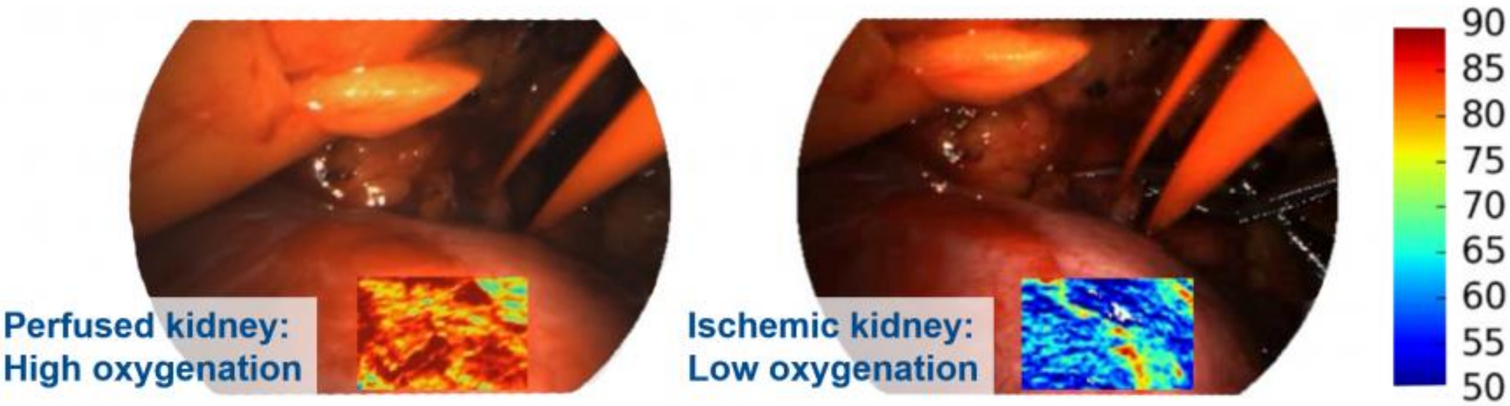
I would call myself a surgical data scientist. We're developing computational methods to advance interventional healthcare. That means surgery, in particular, but it can also refer to interventional radiology, for example.

**You should have been a doctor!**

I think my strengths are more on the theoretical side. I would actually be a



*"Our grandparents have been crucial to enable dual scientific career with kids!"*




**ERC Starting Grant COMBIOSCOPY: Machine Learning meets biophotonics for real-time functional imaging in the operating room.**

bad physician because I'm not very good at quick decision making. I think if I were in charge in the operating theater, patients would frequently die because it would take me too long to figure out how to proceed best.

**Have you been in an operating room?**

It's part of my job to watch surgeries. It's part of my job to apply our methods in the operating room. Also, I was a patient a few times already.

**How does it feel to know that technology that you have developed is helping people?**

It's very hard to bring a method all the way to the patient. We call it the translational gap. There are a lot of issues involved there. One is when you do research, sometimes you do it up to a point when it's good, but then you don't have the 100% robustness or reliability you need. This is a technical challenge. The other problem is, even if you have a good solution, someone needs to bring it to the market. It's basically impossible for a startup, for example, to start from scratch and bring something to the operating room. Say you work on endoscopic imaging data, and you have a great method to detect tumors in the data or to predict a complication. Then you have a niche expertise, but you don't have the expertise to get clinical data access to sell your products. Your innovation may be really good and really helpful, but there are a lot of things involved to translate it. What happens, in the end, is that some of the big companies need to be interested in your work, and then they can bring your innovation to the market. That's a huge problem, I think.

***"It is a flaw in the system. Definitely!"***

**It sounds like a flaw in the system.**

It is a flaw in the system. Definitely!

**A flaw that you and I cannot really solve.**

To some extent, we are working on it. For the first time, we got a project funded by the German Ministry of Economy – not by the Ministry of Research as usual. The aim of the project OP 4.1 is to reduce the market entry barrier for startups that work in computer aided surgery. In analogy to the operating system of a mobile phone, we aim to create a platform that will allow you to bring technical innovation to the operating room via something like apps. Still, we have a long way to go.

***"He said that it's the stupidest idea that he has ever heard because it will never work"***

**You have dedicated your career to healthcare. Did you choose this field for a reason?**

Healthcare was definitely not by chance. My background is computer science and I wanted to apply computer science to something useful for the community. When I heard about microchips that you could implant in patients with disabilities so that they would be able to grasp again, I decided to go for healthcare. But I wanted to do something more computational. In 2005, I told my PhD advisor that I would like to apply machine learning to medical imaging. He said that it's the stupidest idea that he has ever heard because it will never work. [*laughs*] He convinced me to do something different. I started my PhD working in surgery, more on the engineering side of things, not involving machine learning or any of these things. I became a researcher of computer-assisted surgery, working a lot with hardware and devices. At some point, I then became really excited about biophotonics techniques, which are basically methods that deal with the interaction between light and tissue and exploit the properties of how light interacts with tissue. I applied for an ERC starting grant to combine such techniques, in particular multispectral imaging and optoacoustic imaging, with machine learning methodology in surgery. In the project, we apply machine learning techniques (finally!) to convert high-dimensional data to information that is relevant for a physician. We approach the problem in a slightly different way than the biophotonics community, who has focused on model-based approaches for a long time. From the results that we currently have, I would say that machine learning approaches are pretty unique for getting the information fast in the operating room. So finally, after many years, I came back to the machine learning methods that I specialized on in my computer science studies, and I'm very excited about it.

**CARS 2019 is coming up in a few weeks. You are General Chair at one of its satellite conferences (IPCAI) and you are also organizing the Surgical Data Science Workshop.**

**With husband and kids:** ***"More than worth taking on the challenge to balance work and family!"***


***"A good network of friends is so helpful when combining family and science!"***

## *"I'm here because my supervisor made me!"*

The first international workshop on surgical data science was created in 2016 when we saw more and more success stories in all sorts of fields of data science, for example, radiological data science. We realized that we don't see these success stories in surgery. We figured that we should try to advance this field, identify the challenges, and define what we actually mean by "surgical data science" (in comparison to biomedical data science). Together with my colleagues, Stefanie Speidel and Pierre Jannin, we organized the first workshop where we invited leading people from the field of computer-assisted surgery, robotics, etc. We had an interactive workshop to define what we mean by surgical data science, what the problems are, and how to move the field forward. At the beginning of the event, we performed an anonymous voting: we had about 70 or 80 participants and we asked them for their opinions and why they were attending. There was quite a big portion of people that said: "*I'm here because my supervisor made me [attend the workshop]!*" [*smiles*] After the workshop, when we asked people anonymously again, everyone said that they would like to do it again. There was 100% agreement. Most of them voted for meeting annually or every two years. That was a success for us. People enjoyed the workshop after having attended it: and the workshop resulted in a Nature Biomedical Engineering paper 2017.

## *"In surgery, the situation is so much more complex…"*

Now we will have the second edition of the workshop. Personally, I don't know what the low hanging fruits related to surgical data science are. Nobody seems to know. A reason may be that surgical data science is comparatively complex from an infrastructure point of view. In radiological data science, for example, you have digital data anyway. And you have the radiologist who makes a diagnosis. All you have to do is place the algorithm between the two, right? And you can annotate the data that is digitized anyway to train the algorithm. After that, you can show a new radiological image to the algorithm to support the physician. In surgery, the situation is so much more complex. It starts with the fact that most of the information that is used by the physician is not digitized. Also, there is a whole team involved: the surgeons, the patient, the anesthesia team, the nurses, and maybe robots. All of them are involved in manipulating the patient and even making decisions during the intervention. They use tactile feedback. They listen to the acoustic signals. They see something, not on the screen, but on the real patient. All of this data is not being captured yet. There is a lot of variation in the procedure. It's not one 3D image that we're talking about. It's hours of surgery. It's multiple signals that are not being captured. It's really complex. In the end, I don't think we really have success stories, yet. This is one of the things that I would like to discuss at the workshop.



**It seems that you and your colleagues will be able to do great things soon, as a result of your research. You still sound very humble and do not give your research much credit. How can you connect these two things?**

I don't like overselling because I want people to know they can trust my words. I will not say it really works until the point when I'm 100% certain about it.

**I would like to ask you now about the Medical Decathlon. I know that you were involved in this challenge, which was sponsored by DeepMind, NVIDIA, and RSIP Vision. Our readers know about it because we interviewed Michela Antonelli and Jorge Cardoso a few months ago. What are your key take home thoughts?**

I think that the outcome was really interesting. I was recruited to the organizing team due to my team's expertise in challenges in general. We analyzed a lot of challenges, how rankings are performed, and what the problems are. Based on all my experience that I had before, I was very afraid that our ranking would result in many shared places such that we would not be able to make any strong conclusions. However, what we found was quite interesting. The winning algorithm had the hypothesis that you don't need to invent a new net architecture, but you just need to know how to use existing architectures and how to properly train your algorithm. It's an entirely new hypothesis. The title of the paper was actually quite provocative, which is "*No New Net*" (nnU-net because it builds upon the widely used U-net proposed by Olaf Ronneberger and colleagues, 2015).

It's not about inventing something entirely novel, but it's about doing things right. I like this hypothesis. I was really surprised and excited to see that the algorithm won by a large margin. I have to say though that I have a kind of a conflict of interest here because the winning team is part of my husband's group [*laughs*]. The algorithm is open-source though and was uploaded as a docker container, so there was definitely no cheating involved [*laughs again*].

***"The backbone of my work. Amazing team with great team spirit!"***

**"No New Net" is kinda cool! Doing things right is definitely crucial in science. What else is particularly important from your point of view?**

I want to mention how important the contribution from everyone is. There have been some articles about me and my group. Typically, the interviewer wanted the article to only be about me as the PI, but I don't think that's good. It's a team effort. The collaborators are important. The team is important. In surgical data science, you cannot be successful by yourself at all. It really needs a lot of different characters, talents, and people. I'm very grateful to my collaborators, many of them from the University Clinic Heidelberg, and my team!

[Read more interviews with women scientists]

*"In surgical data science, you cannot be successful by yourself at all!"*



*"My basketball team: sports is my preferred option for reducing the stress level!"*

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

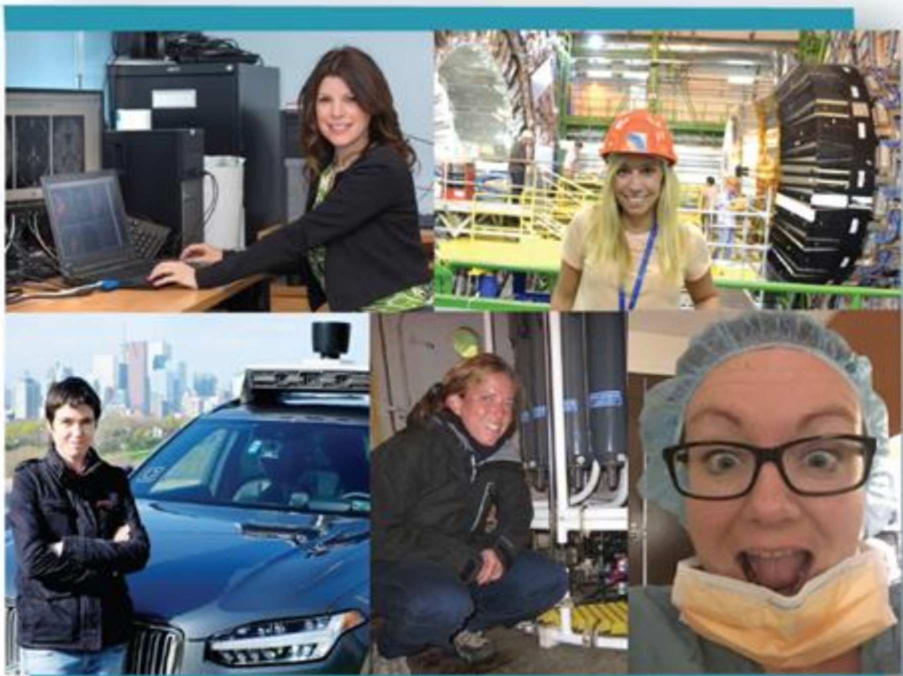
Confidence

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**

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

## Google releases AI training data set with 5M images:

Let's start with Google: they open-sourced a database called **Google-Landmarks-v2**, which follows a previous one from last year. This database contains over 5 million images of more than 200,000 different landmarks collected from photographers around the world. They also announce two new **Kaggle challenges**: Landmark Recognition 2019 and Landmark Retrieval 2019. **Read...**

## Google's lung cancer detection outperforms 6 radiologists:

This powerful AI uses a patient's current and prior CT volumes to predict the **risk of lung cancer**, and it allows to optimize the screening process via computer assistance and automation. RSIP Vision also detects and segments lung tumors! See page 16 in this mag. **Read...**


## LeCun - self-supervised learning for images and videos:

If you missed **Yann LeCun**'s now famous post about "unsupervised" and "**self-supervised learning**" (how could you?), here it is. And if you missed also Yann **LeCun's interview**, you should read it now. **Read More...**

## MATLAB 2019a - More AI, Systems-Engineering Support:

people are starting to use MATLAB's deep learning platform. The guys at electronicdesign.com have reviewed the "*quite a few enhancements and additions to an already formidable development package*". After this intro, you can't but **Read It!**

## Scientists help Artificial Intelligence outsmart hackers:

AI needs to outsmart hackers and neutralize hostile adversarial attacks. The alternative, when AI is vulnerable to patterns added by attackers, is to see this threat become commonplace. Here is how researchers at ICLR want to give AI a defensive edge. **Read More...**

**3 more links worth clicking:**

**This neural net would like to deliver these petitions:**

**TensorFlow Model Optimization Toolkit - Pruning API:**

**Talk to Transformer: how a modern neural network completes your text:**

# Women in AI Dinner




**At the RE•WORK Women in AI Dinner, held in Boston in May, some of the brightest minds in AI came together to share their research, applications and advancements in AI. The dinner welcomed attendees from various countries as well as industries, including: financial services, automotive and engineering, consumer goods, healthcare and more. Speakers included leading female experts from Google Brain, QuantumBlack, TripAdvisor and MIT Tech Review. All upcoming AI & Deep Learning [events from RE•WORK]()**


**UKIVA Machine Vision Conference and Exhibition**
Milton Keynes, UK — Jun 6 — Website and Registration

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

**CVPR - Computer Vision and Patter Recognition**
Long Beach, CA — Jun 15-21 — Website and Registration — MEET US!

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

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

**MIDL - Intern. Conf. on Medical Imaging with Deep Learning**
London, UK — Jul 8-10 — Website and Registration

**IEEE ICCI*CC Int. Conf. on Cognitive Informatics & Computing**
Milano, Italy — Jul 23-25 — Website and Registration

**MIUA - Medical Image Understanding and Analysis**
Liverpool, UK — Jul 24-26 — Website and Registration

**Medical Augmented Reality Summer School**
Zurich, Switzerland — Aug 5-16 — Website and Registration

**CMBBE Computer Methods in Biomechanics and Biomedical Eng.**
New York City, NY — Aug 14-16 — Website and Registration

**ICIAR International Conf. on Image Analysis and Recognition**
Waterloo, Canada — Aug 27-29 — Website and Registration

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

# Feel at CVPR as if you were at CVPR!

Subscribe to the CVPR Daily 2019

# 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!

# FREE SUBSCRIPTION

Dear reader,

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

**Subscription Form**
**(click here, it's free)**

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


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