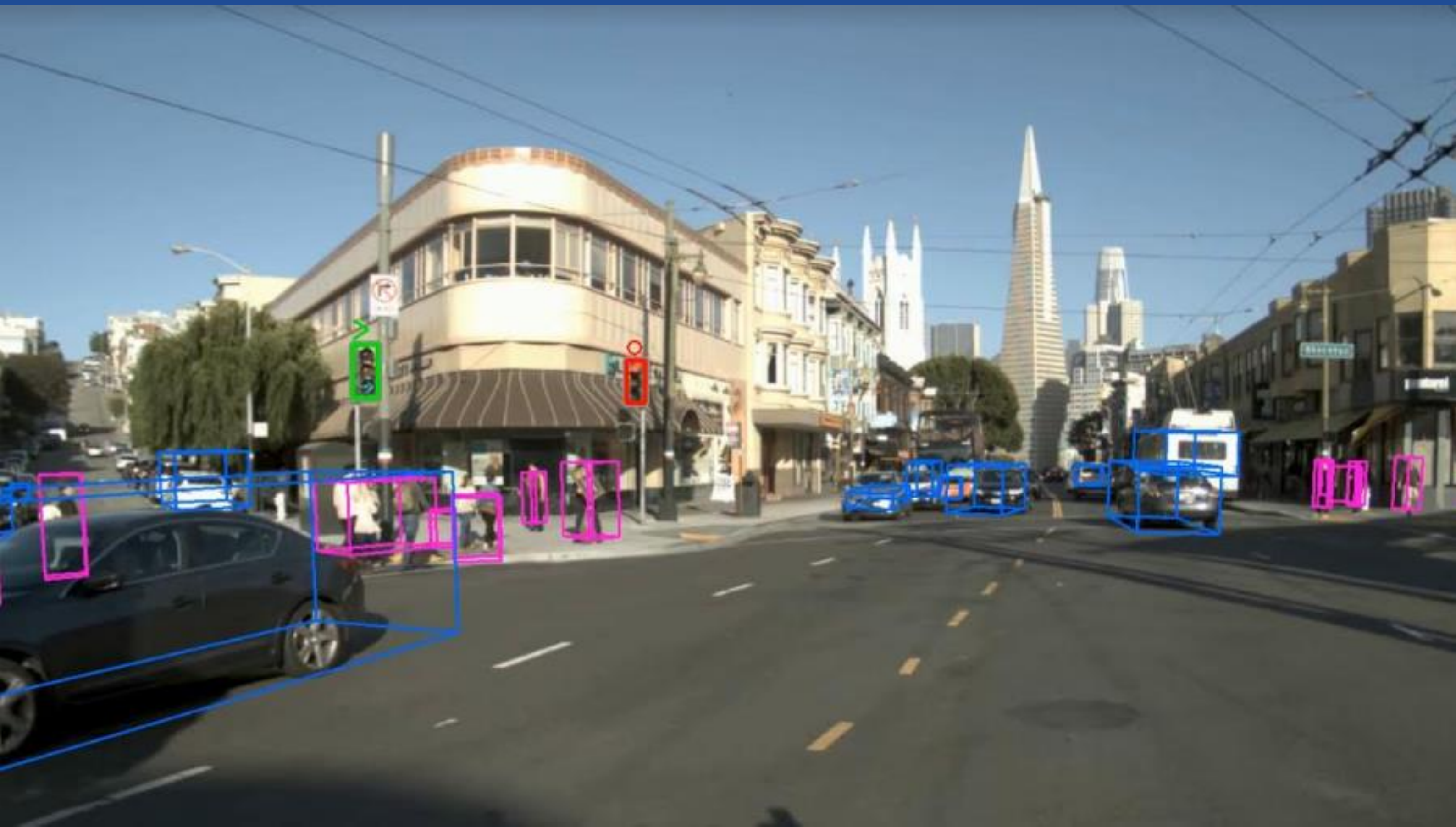
Computer Vision News

The magazine of the algorithm community

A publication by RSIP VISION

September 2018



Application of the month:
Zoox

Upcoming Events

Project:
Deep Learning for Medical Segmentation

Review of Research Paper
Beyond a Gaussian Denoiser:
Residual Learning of Deep CNN for Image Denoising

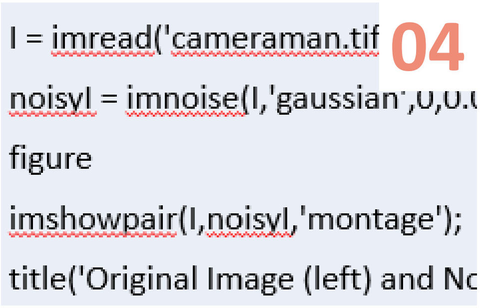
Focus On:
Aligned Feature Visualization
Interpolation for Deep Neural Networks

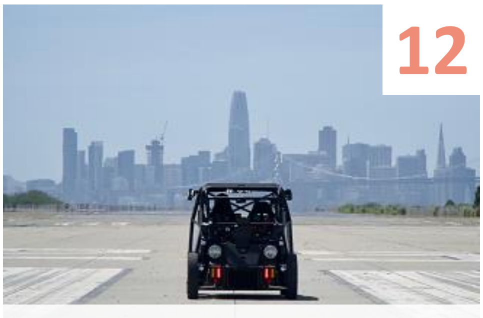
Spotlight News

Women in Computer Vision:
Iro Armeni (Stanford)

Project Management:
Splitting the Work in AI Projects

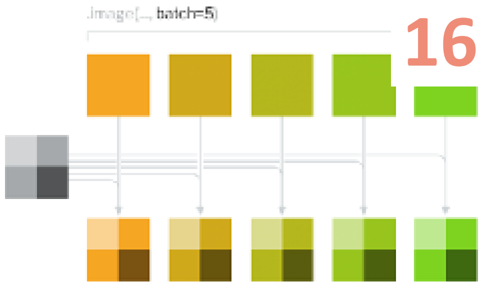
Contents

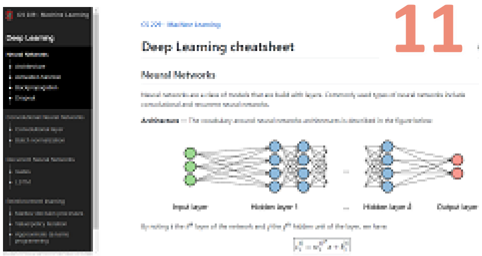

I = imread('cameraman.tif
noisyI = imnoise(I,'gaussian',0,0.0
figure
imshowpair(I,noisyI,'montage');
title('Original Image (left) and No




MICCAI2018
Granada
SPAIN


Hemorrhage & Edema
Annotate
ReTrain




Deep Learning cheatsheet
Neural Networks

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,

If you are going to **ECCV2018** and/or **MICCAI2018**, which will be held later this month, it will be our pleasure to meet you there. We will publish four issues of **ECCV Daily** and three issues of **MICCAI Daily**. If you cannot come, you can still subscribe for free to receive the ECCV Daily and the MICCAI Daily, almost in real time.

The people at **Zoox** shared interesting information with us about their original and intriguing project. Read the report on page 12. I could write the same about the story of **Iro Armeni** and her work bridging the gap between **civil engineering** and **computer vision**. On page 24, you find her fascinating interview with us.

Again, **Assaf Spanier** has prepared for us two more technical articles, **with codes!** On page 4, you can read the review of an impressive recently published paper; and on page 16, you find new insights into the behavior of **neural networks**.

As usual, **Ron Soferman** offers us an enriching lecture about project management in **Artificial Intelligence**. Don’t miss it on page 22.

Enjoy the reading and, as always, take us along for your next Deep Learning project!

Ralph Anzarouth
Editor, **Computer Vision News**

LAST MINUTE: Google releases an open source reinforcement learning framework for training AI models. Read the blog post by Pablo Samuel Castro and Marc G. Bellemare, Google Brain: **Introducing a New Framework for Flexible and Reproducible Reinforcement Learning Research**

by Assaf Spanier

Every month, Computer Vision News reviews a research paper from our field. This month we have chosen: **Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising**. We are indebted to the authors (**Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhan**), for providing us with great new images to illustrate our review. Their article is here.

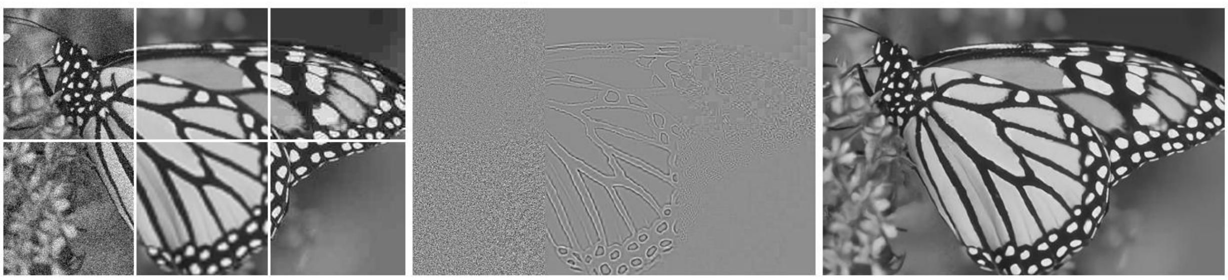
DnCNN outperforms state-of-the-art methods handling both blind gaussian denoising (with unknown noise level), SISR and JPEG image deblocking.

Introduction:

Image restoration is a preliminary step in low-level computer vision with many applications. Although many very good approaches have been proposed, active research into better methods continues. The goal of image denoising is to recover clean image x from corresponding noisy image y, assuming y = x + v, v is commonly assumed to be additive white gaussian noise with standard deviation σ. Two special cases are single image super-resolution (SISR) and JPEG image deblocking.

In this paper the authors investigated the construction of feed-forward denoising convolutional neural networks (DnCNNs) incorporating the progress in very deep architecture. DnCNN outperforms state-of-the-art methods handling both blind gaussian denoising (with unknown noise level), SISR and JPEG image deblocking.

To give you a general idea of the capabilities of the DnCNN model for dealing with different denoising tasks, the Input Image below (a) is put together with 3 types of noise, each at two different noise levels. Gaussian noise with σ=15 (top left) and σ=25 (bottom left), low-resolution interpolation images with an upscaling factor of 2 (top center) and 3 (bottom center), and JPEG images with a quality factor 10 (top right) and 30 (bottom right). The white lines in the Input Image are only for us to differentiate the 6 different noise-type areas. The Output Residual Image (b) was normalized to the range [0,1] for illustrative purposes only. (c) is the denoised – restored Image.



(a) Input Image (b) Output Residual Image (c) Restored Image

Aim & Motivation:

Classical denoising methods used image prior modeling as a stepping stone for denoising. Although those methods achieve high denoising quality, they suffer from two main drawbacks: (a) they require a complex, computationally demanding, optimization problem for testing. (b) the non-convex models require manual selection of several parameters, leaving room for performance improvement.

Advantages:

The authors propose an end-to-end trainable feed-forward denoising convolutional neural network (DnCNN), taking advantage of the progress in deep learning methods (i.e. batch normalization, residual learning). The network doesn't directly output the denoised image ^x, instead it is designed to predict the residual image (the noise itself). And rather than learning a discriminative model using an explicitly predicted image prior, it treats image denoising as a plain discriminative learning problem, that is, separating the noise from a noisy image. The authors propose a single network to solve three general image denoising tasks: blind gaussian denoising, SISR, and JPEG deblocking.

Background:

Image denoising, like every other field in computer vision, can be divided into two periods: the models developed prior to the deep learning revolution, and the models developed since :-).

Prior to the deep learning revolution, image denoising explored various methods for modeling image priors, including gradient models, sparse models, Markov random field (MRF) models and nonlocal self-similarity (NSS) models, which are implemented in state-of-the-art methods such as BM3D, LSSC, NCSR and WNNM. As mentioned, these methods' main disadvantage is a complex, computationally demanding optimization problem at the testing stage.

Super-resolution is the process of creating high-resolution images from low-resolution images, one of the best known approaches being SISR, where the purpose is the reconstruction of a high-res image from one single low-res image. SISR is challenging because without any data from a high-res image, the quality of the reconstructed image is necessarily limited. Moreover, SISR is an ill-posed problem, since an infinite catalogue of high-res images can be produced for every low-res image.

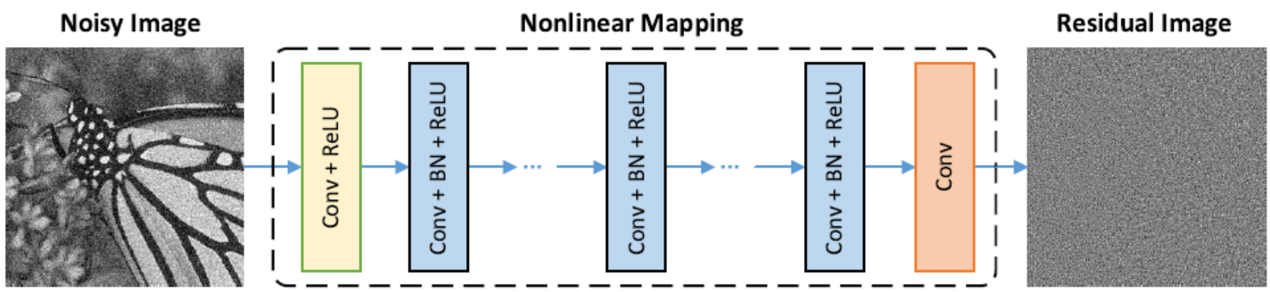
TNRD (Trainable Nonlinear Reaction Diffusion model), despite being proposed in 2016 (after the deep learning revolution), still uses prior models. Developed by Chen et al., its critical point lies in the additional training of the influence functions. The effectiveness of the trained diffusion models is attributed to the following properties of the trained filters:

(a) find rotated derivative filters in different directions; (b) contain first, second and higher-order derivative filters; (c) adaptive diffusion learned through the nonlinear functions.

Since the revolution in deep learning, several methods have been proposed to overcome the limitations of prior-based approaches, by eliminating iterative optimization at the testing stage. One of them is **VDSR** a neural network architecture whose purpose is single image super resolution (SISR). The VDSR network learns a mapping between low resolution images and their respective high resolution images, learning to predict their residual image (noise). However, unlike DnCNN, this network only deals with and was only tested on SISR tasks, and doesn't use newer techniques such as batch normalization.

Method:

The DnCNN network the authors developed has a simple basic structure made up of three types of layers: (1) Conv+ReLU, (2) Conv+BN+ReLU, and (3) Conv, shown in the figure below in yellow, blue and orange, respectively.



Training details	**Network name**
For gaussian denoising with known noise level: the authors considered three noise levels: σ = 15, 25 and 50. Patch size was 40x40 Network depth (number of blue layers): 17	DnCNN-S.
train the DnCNN model for a range of the noise levels (σ) as 0 through 55, and the patch size as 50X50.	DnCNN-B B is for blind (noise level is unknown)
Trained on color version of the BSD68 dataset	CDnCNN-B B is for blind / C is for color
Trained simultaneously on 3 image denoising tasks: blind gaussian denoising, SISR, and JPEG deblocking.	DnCNN-3

Code and Implementation:

Implementing the DnCNN network using the Keras software package will look as follows:

- A first layer of 64 3x3 filters with ReLU activation.
- Followed by 15 identical convolutional layers with batch normalization and ReLU activation.
- With another 3x3 filter

```
def DnCNN():

  inpt = Input(shape=(None,None,1))
  # 1st layer, Conv+relu
  x = Conv2D(filters=64, kernel_size=(3,3),
        strides=(1,1), adding='same')(inpt)
  x = Activation('relu')(x)
  # 15 layers, Conv+BN+relu
  for i in range(15):
      x = Conv2D(filters=64, kernel_size=(3,3),
                 strides=(1,1), padding='same')(x)
      x = BatchNormalization(axis=-1, epsilon=1e-3)(x)
      x = Activation('relu')(x)

  # last layer, Conv
  x = Conv2D(filters=1, kernel_size=(3,3),
               strides=(1,1), padding='same')(x)
  x = Subtract()([inpt, x])   # input - noise
  model = Model(inputs=inpt, outputs=x)

  return model
```

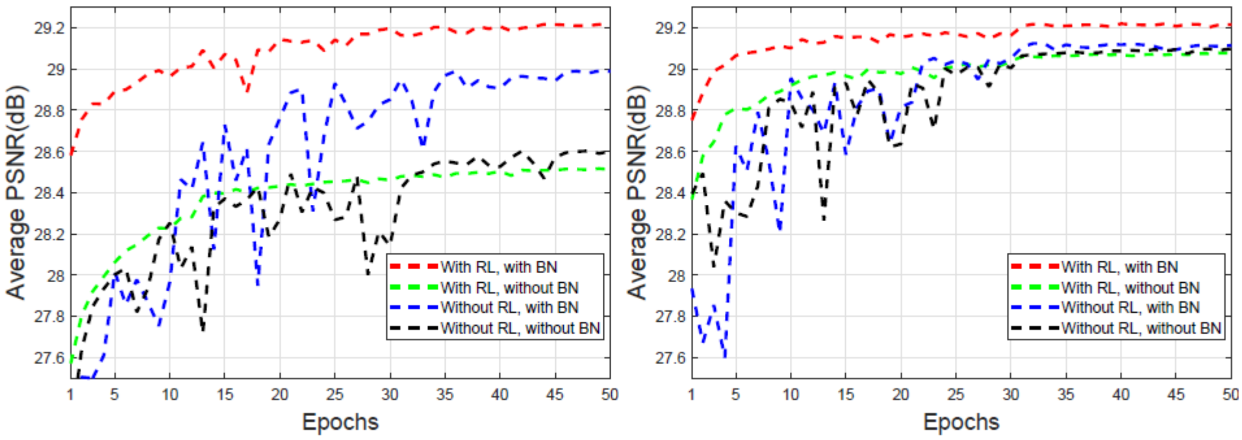
A pre-trained DnCNN network is now included in Matlab out-of-the-box -- you can just run:

```
net = denoisingNetwork('DnCNN');
```

And test it using the following code:

```
I = imread('cameraman.tif');
noisyI = imnoise(I,'gaussian',0,0.01);
figure
imshowpair(I,noisyI,'montage');
title('Original Image (left) and Noisy Image (right)')
denoisedI = denoiseImage(noisyI, net);
figure
imshow(denoisedI)
title('Denoised Image')
```

Quantitative Results:



(a) SGD (b) Adam

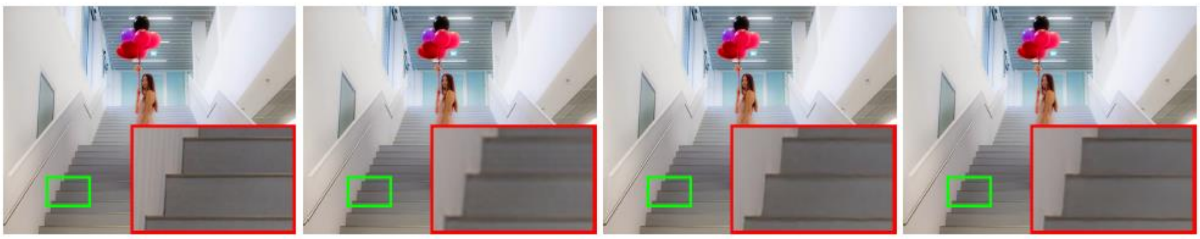
The authors evaluated several configurations for training the network: (1) They compared the performance using different optimization methods -- SGD vs. Adam. (2) They tested the effect of batch normalization (BN) on network performance, by comparing the network including BN in the blue blocks, to one without them. (3) They compared the residual learning (labeled 'with RL') plain discriminative training model they were proposing, where the model is trained to output the noise, with the model trained to output the denoised image ^x (labeled 'without RL').

You can see in the figure above, that the best performance and fastest training time was achieved using Adam, with the network trained to output noise (residual learning), and the network including batch normalization (BN).

You can also clearly see the importance of including batch normalization for residual learning, which gains highly in both performance and speed of network convergence when BN is present. The disparity in 'with RL' performance between 'with BN' and 'without BN' is especially high, which makes sense, since that residual learning aims at predicting gaussian noise and batch normalization helps us normalize the activation function's output exactly to this range.

With highly impressive qualitative results, as you can see below:

(a) Noisy / 14.76dB (b) BM3D / 26.21dB (c) WNNM / 26.51dB (d) EPLL / 26.36dB
(e) MLP / 26.54dB (f) TNRD / 26.59dB (g) DnCNN-S / 26.90dB (h) DnCNN-B / 26.92dB



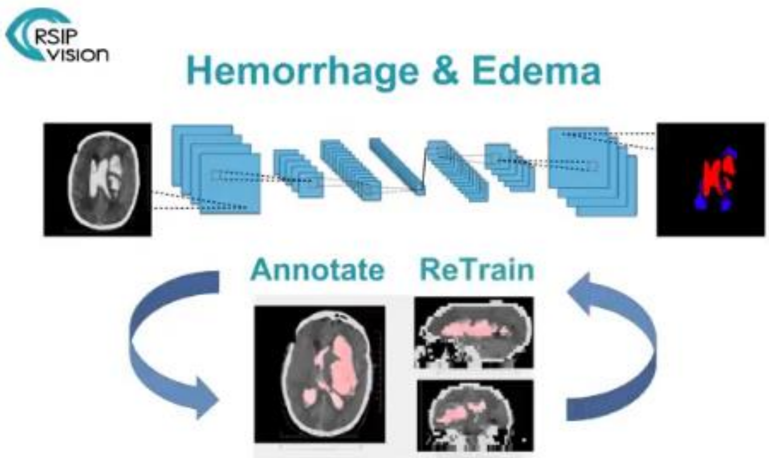
(a) Ground-truth (b) TNRD / 32.00dB (c) VDSR / 32.58dB (d) DnCNN-3 / 32.73dB

(a) Ground-truth (b) TNRD / 28.91dB (c) VDSR / 29.95dB (d) DnCNN-3 / 30.02dB

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 physical difficulties. This month we review **RSIP Vision's Deep Learning solution for a Fully Automated Intracranial Hemorrhage and Edema Segmentation**. RSIP Vision's engineers can assist you in countless application fields.

A fully automated medical segmentation for many types of Intracranial Hemorrhage, during which running time is constant.

Blood vessel can erupt inside the brain, causing internal bleeding. This condition (called **ICH** or **IntraCranial Hemorrhage**) can be deadly when it is not treated correctly and timely. Some brain hemorrhages come with **cerebral edema**: accumulating fluid in intracellular or extracellular spaces within the brain.



ICH is generally diagnosed using **MRI** or **CT scans**. However, edema appears as a subtle dark area around the hemorrhage; it is therefore difficult to detect, requiring multiple scans. Successful segmentation can be achieved by expert computer vision techniques, such as performed by **RSIP Vision** for all **fully automated medical segmentations**: these techniques may be classical or built on **deep learning** state-of-the-art procedures such as RSIP Vision's **CNN-based solutions**. Watch the video above to find out more details about this procedure.

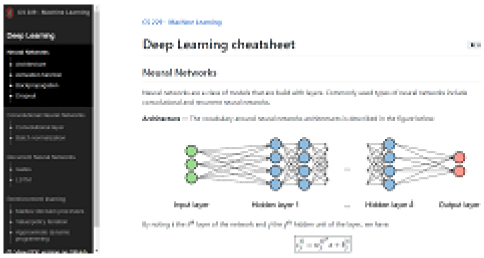
It is rare to find properly annotated datasets in sufficient quantity for training. Instead of recurring to a huge manual annotation task, experts at RSIP Vision prefer to augment the dataset using an **optimized semi-automatic segmentation** task, by which classical computer vision techniques (**superpixels**, **graph cuts** and more) enable a human expert to easily annotate a training dataset, subsequently used to train a small neural network.

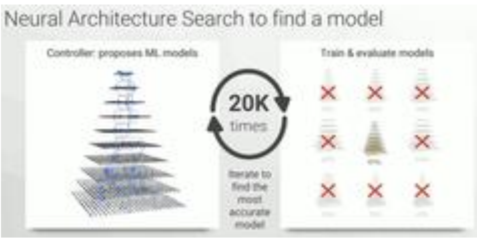
Once these tasks are completed, new techniques are called by RSIP Vision's engineers to train larger and better neural networks. The result is that RSIP can provide its clients with a fully automated medical segmentation for many types of Intracranial Hemorrhage, during which **running time is constant**.

This **Artificial Intelligence approach** can be called hybrid and is most suited when training datasets are not sufficient to use the most effective deep learning methods available today. This method proposed by RSIP Vision has provided an optimal solution to many clients and may solve your own medical segmentation problem. [Contact RSIP Vision today] and find out how its engineers will find the best solution for your case.

[Super VIP Cheatsheet: Machine Learning and others!]

Let's start from the top of the tops: a 15-pages cheatsheet covering the content of Stanford's CS229 class on **Machine Learning**, with content originally taught by **Fei Fei Li**, **Andrew NG** and others. More cheatsheets: Deep Learning, Supervised Learning, Unsupervised Learning, Tips and tricks. Pure Gold!



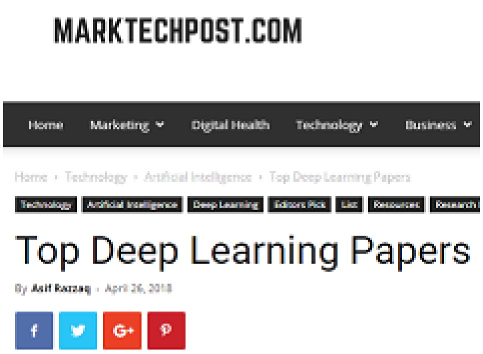


9 Things You Should Know About TensorFlow:

Last month many readers liked the piece by Cassie Kozyrkov. So here we found a new one, even more interesting, reviewing a talk on **Google Cloud Next** about **what's new with TensorFlow**. Cassie's great article is here and Laurence Moroney's video is here.

Why computers are so bad at comparing objects:

Very clear article explaining in simple terms why **computers are so bad** at a class of tasks that almost **no human has problem with**: determining whether two objects in an image are the same or different. Read it Now!



Top 23 Deep Learning Papers:

Why exactly **23 papers** I don't know: you should ask the guys at marktechpost.com who prepared the list. It is not new but we noticed it only very recently. Regardless, it is a remarkable choice of **Deep Learning papers**. Among the authors: Jürgen Schmidhuber, Yoshua Bengio, Yann LeCun, Matt Zeiler, Kaiming He, Lisa Anne Hendricks, Kate Saenko, Tamara Berg, Subhashini Venugopalan, Marcus Rohrbach, and many more friends of Computer Vision News. The 23 Papers

Smaller Collaborative Robots Disrupting Robotics Industry:

Very complete review of current issues and solutions to robots in the industry and how a new category of robots is rapidly becoming mainstream - here is how RSIP Vision does it. BTW, if you want to see how a robot learns dexterity, Watch the Video on the Right.

Artificial Intelligence Draws Faces from Text Descriptions:

It's a brilliant idea: it tries to give a face to book characters we can only figure in our mind or watch in movies. His project called T2F uses a **Generative Adversarial Network (GAN)** to encode text and synthesize facial images. Watch the Video on the Left.

David Pfeiffer

Zoox describes itself as a group of ***"inventors, builders and doers"*** and, as a company that develops fully autonomous vehicles from scratch, they surely live up to the description. The startup has already raised nearly $800 million in funding. Their end goal? To provide self-driving vehicles as a service, i.e. robot taxis.

While certainly not the only company working towards autonomous driving, Zoox stands out compared to others because, rather than building on existing vehicles, they have actually redesigned the entire idea of transportation by building a robotic vehicle from the ground up.

The idea started with Zoox founders, **Tim Kentley-Klay** and **Dr. Jesse Levinson**. Kentley-Klay, a designer originally from Australia, considered the tremendous potential of this type of technology. He reached out to Levinson, who at that time was leading the **Stanford Autonomous Driving Project**. They met, and together, eventually began collaborating until Zoox was finally born.

The Lidar team manager **David Pfeiffer**, along with the company's Director of Detection and Tracking, **Sarah Tariq**, share fascinating insights with us into this new autonomous vehicle and the future of self-driving cars.

David says that technology completely changes the setup of a car as we know it today. Their system allows them to get rid of all extraneous parts needed in a car driven by humans which are now obsolete in a fully autonomous configuration. For example, their model allows passengers to sit facing each other, rather than all facing forward.

The perception technology **fuses three types of modalities: radar, lidar, and cameras**. David explains why the cars require this combination:

"Why all three? Every sensor has different strengths and weaknesses. The camera has magnificent resolution and is really good at classifying objects, but lacks depth. Lidar has excellent depth but is somewhat sparse, and tracking is somewhat complex. Radar doesn't have great lateral resolution, and only limited accuracy in terms of saying where objects are. But it's very good with velocity and works equally well in all weather conditions.

All of them are, in their way, perpendicular to another. You want to guarantee maximum availability of objects when you detect them. It's a bit of a change in paradigm. Let's consider a traditional off-the-shelf collision warning system: even when the car beeps for no reason, the driver is always in charge to decide what to do, right?

Here, it's different. You always have to do the right thing. You can't fail. You just want to get to the point where you leverage all availabilities as good as you can for a system that is both high on recall in detecting an object and doesn't make any mistakes."

Currently, Sarah works more with the vision team while David spends more time focusing on lidar. Both methods use similar techniques, and at the end of the day, they work as **one big perception team**.

Sarah Tariq

How do they do it? Well, therein lies the real challenge. David expands on how they leverage each modality to the fullest so that they can compensate for the others' limitations. For example, lidar has great depth, but sometimes struggles with classifying objects, particularly ones that are far away. During detection, cameras can then figure out the object's type. Then lidar assists further in providing a better understanding of depth. At the same time, radar establishes the velocity and improves tracking of moving objects.

"We try to leverage different sensors in terms of their strengths to enable more robust detection and tracking" David explains.

In terms of the challenges in vision, Sarah adds: *"I think that what might be challenging for all of us is long-tail events. We're using a lot of neural networks. These are obviously very good at things that we've seen and understood well. They work well in detecting most normal situations. It's really planning for these long-tail things that will inevitably happen and that happen so rarely that it's hard to collect data for or plan models for them. But we also have ways to detect whether there's an object present at all, even if we don't know exactly what it is, and that is very important for our safety case."*

Sarah tells about her early years at Zoox. *"I've been here almost three and a half years. When I came, the robots we were driving were not on public roads. We had very limited 3D perception and no computer vision at all. Even that was so amazing! To see this **little golf-cart-looking thing**, this very different looking vehicle that could navigate through these experiments -- for example, a person would jump in front of it and it would stop. We still do that sometimes. It's amazing that it can recognize and stop for all sorts of different situations, people, things, animals… it's very cool! Now we've moved to **San Francisco**. We go out very often. All the software engineers, especially those writing code that runs on the vehicle, are encouraged to and do **go out very often to be able to see and feel what the vehicle is doing**. We are very intimately connected with what the problems are."*

David goes on to talk about the challenges in his work. He explains the difficulty in building a vehicle that can drive on the road autonomously using sensors: **the more the system matures, the more sophisticated and hard it gets to improve it**. You fix things, but at the same time you could make other things worse. It's very hard to make sure you consistently improve and narrow in on the problem. It took a reasonable amount of training effort to actually convert the system to a point where it improved over time without exploding with complexity. Then they began building solutions that didn't require too much effort to maintain and still had the potential to scale up.

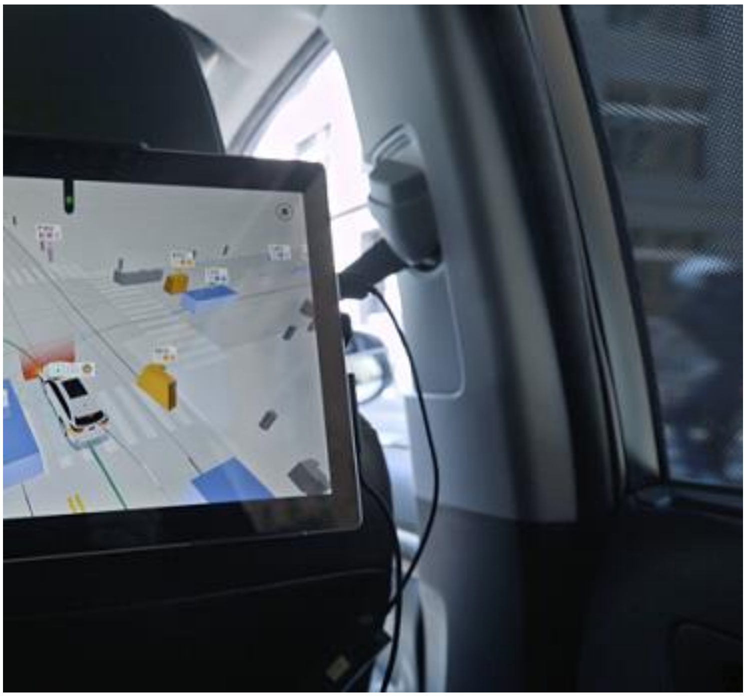
He elaborates: *"You may need to understand more of how we develop. We have a certain regime that is called our **ODD (Operational Design Domain)**, where we define the types of situations we intend to handle. We removed a lot of situations out at the beginning. Otherwise, **we would be drowning in complexity**.*

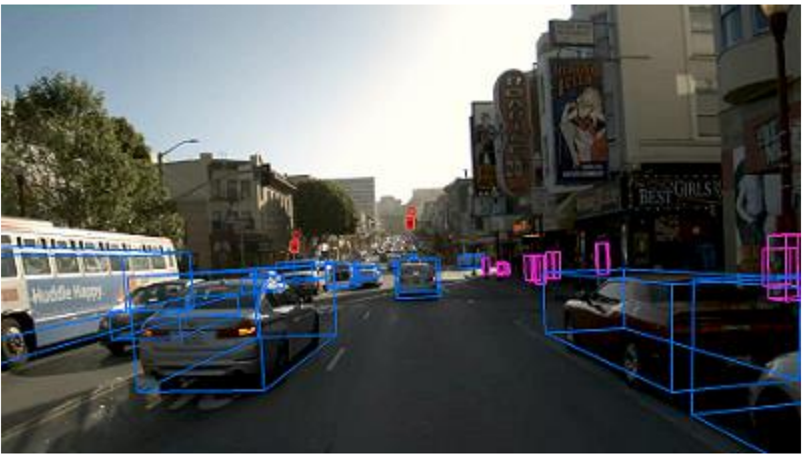
"We take safety very seriously as a company"

Let's use an example. Initially, we took out left turns and other scenarios as well. When we start driving, and we get better, we bring more of these things in because, at the end of the day, we have to solve them. So the system gets more complex. The scenarios that we have to handle get more complex. At the same time, you want to have a solution that you can maintain, you can grow, that doesn't break down once you change out a sensor, use a different type of sensor, change sensor placements, or cameras... This is something I find very challenging. Also, it's very challenging once you change something to validate that nothing breaks. At the same time, you need to provide solutions that are able to run on the vehicle. You can have the best algorithm in theory, but if you can't compute it in real-time, it doesn't help you."

David comments on the public misconceptions of this technology: *"People talk about vehicles on the road today that have driver assistance. They take that technology as an autonomous system, but it's not. It's an assisting system. To **switch to an autonomous system**, there is much, much more to it. An assisting system is aimed mainly on emergency brakes, and you only need to make sure that it doesn't misfire. It never guarantees the safety of a pedestrian. Whereas, we have to do both. It can't misfire. It should never break for no reason, but if there is a pedestrian, it must always break. This a completely different problem, and we take safety very seriously as a company."*

Sarah expounds: "*What David says is very important: people need to realize the massive difference between an **L3 system** and an **L4 or L5 system**, and that the sort of issues that these systems might have are totally different. L3 systems are designed to give the human*





*the warning that “*Hey, you need to take over!" *You might have to take over very, very fast. The problems might come from humans not being able to take over in the time that the system expects. In our vehicles, there’s no such backup, so the system is designed to be safer.”*

Ground-up vehicles will be on public roads in the next couple of years...

Looking ahead, Sarah revealed that Zoox hopes its ground-up vehicles will be on public roads in the next couple of years.

She says: “*It should be deployed in one urban area, say San Francisco. That would be our launch site where people can call a Zoox robot from an app somewhere in the city and be able to go somewhere else. We will have a fleet of these robots deployed, and* ***we’re meeting the whole ODD that we set out****. We should be able to drive completely autonomously, of course with the safety requirements that we set for ourselves. We have set, as an internal goal, to be* ***much safer than humans*** *at driving in these areas.”*

Sarah adds: “*The other thing that I want to bring up is for people to remember where most of the accidents today come from. Approximately 94% of accidents today come from human judgment or human error. When speaking of autonomous vehicles, people sometimes think of really difficult situations that go wrong where maybe there was no good choice, like between hitting one person or the other person and what the robot would do. But it is true that these situations are tricky for human drivers as well and they occur very rarely. The everyday situations that happen are when you are distracted, texting on the phone, drunk driving, falling asleep and similar things - these are all situations we would completely avoid with robots.”*

In essence, their system will avoid the circumstances that lead to most human fatalities and injuries from vehicles. The robots have **full attention all around the vehicle at 360 degrees**, whereas humans have very selective concentration.

“I have a kid myself and every day I look at this vehicle, and think if it’s going to be safe enough to drive on the same street as where my child is”

As a mother, Sarah looks forward to the benefits that Zoox will bring. “*I have a kid myself and every day I look at this vehicle, and think if it’s going to be safe enough to drive on the same street as where my child is. I would like to get to a point when I trust them to pick up my kid and bring him home; that would be great!”*

With the tremendous progress of Zoox, it seems the days of getting picked up by a robotic car may come sooner than we all think. BTW, they are hiring!

Focus on: Aligned Feature Visualization Interpolation for Deep Neural Networks

by Assaf Spanier

One of the most complex issues involved, which is yet to be fully cracked, is the attempt to understand the internal decision mechanisms and processes of deep neural networks

A large neural network made up of hundreds of millions of artificial neurons for image classification? How? How does the network arrive at the specific outcomes it does?

This is a big question. One of the most complex issues involved, which is yet to be fully (some would say, even partly) cracked, is the attempt to understand **the internal decision mechanisms and processes of deep neural networks**. The attempt to interpret these internal processes has become one of the hottest areas of research in the field of **deep learning**. Early research focused on trying to identify the crucial neuron within the network and understanding what it affects as well as how. Later, researchers tried to understand the activity of groups of neurons and the integration between them, recognizing the fact that neuron #123456 being activated five times doesn't really tell us anything truly useful about the network as a whole.

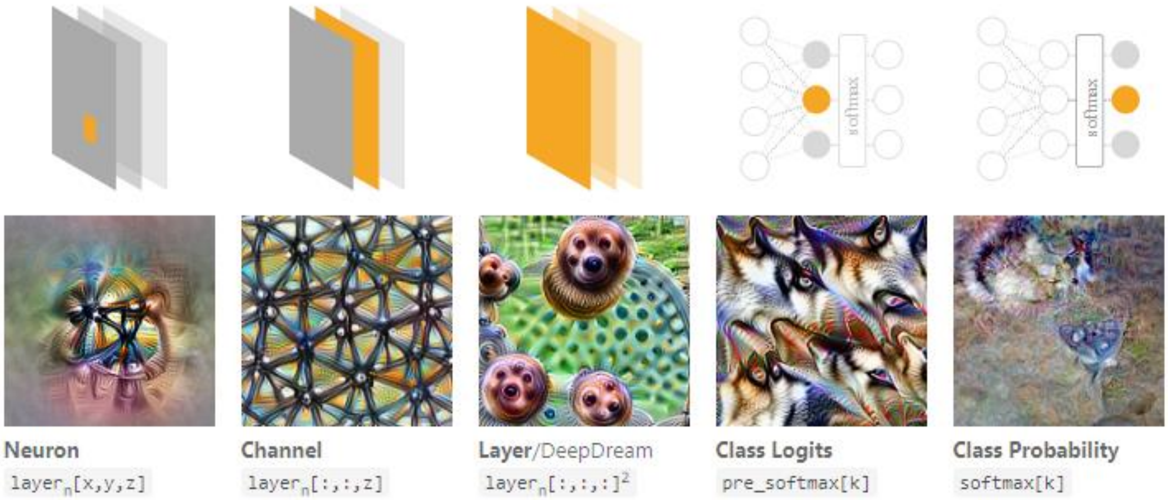
Tool

The research into interpreting the decision-making mechanisms in neural networks focused on three main areas: (1) feature visualization, (2) attribution and (3) dimensionality reduction.

The central insight of the latest research in the field is not seeing these interpretative techniques in isolation, each standing on its own, but as composable building blocks towards more comprehensive models, each helping foster some insight into the behavior of neural networks.

The goal of the integration of these building blocks isn't just to explain what features it is that the network identifies, but to understand the mechanisms by which the network integrates these small pieces to arrive at decisions further down the line, and why/how it arrives at the specific decisions that it does.

In this **'Focus on' article**, we will talk about **feature visualization**: a very effective technique for understanding the processing of data by and between single neurons. Feature visualization can take place at various levels: the individual neuron, channel, layer, class logits, or category, as illustrated below.



Note that visualization at the category level is a technique that, unless you use special regularization constraints, is identical to creating adversarial examples, which we talked about in the May 2018 issue (Adversarial Examples: Attacks on Deep Learning).

From the spectrum above we will focus on the feature visualization at the neuron level, specifically we will demonstrate studying interaction between a few neurons, rather than focus on an individual neuron. We will be using a special technique, detailed below, for aligning the visualizations produced to facilitate interpretation and analysis of the results. We will be using **Lucid, the software library by Google**. We've already seen some visualizations produced by Lucid, in our June 2018 issue (Focus on: Debug and Analysis Mechanisms for Deep Learning in TensorFlow and Keras). And I would guess this is not the last time we'll be seeing them in these articles, since the visualizations are fascinating and have a major contribution in our capability to conceptualize the internal processes of deep neural networks.

As stated, early feature visualization methods studied the effects of **single neurons within the network**. However, later, this technique was expanded to **groups of neurons**, maximizing the overall activation of the group, rather than the activation of a single neuron, in order to study, visualize and interpret the effect of the interaction of two or more neurons.

This approach faces a 'little' challenge. Because the feature visualization process is based on a random initial state, despite the fact we optimize and visualize for the same function and the same object, the visualization produced will be slightly different each iteration (a different output, like a different angle or spread of features). This was not a problem for classic feature visualization research which only studied a single neuron at a time, but becomes a problem for visualizing the interaction among several neurons. If we run the visualization without special constraints, the visualizations arrived at won't align -- recognizable visual landmarks crucial to successful interpretation (the eye-like features, in our example) will appear at different locations in each image, which will greatly reduce the usefulness of the visualization technique for analysis.

How can we arrive at an aligned interpolation, which will have stable locations for visual landmarks? A shared parameterization is one approach to tackle this. In shared parameterization, each epoch (to be explained below) has a number of unique parameters and one parameter in common.

By making some of the parameters shared between neurons this way, we make the visualizations produced naturally align. This technique constrains nearby regions to share their parameters and as a result greatly reduces localization-changes between visualizations.

Let’s go deeper into this with an example of some coding. Using Lucid’s custom parameterization capability to build a shared parameterization that encourages alignment of visual landmarks, we’ll create visualizations that are interpolations of two feature visualizations. Let's take a look at the two neurons:

```
neuron1 = ("mixed4a_pre_relu", 476)
neuron2 = ("mixed4a_pre_relu", 460)

for neuron in [neuron1, neuron2]:
  param_f = lambda: param.image(128)
  objective = objectives.channel(*neuron)
  _ = render.render_vis(model, objective, param_f)
```

Lucid `render.render_vis` controls visualization using a few components which you can fiddle with in a completely independent way:

- objectives -- What do you want the model to visualize?
- parameterization -- How do you describe the image?
- transforms -- What transformations do you want your visualization to be robust to?

Now, we want to see the optimization for several neurons -- in our case, two -- we can use the batch parameter with a value of 2 for this. We concatenate 2 objective functions in parallel, one for each batch -- the first batch will produce a visualization representing the first neuron and the second batch will produce a visualization representing the second neuron, as you can see in the code below.

```
param_f = lambda: param.image(128, batch=2)
objective = objectives.channel(*neuron1, batch=0) + objectives.channel(*neuron2, batch=1)
_ = render.render_vis(model, objective, param_f)
```

This technique will allow us to study a variety of interactions between these two neurons. We will study 5 of them: A 1-0 ratio (that is, the first neuron only), 0.75-0.25 ratio, 0.5-0.5 ratio, 0.25-0.75 and 1-0. We will output the result of each of these 5 combinations:

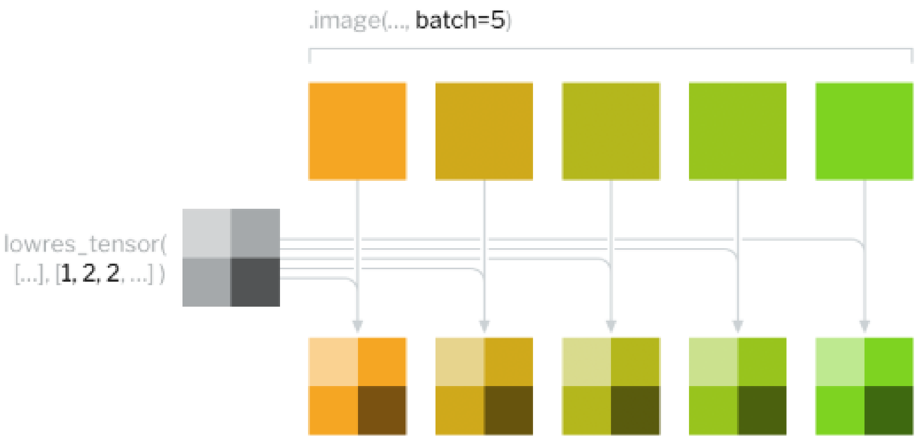
```
param_f = lambda: param.image(128, batch=5)
objective = objectives.Objective.sum([
   # neuron 1 objectives, orange row:
   1.00 * objectives.channel(*neuron1, batch=0),
   0.75 * objectives.channel(*neuron1, batch=1),
   0.50 * objectives.channel(*neuron1, batch=2),
   0.25 * objectives.channel(*neuron1, batch=3),
   0.00 * objectives.channel(*neuron1, batch=4),
   # neuron 2 objectives, green row:
   0.00 * objectives.channel(*neuron2, batch=0),
   0.25 * objectives.channel(*neuron2, batch=1),
   0.50 * objectives.channel(*neuron2, batch=2),
   0.75 * objectives.channel(*neuron2, batch=3),
   1.00 * objectives.channel(*neuron2, batch=4),
])
_ = render.render_vis(model, objective, param_f)
```

And the result is five images, one for each combination we set under the objective parameter.

That’s all well and good; however, to be able to better analyze the interaction between neurons we now want to align features. We’ll do this by using the function `lowres-tensor`. This function gets two shape objects and produces a per-dimension bilinear interpolation between those shapes. Let’s get an idea of what this means with a simple case:

Let's say we have an image parameterization like before, with a shape of `(5, 128, 128, 3)`. Then we could ask for a `lowres_tensor` of shape `(1, 128, 128, 3)` to get a shared parameterization. Here, for instance, is the fixed parameterization area, each taking up one quarter of the image.

```
def lowres_param_f():
  shared = lowres.lowres_tensor((6, 128, 128, 3), (1, 128//16, 128//16, 3))
  return color.to_valid_rgb(shared, decorrelate=True)
```



It is now expanded to a pyramid of spaces in a manner reminiscent of a generalized Laplacian pyramid, as demonstrated in the code below.

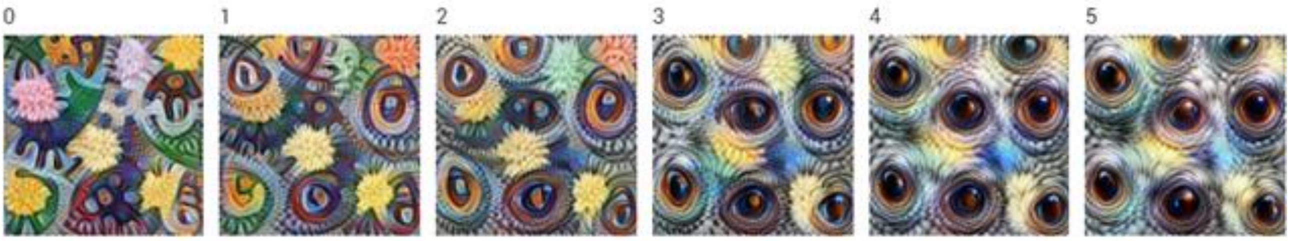
```
def interpolate_f():
 unique = spatial.fft_image((6, 128, 128, 3))
 shared = [
  lowres.lowres_tensor((6, 128, 128, 3), (1, 128//2, 128//2, 3)),
  lowres.lowres_tensor((6, 128, 128, 3), (1, 128//4, 128//4, 3)),
  lowres.lowres_tensor((6, 128, 128, 3), (1, 128//8, 128//8, 3)),
  lowres.lowres_tensor((6, 128, 128, 3), (2, 128//8, 128//8, 3)),
  lowres.lowres_tensor((6, 128, 128, 3), (1, 128//16, 128//16, 3)),
  lowres.lowres_tensor((6, 128, 128, 3), (2, 128//16, 128//16, 3)),
 ]
 return color.to_valid_rgb(unique + sum(shared), decorrelate=True)
```

Now that the space of shared parameters between the neurons is in a number of resolutions, the sharing of the image space is much more tightly aligned and therefore the movement of visual landmarks between the differently combined images we are studying will be largely reduced. And when we run the following code:

```
objective = objectives.channel_interpolate(*neuron1, *neuron2)
images = render.render_vis(model, objective, param_f=interpolate_f,
verbose=False)
```

We'll get the following result:

Images from https://distill.pub/ (CC-BY 4.0)



Read at page 24 our interview with **Iro Armeni**, our **Woman in Computer Vision** of September.

Read also many more interviews with [Women in Science](#) on our online archive.

Tool

Team - How to Split the Work in AI 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 we learn about **Splitting the Work in AI Projects**. It's another tip by **RSIP Vision** for **Project Management in Computer Vision**.

Split the work among the team in such a way that everyone gets an interesting role and gives a significant contribution to the project

Some AI projects are big enough to require a team of 4 to 5 programmers. If the project manager wants to start the coding as soon as possible, it is necessary to allocate resources efficiently. At the same time, everybody is eager to work at the Deep Learning part, get a feel of the results and solve the problem.

Here are options to split the work among the team in such a way that everyone gets an interesting role and gives a significant contribution to the project:

1) **Different approaches**: we can try to use different architectures for the network and even compare the work in Patch-based vs full convolutional neural network or other options.

2) **Augmentation**: datasets are rarely found in sufficient size and overfitting may occur even with large datasets. It is often necessary to perform data augmentation in order to obtain the needed dataset for the learning phase.

3) **Annotation**: sometimes it's easier to give the work to annotators; but it is often possible to perform a semi-automated annotation using computer vision technique. The outcome of this might be increased speed and/or a larger dataset.

4) **Feature extraction**: this new idea is based on the thought that good features can provide a better analysis of the results. For instance, going deeper through an average result of 90%, we might find that most failure come from specific cases and features, shedding light upon the weaker points of the network or the dataset.

5) **Sanity check**: sometimes, when rare cases of bad example cannot be supplied to the database, we can use tools to detect failures that we do not want to transform into false positives.

6) **Different Pre-processing**: in many cases, pre-processing might give much better images, that can yield better results. It will be worthwhile to allocate resources in this direction as well.

7) **Open source options**: scientific literature provides alternatives which are offered with open source code. This enables testing different approaches in an efficient way, with the goal of finding the most valuable option.

I am confident that you can successfully apply these ideas and principles!

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.

3DV - International Conference on 3D Vision
Verona, Italy — Sep 5-8 — Website and Registration

ECCV 2018- European Conference on Computer Vision — MEET US!
Munich, Germany — Sep 8-14 — Website and Registration

ImageCLEF 2018 - CLEF initiative labs [*Read More…*]
Avignon, France — Sep 10-14 — Website and Registration

CGVC2018 - Computer Graphics and Visual Computing
Swansea, UK — Sep 13-14 — Website and Registration

CSCS 2018 - ACM Computer Science In Cars Symposium
Munich, Germany — Sep 13-14 — Website and Registration

MICCAI - Medical Image Computing and Comp. Assisted Intervention — MEET US!
Granada, Spain — Sep 16-20 — Website and Registration

RE•WORK Deep Learning Summit + Deep Learning in Health Care
London, UK — Sep 20-21 — Website and Registration

IROS 2018 - IEEE/RSJ conf. on Intelligent Robots and Systems
Madrid, Spain — Oct 1-5 — Website and Registration

ICIP 2018 IEEE International Conference on Image Processing
Athens, Greece — Oct 7-10 — Website and Registration

ISMAR 2018 IEEE Symposium on Mixed and Augmented Reality
Munich, Germany — Oct 16-20 — Website and Registration

RE•WORK Deep Learning Summit
Toronto, Canada — Oct 25-26 — Website and Registration

BENAIC / BENELEARN 2018 - Benelux Conf. AI / Machine Learning
's-Hertogenbosch, NL — Nov 8-9 — Website and Registration

AI & Big Data Expo North America — MEET US!
S.ta Clara, CA — Nov 28-29 — Website and Registration

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

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!

Iro Armeni 色

Iro Armeni is a Ph.D. candidate at Stanford University, Civil and Environmental Engineering (CEE) Department, Sustainable Design and Construction (SDC) Program. She is interested in interdisciplinary research between Civil Engineering and Machine Perception with an area of focus on automated semantic and operational understanding of buildings throughout their life cycle using visual data.

Iro, what are you doing at Stanford?

I am doing my Ph.D. at the Civil and Environmental Engineering (CEE) Department. However, I work in between two departments with Martin Fischer being my advisor in the Center for Integrated Facility Engineering (CIFE) and Silvio Savarese being my advisor at the Computer Science Department, Stanford Vision and Learning Lab (SVL).

Are you a scientist, an architect or both?

That's a big question, in terms of architecture, whether it's more of an art or a science. I find myself being in a "*creative science*". I find a lot of inspiration from architecture and the world of engineering on how to address computer vision problems and how I can bridge the gap between these two domains and see what I can find between them.

Why did I ask you this question?

I guess it's because I am doing interdisciplinary research between civil engineering or the construction industry and computer vision.

My background is in architecture. I spent many years studying and working as an architect. The reason I turned to AI and computer vision (I knew nothing about it beforehand) is because of issues in my work environment. I was dealing with renovation, at that point, and I needed to understand the building before being able to do any new design. This process, which at that point would usually be done with a tape measure, takes a lot of time. What I wanted to do is to be able to jump into the creative process, having in my hands some information about the existing building. That was hard to do. There should be a better way to do it, and not just go around with a tape measure to try and figure out what is going on in the building. I got introduced into 3D sensors, which was great. Then I understood the issue behind them which is, now what do we do with their output, how to get the information we need out of it. We have to semantically understand it in order to automate the process. Otherwise, we just go back again to a very tedious process of manual modeling. That's how I started looking into computer vision, about all of these 2D and 3D modalities, how we can automatically extract semantic information out of them by creating algorithms and "teaching" them. That was very exciting for me.

It was that problem that stole you from one community into another.

Well, I don't think I was stolen. I work in between the two. Sometimes, I find myself towards one side or the other side. I'm always in between, trying to bridge the two. No matter how much I love artificial intelligence, I will never forget my architectural background. I know that the construction industry has so many problems with being able to introduce technology in a sustainable way. I find myself wanting to help move this industry a step forward. Being able to do that would be awesome.

I think I know why you don't want to forget your roots. Is it because you are Greek? I learned from Nikos Paragios that there is a saying in Greek: "*You should never forget your roots*". And now you are in the USA…

Yes, for my work, I've been to different places all over the world. Of course, home is always home. Also, home is

where you are living.

You are actually living the same way you work... two different communities that you want to belong to and two different homes that you feel you belong to in a way. Is there any other complexity that we should know about?

[*laughs*] I never thought of it like that! I think there must be more complexities in my life. [*both laugh*]

What was one of the most complex things that you have done until now?

I'm not sure if it's complex, but it's something that has been a big change. Coming from the architectural field into the AI field definitely creates some kind of difference in how your mind works. Still, I wanted to go back and do more math and science. I'm trying to keep both in balance. Both the creative side, which comes out in different ways inside the work I am doing now, and being able to do more math as well.

What is your current work?

I continue working on being able to understand semantics in 3D scenes. Semantics not only means specific objects, like walls, chairs and tables, but being able to expand them into understanding properties, relationships, how humans interact with these objects... Being able to understand the affordances within the buildings and the elements. I am also interested in the spatiotemporal understanding of buildings, from the moment they get constructed to the moment they get demolished.

What will you be doing ten years from now?

[*laughs*] I have no idea! I don't know where I'll be in terms of location. I know that I'll be working on this problem and being able to introduce this thinking and technology into the construction industry, being able to help and introduce more students and people into how we can, in a good way, leverage AI in our domain. Last quarter, we taught with my advisor a seminar class in the Civil and Environmental Engineering department about what kind of applications of AI exist in the construction industry. I saw that there were a lot of students and people in the industry that don't understand how AI can be leveraged. They see it more as a trend that they have to chase. I think that the industry lacks a lot of foundations before it can go forward. That's something that I would like to introduce to more people so that they can understand what we're missing and what we can do next.

If you did not have this background in engineering and architecture, and you worked in computer vision and artificial intelligence, what subject would you have chosen?

I can't imagine working on anything else. I love 3D space geometry and semantics. I couldn't imagine working on any other subject.

Which is the nicest building in the world, seen with the eyes of an architect?
That is a very tricky question! I would say that there are so many different cultures. It is culture that gets reflected in buildings. There are buildings that I like from different cultures or eras; architecture reflects the changes in society. When I go back to Athens, it's much fun for me to visit the Parthenon and Acropolis. Every time that I go there, I feel some kind of energy, the light and energy of the building. That is, for me, incredible. I also really like the Japanese culture and the traditional Japanese buildings. I imagine all of the things that took place in these traditional homes. I think that's the most important part. I don't have a favorite building, but I like how our culture and all of our activities take place in buildings from different eras and different parts of our thinking and of our society. In terms of architecture, function, shape and all the things that you could talk about, it still impresses me.
What is the most important home improvement in modern times?
I think more information in the way we use our houses, from understanding more about the air quality to being able to have more sustainable energy performance.
You are not afraid of automation in your house, like not being able to get in or get out because some robot decided that you do not belong there?
[*both laugh*] Well, there might be glitches in going forward. I don't know. In going forward, I don't believe that AI will kill us all…
So we are not in a movie…
[*both laugh*] There are things that we have to be aware of, we have to sort out, of course. I am working on complications.
Tell your American friends something good about Greece or your American friends something good about Greece that they don't know about. You can choose!
All these are things I haven't thought about for a long time! As for America to my Greek friends, I would say that things in Greece are not in the best shape. Financially, people are struggling to find a job. Here (in the States), there are more opportunities. If you want to do things, you can work hard, and you can achieve something. Whereas, right now in Greece, it's harder to do it. You work a lot and try a lot, but it's harder to achieve anything. Young people with dreams have these issues, and we need to be able to change that.
So if you had a little sister, ten years younger than you, you would tell her to come to America.
Well, they can go wherever they want. I'm not telling them to come to the US. I'm just telling them that we need to somehow change our mentality. In Greece, we do work hard. I think that we should be able to do better. We should be able to go back to when things were better than that and restructure.
Did you read any of the "Women in Computer Vision" interviews that we published before yours?
I did look at them, yes.
Which one do you prefer?
I don't think I have a favorite one: I liked most of them. There were maybe one or two that were not my favorite...
[*both laugh*] Definitely, you are a great interviewer!


Improve your vision with
Computer Vision News
The Magazine Of The Algorithm Community
The only magazine covering all the fields of the computer vision and image processing industry
Subscribe
(click here, it's free)
RE•WORK
% invoke the matlab debugger
function STOP_HERE()
[ST,~] = dbstack;
file_name = ST(2).file; fline = ST(2).line;
stop_str = ['dbstop in ' file_name ' at ' num2str(fline+1)];
eval(stop_str)
Mapillary
Building
Traffic Sign
Car
License Plate
Gauss Surgical
A publication by
RSIP VISION