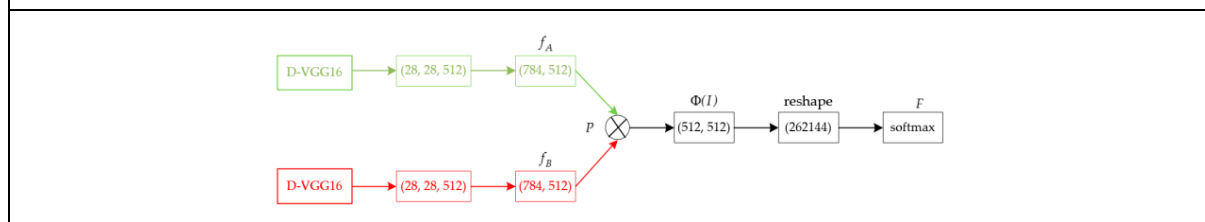
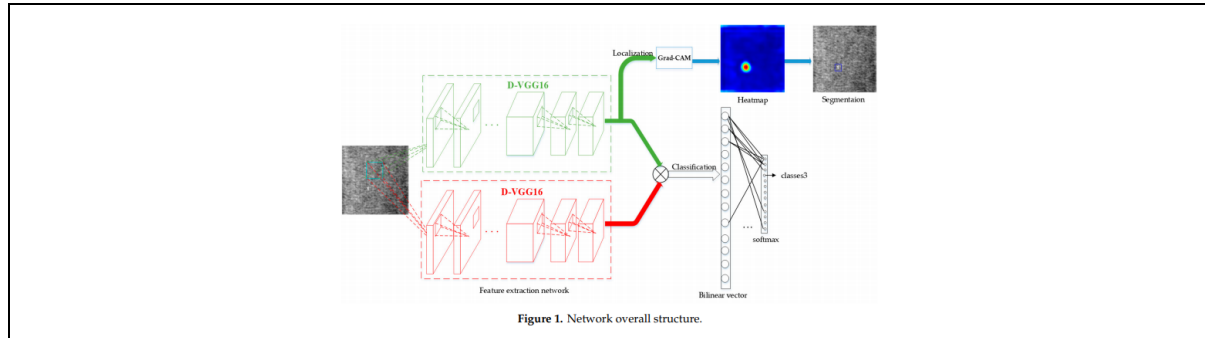


# #1. A Generic Automated Surface Defect Detection Based on a Bilinear Model By Fei Zhou, Guihua Liu \*, Feng Xu and Hao Deng

① This paper used a new Double-Visual Geometry Group 16(D-VGG16) that is firstly designed as feature functions of the bilinear model so that to realize the automatic classification and localization of surface defects.



→ D-VGG16 is a network structure model that has two fully symmetric(shape, width, height, number of channels and convolutional kernel) bilinear structures.

→ For the classification task using the Convolutional Neural Network (CNN), the simplest way to improve the accuracy of small sample training and avoid overfitting is to reduce the feature map of the last layer of CNN without decreasing the receptive field of the network.

② The global and local features fully extracted from the bilinear model by D-VGG16 are output to the soft-max function to realize the automatic classification of surface defects.

③ Then the heat map of the original image is obtained by applying Gradient-weighted Class Activation Mapping(Grad-CAM) to the output features of D-VGG16.

④ The defects in the original input image can be located automatically after processing the heat map with a threshold segmentation method.

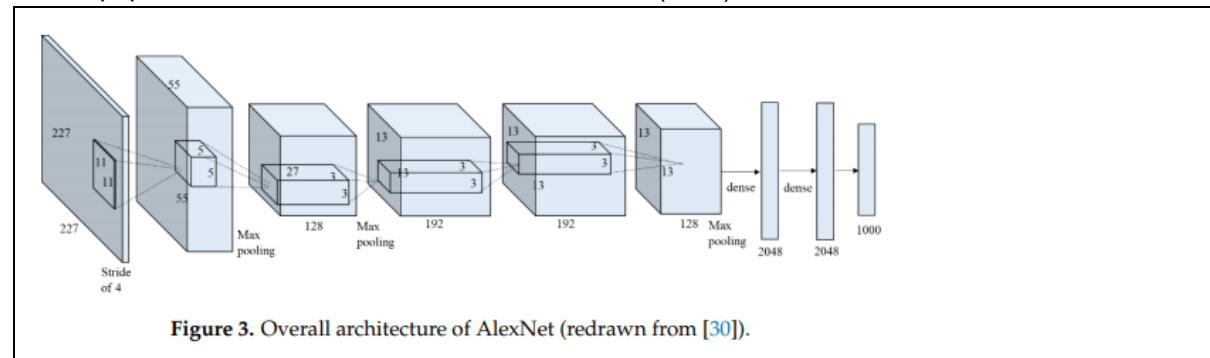
→ the heat map of the original image is obtained by applying Gradient-weighted Class Activation Mapping (Grad-CAM) to one of the output features of D-VGG16, and then the corresponding defect location can be determined in the input image by a threshold segmentation to the heat map.

⑤ Similar to most deep convolutional neural networks, the back-propagation algorithm was used as the training rule, and we then minimized the loss function with respect to the network parameters using Adam.

## #2. Automated Vision-Based Detection of Cracks on Concrete Surfaces Using Deep Learning Technique

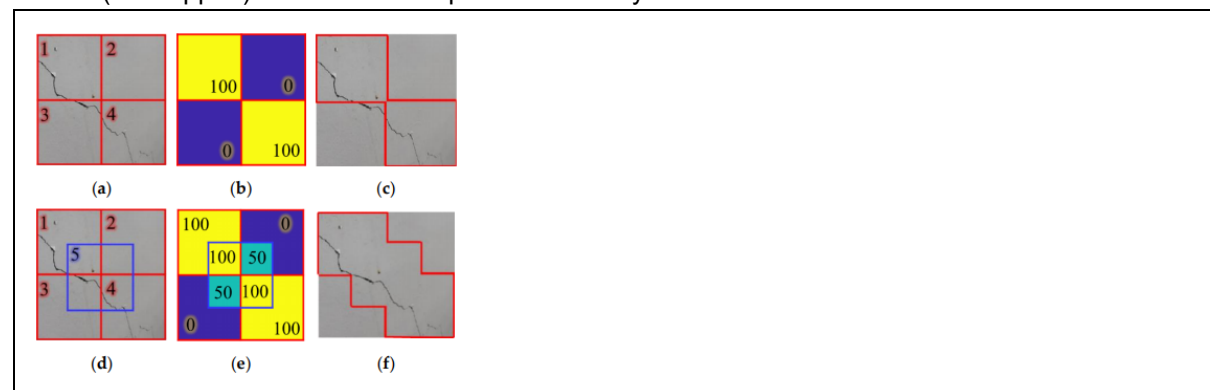
By Byunghyun Kim and Soojin Cho

① This paper is based on convolutional neural network(CNN), AlexNet



② The training set is divided into five classes involving cracks, intact surfaces, two types of similar patterns of cracks and plants.

③ A probability map is developed using a softmax layer value to add robustness to sliding window(overlapped) detection and a parametric study was carried out to determine its threshold.



④ I need to analyze TP, TN, FP, and FN on images that have detected cracks.

⑤ Grouping of FPs and FNs, and their possible solutions

False-Positive(FP)		False-Negative(FN)	
Groups	Solution	Group	Solution
crack-shaped contaminants	1, 3	crack hidden behind object	4
overlaid cement paste	3	crack having straight line	2
continuously-distributed concrete pores	2	crack obscured by dark surface	1
edge of linear-shaped construction material	2	crack on the corner of detecting window	4

- Solution 1: Enough illumination
- Solution 2: Consideration of region areas, shapes, and continuities
- Solution 3: Additional vision techniques(stereo vision, IR, etc)
- Solution 4: Moving image FOVs(e.g., using a video stream)

### #3. Image-Based Concrete Crack Detection Using Convolutional Neural network and Exhaustive Search Technique

By Shengyuan Li and Xuefeng Zhao

- A CNN is designed through modifying AlexNet. It helps to detect concrete cracks, concrete spalling, and potholes and cracks taken from a concrete structure may be affected by factors such as light, shadows, and rusty and rough surface in real-world situations.
  - AlexNet is a remarkable CNN for image classification.
- It is trained and validated using a built database with 60000 images.
- This paper explains the operation principle of all layers in AlexNet details like convolution layer, pooling layer, ReLu, LRN, dropout, full connection layer, and softmax layer.

### #4. Deep Learning with Spatial Constraint for Tunnel Crack Detection

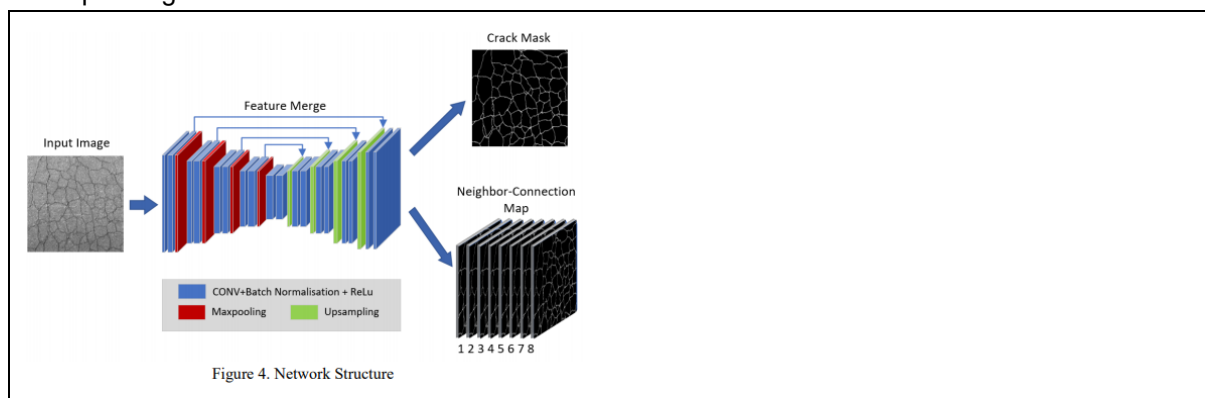
By Li, Ph.D., Qin Zou, Ph.D., Jianghai Liao, Yuanhao Yue and Song Wang, Ph.D

[Background]

: The traditional method required labor intensive by professional workers and time-wasting.

[Network Architecture]

: Unet network, a fully convolutional network is a deep convolutional encoder-decoder architecture designed for pixel-wise semantic segmentation, which contains an encoder network and a corresponding decoder network.



① For image-based crack detection, a larger receptive field obtained by down-sampling convolution feature is useful to overcome the influence of noise, and the decoder part can refine crack edges with higher precision by using the encoded features.

② The encoder network is similar to the convolutional layers in the VGG network, but is constructed with less convolution channel. It consists of the repeated application of two 3x3 convolutions. Each of them is followed by a rectified linear unit(ReLU) and a 2x2 max-pooling operation with a stride of 2 for down-sampling. Different from VGG networks, they double the number of feature channels before the down-sampling step such that the loss of feature information can be reduced.

③ In the decoder part, they use nearest neighbor up-sampling to increase the size of feature and merge corresponding encode layer features using point multiplication to reduce the amount of parameters.

④ At the final layer, a 1x1 convolution is used to map each 32-component feature vector to the crack mask and the 8 neighbor-connection maps.

⑤ After each convolution operation, a batch-normalization step is applied to the feature maps, except for the final convolution layer.

⑥ The number of model parameters are only one-fifth of Unet.

- Equipment used to collect data: line-scan CCD cameras, LED light, infrared thermography and controller mount → These equipment can identify cracks with 0.2mm width at a driving speed of 0-80 km/h.

## #5. Structural Damage Detection using Deep Convolution Neural Network and Transfer Learning

By Chuncheng Feng, Hua Zhang, Shuang Wang, Yonglong Li, Haoran Wang, and Fei Yan

### ① Dataset

- The dataset is labelled as one of 5 types: crack, seepage, spalling, Intact and rebar exposure.
- The dataset is separated into 8:1:1(training: validation:testing)
- With the explosive development of convolution neural networks, some datasets are indispensable, such as ImageNet, COCO and PASCAL VOC.
- This paper used a UAV equipped with an HD camera, a real-time kinematic(RTK)

### ② Convolutional Neural Network

- In this paper, they transfer learning and the Inception-v3 deep learning model to the damage detection task for hydro-junction infrastructure.
- To achieve multiple types of damage detection of hydro-junction infrastructure, they slightly altered the fully connected output layer of the network.
- The advantage of deep convolutional neural networks is that they can automatically extract rich defect image features.
- The DCNN with Transfer Learning includes layers: Convolution Layer, Activation Layer, Pooling Layer and Softmax Layer

Table 2. Detailed Specification of Inception-v3 Network

Type	Kernel size/stride	Input size
Convolution	$3 \times 3/2$	$299 \times 299 \times 3$
Convolution	$3 \times 3/1$	$149 \times 149 \times 32$
Convolution	$3 \times 3/1$	$147 \times 147 \times 32$
Pooling	$3 \times 3/2$	$147 \times 147 \times 64$
Convolution	$3 \times 3/1$	$73 \times 73 \times 64$
Convolution	$3 \times 3/2$	$71 \times 71 \times 80$
Convolution	$3 \times 3/1$	$35 \times 35 \times 192$
Inception module	Three modules	$35 \times 35 \times 288$
Inception module	Five modules	$17 \times 17 \times 768$
Inception module	Two modules	$8 \times 8 \times 1,280$
Pooling	$8 \times 8$	$8 \times 8 \times 2,048$
Linear	Logits	$1 \times 1 \times 2,048$
Softmax	Output	$1 \times 1 \times 1,000$

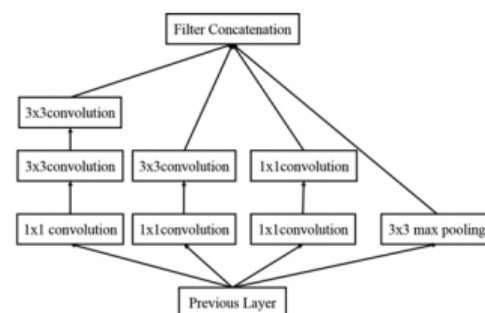


Fig. 6. Sample of the Inception Module

## #6. Crack Detection on a Retaining Wall with an Innovative, Ensemble Learning Method in a Dynamic Imaging System

By Chern-Sheng Lin, Shih-Hua Chen, Che-Ming Chang and Tsu-Wang Shen

**[Keywords]** unmanned vehicle, innovative ensemble learning, cascade classifier, edge feature comparison

### [Summary]

#### ① Dataset Building

- A company vision-based method is often used to automatically detect cracks and record crack information, such as the shape, location, and width.
- The unmanned vehicle can provide a lightweight and remote cruise routine with a Geographic Information System(GIS) sensor, a Gyro sensor, and a charge-coupled device(CCD) camera for the imaging task of the retaining wall in a climbing area or a mountain road.
- The innovation ensemble learning classifier was used to analyze the image and determine the location of the crack for risk assessment

② Canny Detection: In this study, Canny detection was used, because Canny detection is an ideal balance between the detection effect and the amount of computation compared to other edge-detection algorithms.

- The edge detection operation is mainly based on the first-order and second-order derivatives of the image intensity, but the derivative is very sensitive to noise, so the Gaussian filter equation is used to smooth the image in order to reduce noise, to avoid the impact of noise on the test result.
- In order to obtain a characteristic and closed edge with a double threshold, when the image gradient is greater than the high threshold, an image with a small number of false edges and sharp features can be obtained, but the image edges may not be closed.
- set a low threshold and find the point at the edge of the edge that meets the low threshold to complete the image edge.

③ Ensemble Learning: It is the process of bringing together a number of different classifiers, each with different weight, and combining the different weighted classifier results as the final classifier.

→ Local binary patterns(LBP) mark the pixels of an image by thresholding each pixel's neighborhood using the value of the center pixel, and treat the result as a binary number.

$$\text{LBP}(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p - i_c)$$

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases},$$

## #7. DEEP LEARNING AND IMAGE PROCESSING FOR AUTOMATED CRACK DETECTION AND DEFECT MEASUREMENT IN UNDERGROUND STRUCTURES

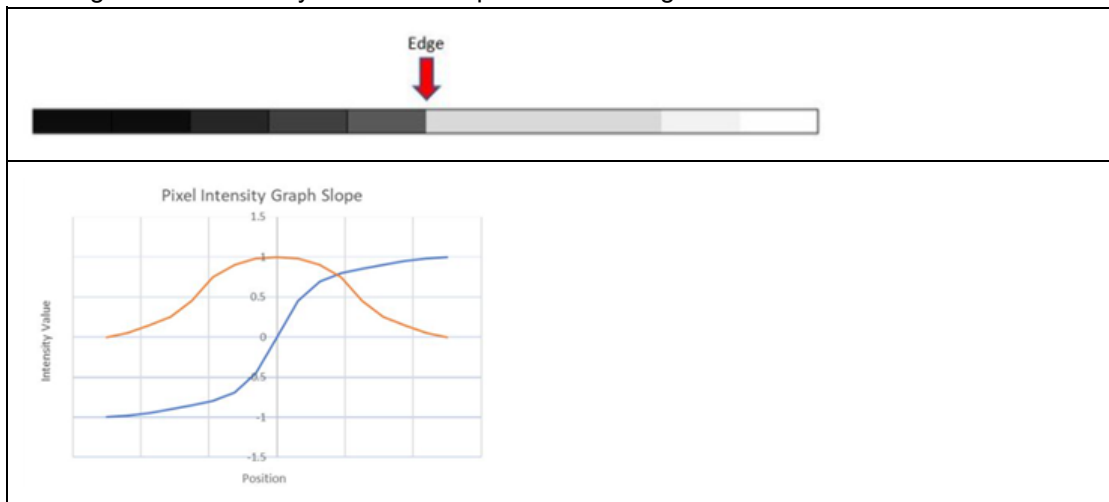
By F. Panella, J. Boehm, Y. Loo, A. Kaushik, D. Gonzalez

### ① Image processing based crack detection

- It propose a high-level image processing based on an initial contrast enhancement to highlight the dark pixel
- A morphological transformations to clean the image from small connected objects, Laplacian of Gaussian as an object detector

### [ Laplacian of Gaussian(LoG) for processing edge on images ]

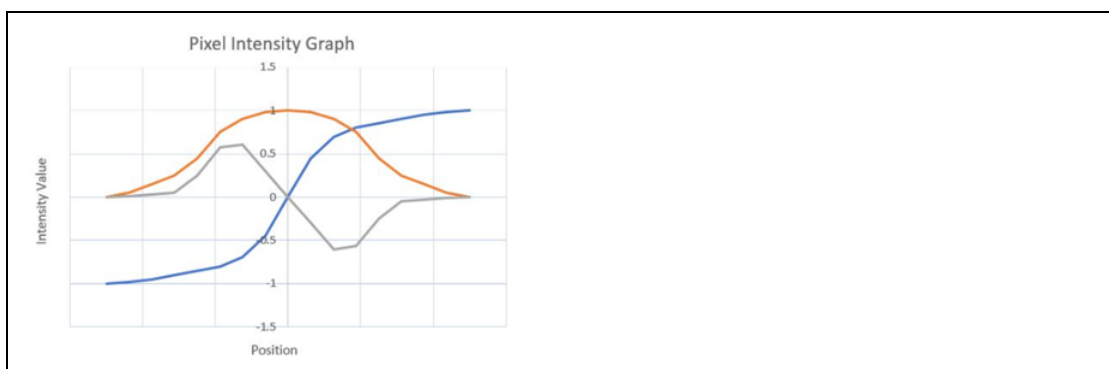
- Edge detection is an important part of image processing and it is used to detect objects, locate boundaries, and extract features. Edge detection is about identifying sudden, local challenges in the intensity values of the pixels in an image.



- For example, for our 10 x 1 pixel image, the blue curve is **a plot of the intensity values**, and the orange curve is **the plot of the first derivative of the blue curve**. In layman's terms, the orange curve is a plot of the slop.

The Orange curve peaks in the middle, so we know that is likely an edge. When we look at the original source image, so confirm that it is an edge.

One limitation with the approach above is that the first derivative of an image might be subject to a lot of noise. Local peaks in the slope of the intensity value might be due to shadows or tiny color changes that are not edges at all



- An alternative to using the first derivative of an image is to use the second derivative, which is the slope of the slope of the first derivative curve(orange). Such a curve looks something like above.

An edge occurs where the graph of the second derivative crosses zero. This second derivative-based method is called the Laplacian algorithm.

For example, a cat hair or whisker might register as an edge because it is an area of a sharp change in intensity. However, it is not an edge. It is just noise. To solve this problem a Gaussian smoothing filter is commonly applied to an image to reduce noise before the Laplacian of Gaussian(LoG).

- Mathematical Formulation of LoG

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

More formally, given a pixel(x, y), the Laplacian L(x, y) of an image with intensity values can be written mathematically as (1).

Just like in the case of the Sobel Operator, we cannot calculate the second derivative directly because pixels in an image are discrete. We need to approximate it using the convolution operator. The two most common kernels like (2) and (3).

Calculating just the Laplacian will result in a lot of noise, so we need to convolve a Gaussian smoothing filter with the Laplacian filter to reduce noise prior to computing the second derivatives. The equation that combines both of these filters is called the Laplacian of Gaussian and is as follows:

$$LoG = -\frac{1}{\pi\sigma^4} \left[ 1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

The above equation is continuous, so we need to discretize it so that we can use it on discrete pixels in an image.

0	1	1	2	2	2	1	1	0
1	2	4	5	5	5	4	2	1
1	4	5	3	0	3	5	4	1
2	5	3	-12	-24	-12	3	5	2
2	5	0	-24	-40	-24	0	5	2
2	5	3	-12	-24	-12	3	5	2
1	4	5	3	0	3	5	4	1
1	2	4	5	5	5	4	2	1
0	1	1	2	2	2	1	1	0

Here is an example of a LoG approximation kernel where  $\sigma = 1.4$ . This is just an example of one convolution kernel that can be used. There are others that would work as well.

→ This Log kernel is convolved with a grayscale input image not to detect the zero crossings of the second derivative. We set a threshold for these zero crossings and retain only those

zero crossings that exceed the threshold. Strong zero crossings are ones that have a big difference between the positive maximum and the negative minimum on either side of the zero crossing. Weak zero crossings are most likely noise, so they are ignored due to the thresholding we apply.

- Difference of Gaussian(DoG) Edge Detection  
This is an alternative that has a shorter computing time than LoG.

#### [Reference for LoG]

Laplacian of Gaussian Filter

<https://automaticaddison.com/how-the-laplacian-of-gaussian-filter-works/>

Laplacian Code on OpenCV

<http://fourier.eng.hmc.edu/e161/lectures/gradient/node8.html>

LoG and DoG Filters

[https://docs.opencv.org/master/d5/db5/tutorial\\_laplace\\_operator.html](https://docs.opencv.org/master/d5/db5/tutorial_laplace_operator.html)

Edge processing

[http://www.cse.psu.edu/~rtc12/CSE486/lecture11\\_6pp.pdf](http://www.cse.psu.edu/~rtc12/CSE486/lecture11_6pp.pdf)

Image Processing on MathWorks

<https://bskyvision.com/132>

<https://kr.mathworks.com/help/images/deep-learning.html>

<https://kr.mathworks.com/help/images/index.html>

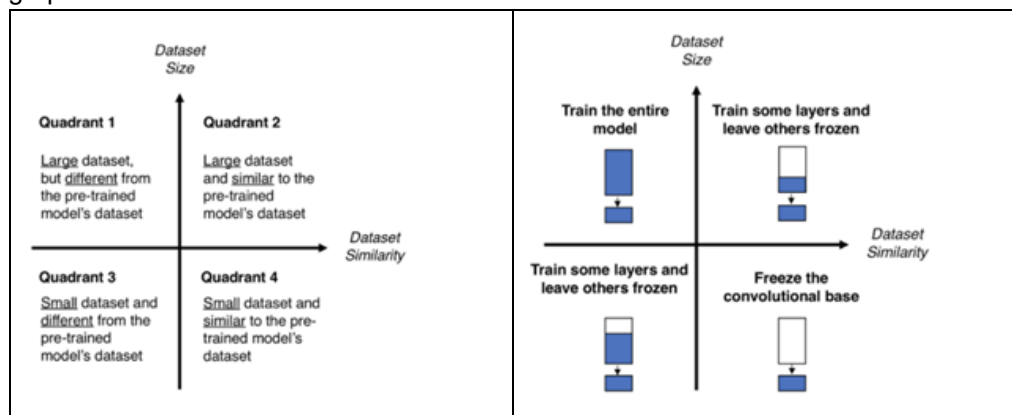
<https://kr.mathworks.com/help/stats/classificationkernel.edge.html>

<https://kr.mathworks.com/help/images/ref/edge.html>

<https://kr.mathworks.com/help/images/ref/fspecial.html?refresh=true>

#### [Transfer Learning]

- The general idea of transfer learning is to use knowledge learned from tasks for which a lot of labelled data is available in settings where only little labelled data is available. Transfer Learning is used to improve prediction accuracy using pre-trained models when building a new model.
- The reason to use it follow:
  - Training a Convolution Network from scratch is not good in terms of time and cost. In most cases, the problems can be solved using models that have already been trained.
  - The more complex the model, the higher the training cost
  - To learn from the beginning, there are many things to consider, such as the number of layers and hyper parameters, and many attempts to increase the performance are required
  - When I already have a well-trained model and need to solve a problem similar to that model, I can get high learning performance by using transfer learning.
- Step for Transfer Learning
  - Choice a pre-trained model
  - Find out where my problem lies in the data size-similarity graph. For example, when the situation needs to classify dogs and cats, the ImageNet is a data size-similarity graph.





Data size-similarity graph	4 Strategy Fine-tuning according to situation
----------------------------	---

**[Reference for Transfer Learning]**<https://velog.io/@chy0428/DL-Transfer-learning>[https://www.datacamp.com/community/tutorials/transfer-learning?utm\\_source=adwords\\_ppc&utm\\_campaignid=1455363063&utm\\_adgroupid=65083631748&utm\\_device=c&utm\\_keyword=&utm\\_matchtype=b&utm\\_network=g&utm\\_adposition=&utm\\_creative=332602034364&utm\\_targetid=dsa-429603003980&utm\\_loc\\_interest\\_ms=&utm\\_loc\\_physical\\_ms=1009866&gclid=EAlaIQobChMliYqcyZ\\_w6wIVSGoqCh25\\_QOBEAAYASAAEgly-D\\_BwE](https://www.datacamp.com/community/tutorials/transfer-learning?utm_source=adwords_ppc&utm_campaignid=1455363063&utm_adgroupid=65083631748&utm_device=c&utm_keyword=&utm_matchtype=b&utm_network=g&utm_adposition=&utm_creative=332602034364&utm_targetid=dsa-429603003980&utm_loc_interest_ms=&utm_loc_physical_ms=1009866&gclid=EAlaIQobChMliYqcyZ_w6wIVSGoqCh25_QOBEAAYASAAEgly-D_BwE)<https://jeinalog.tistory.com/13>

// This is so detail and explains well to understand

<https://keras.io/api/applications/>

// Keras Application are deep learning models that are made available alongside pretrained wright

**[Fine-tuning]**

: It refers to a method of transforming the architecture for a new purpose (according to my image data) based on the model that has been previously trained and updating the training from the model weights that have already been trained. This means to adjust the parameters of a model.

<https://eehoeskrapp.tistory.com/186>**#8. Deep Learning for Detecting Building Defects Using Convolutional Neural Networks****by Husein Perez, Joseph H. M. Tah and Amir Mosavi****[Bayesian Method]**

: Bayesian analysis, a method of statistical inference(named for English mathematician Thomas Bayes) that allows one to combine prior information about a population parameter with evidence from information contained in a sample to guide the statistical inference process. A prior probability distribution for a parameter of interest is specified first. The evidence is then obtained and combined through an application of Bayes' Theorem to provide a posterior probability distribution for the parameter. The posterior distribution provides the basis for statistical inferences concerning the parameter.

$$P(A_i|B) = (P(B|A_i) \times P(A_i)) / (\sum_{i=1}^n (P(B|A_i) \times P(A_i)))$$

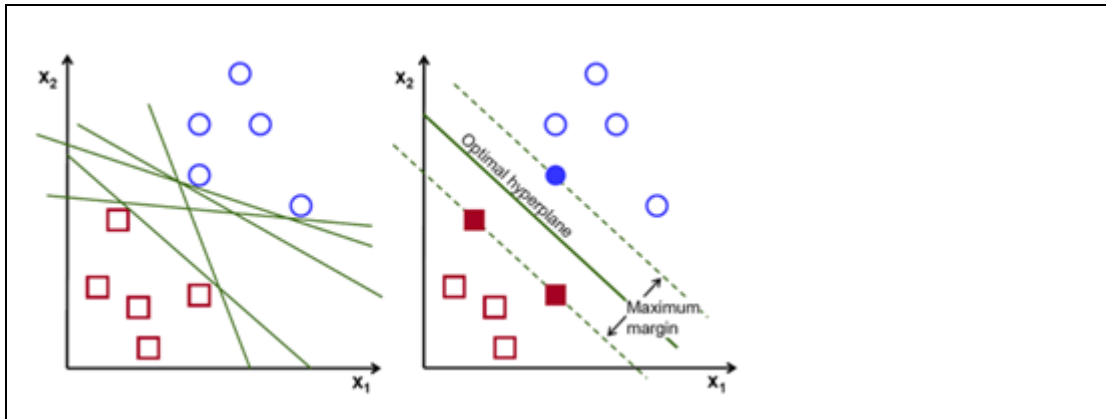
Bayes Theorem

- An important part of Bayesian inference is the establishment of parameters and models. Models are the mathematical formulation of the observed events. Parameters are the factors in the models affecting the observed data.

[http://www.scholarpedia.org/article/Bayesian\\_statistics](http://www.scholarpedia.org/article/Bayesian_statistics)<https://www.britannica.com/science/Bayesian-analysis>**[Support Vector Machine(SVM)]**

: The SVM is a supervised model that changes non-linear problems into linear problems and is used for classification and regression analysis.

- The main objective is to find a hyperplane in an N-dimensional space(N: the number of features) that distinctly classifies the data points like below:



→ To separate the two classes of data points, there are many possible hyperplanes that could be chosen. This objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

→ Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of features is 3, then the hyperplane becomes a two-dimensional plane.

→ SVM's pros and cons

- Nonlinear separation data can be modeled using kernel tricks
- Works smoothly in high cooler spaces (when there are many predictors)
- Effective for text distribution and image distribution
- Avoiding multicollinearity problems
- It takes a lot of time to process large data sets
- Does not directly return probability estimates

→ Linear kernel is almost similar to logistic regression for linear separable data

<https://bioinformaticsandme.tistory.com/304>

<https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>

### [Deep Belief Network(DBN)]

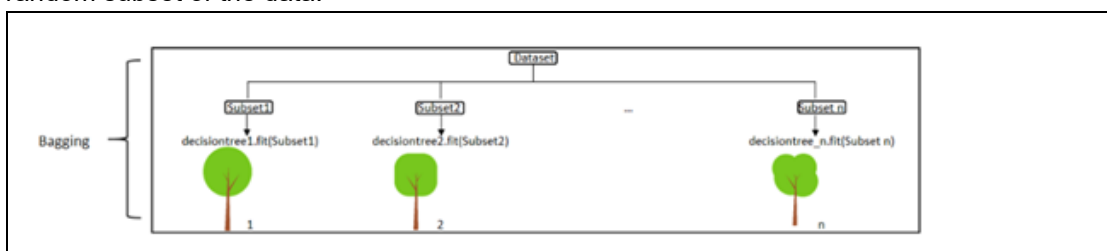
: In machine learning, a deep belief network(DBN) is a generative graphical model, or alternatively a class of deep neural network, composed of multiple layers of latent variables('hidden units'), with connections between the layers but not between units within each layer.

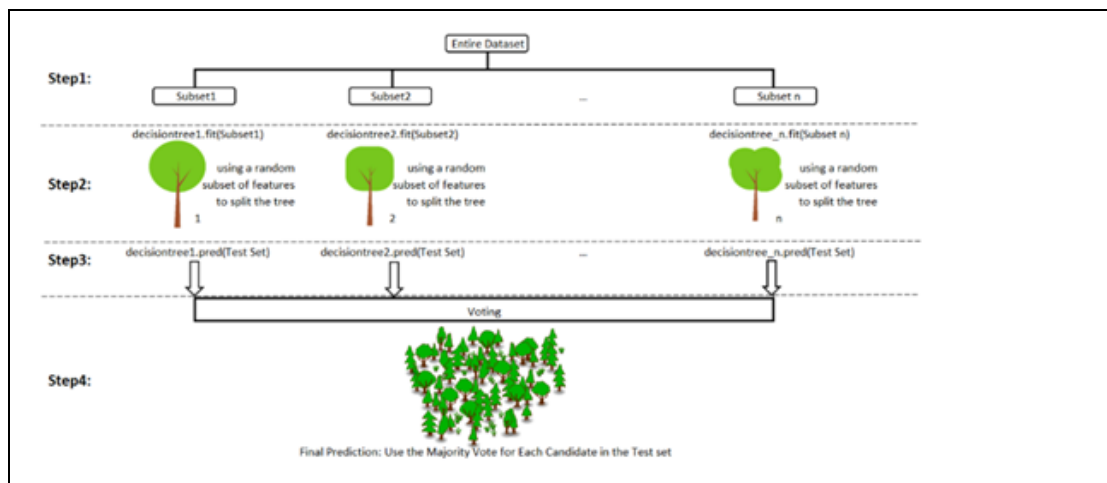
- It just same with Deep Neural Network

### [Ensemble Methods of Random Forest(RF)]

: Random forest is an ensemble model using bagging as the ensemble method and decision tree as the individual model.

- Bagging: Training a bunch of individual models in parallel ways. Each model is trained by a random subset of the data.





Step 1> Select n (e.g 1000) random subsets from the training set

Step 2> Train n (e.g. 1000) decision trees

- one random subset is used to train one decision tree
- the optimal splits for each decision tree are based on a random subset of features(e.g features in total, randomly select 5 out of 10 features to split)

Step 3> Each individual tree predicts the records/ candidates in the test set, independently

Step 4> Make the final prediction

⇒ For each candidate in the test set, Random Forest uses the class(e.g. cat or dog) with the majority vote as this candidate's final prediction.

<https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725>

## [Fuzzy Logic]

: Fuzzy Logic is an approach to computing based on “degrees of truth” rather than the usual “true or false” (1 or 0) Boolean logic on which the modern computer is based.

- It turns out that the useful applications of fuzzy logic are not in high-level artificial intelligence but rather in lower-level machine control, especially in consumer products. Usually, fuzzy controllers are implemented as software running on standard microprocessors. A few special-purpose microprocessors have been built that do fuzzy operation directly in hardware, but even these use digital binary(0 or 1) signals at the lowest hardware level. There are some research prototypes of computer chips that use analog signals at the lowest level, but these chips simulate the operation of neurons rather than fuzzy logic.

<https://www.scientificamerican.com/article/what-is-fuzzy-logic-are-t/>

## #9. Advances in Computer Vision - Based Civil Infrastructure Inspection and Monitoring

By Husein Perez, Joseph H. M. Tah and Amir Mosavi

### ① Difference between Deep Learning and Computer Vision

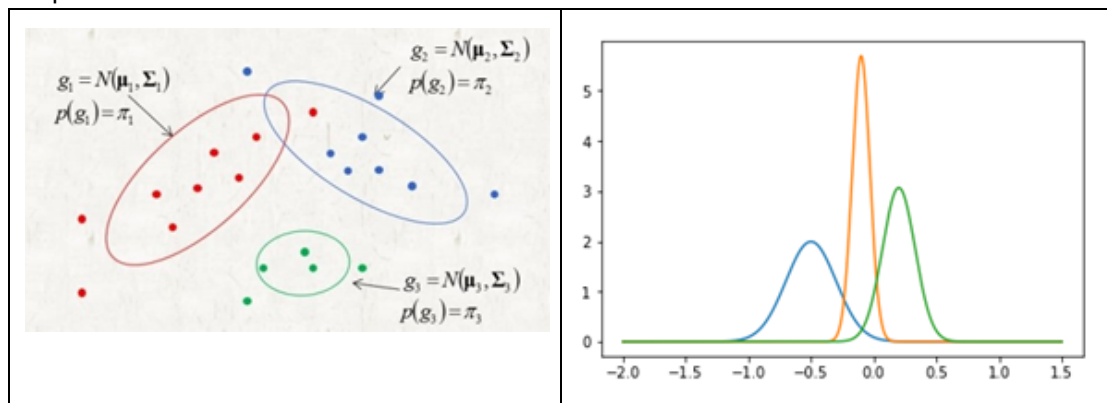
: Computer Vision can find and tell features from images to help discriminate objects and/or classes of objects. The main process of computer vision was extracting the features of the images. Detecting the color, edges, corners and objects were the first step to do when performing a computer vision task. These features are human engineered and accuracy and the reliability of the models directly depend on the extracted features and on the methods used for feature extraction. In the traditional vision scope, the algorithms like SIFT(Scale-Invariant Feature Transform), SURF(Speeded-Up Robust Features), BRIEF(Binary Robust Independent Elementary Features) plays the major role of extracting the features from the raw image.

- Deep Learning: The main difference in the deep learning approach of computer vision is the concept of end-to-end learning. There is no longer a need of defining the features and doing feature engineering.

### ② GMM(Gaussian mixture model)

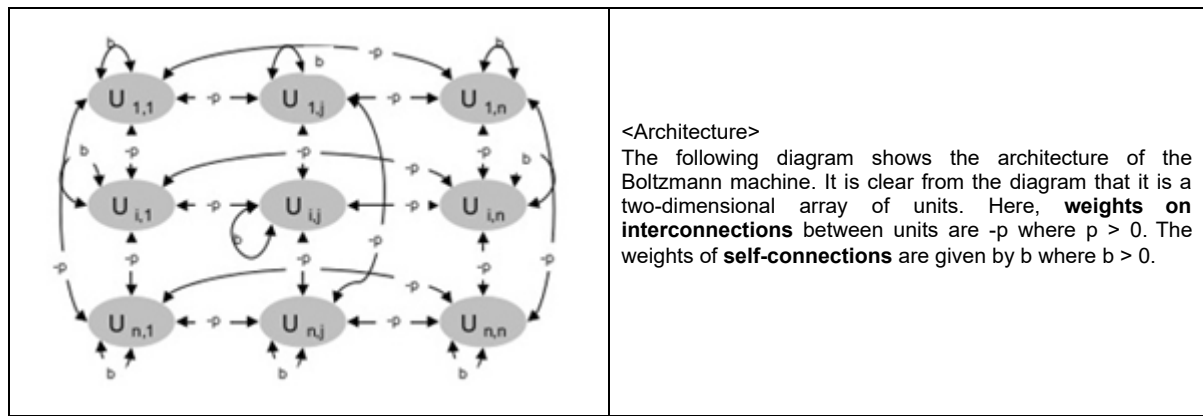
: A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. This model is one of clustering non-parametric methods in unsupervised learning algorithms.

- A Gaussian Distribution is known as a standard distribution well.
- Here is one example. I need to know the average, parameter and weight for 3 normal distributions. A parameter has weight, which indicates where it belongs to one of the 3 normal distributions, and parameter(average and distribution) of each normal distribution.
- When I know the probability to being chosen and the parameter each, it is possible to know where the given data is from Gaussian. Conversely, knowing in which Gaussian the given data is generated, it is possible to estimate the probability of a Gaussian being selected and the parameters of Gaussian.



[http://norman3.github.io/prml/docs/chapter02/3\\_1.html](http://norman3.github.io/prml/docs/chapter02/3_1.html)  
<https://m.blog.naver.com/PostView.nhn?blogId=kmkim1222&logNo=10187825620&proxyReferer=https:%2F%2Fwww.google.com%2F>  
<https://3months.tistory.com/154>  
<https://scikit-learn.org/stable/modules/mixture.html>

③ A Boltzmann machine is a stochastic learning process having recurrent structure and is the basis of the early optimization techniques used in ANN. The main purpose of Boltzmann Machine is to optimize the solution of a problem. It is the work to optimize the weight and quantity related to that particular problem.



[https://www.tutorialspoint.com/artificial\\_neural\\_network/artificial\\_neural\\_network\\_boltzmann\\_machine.htm](https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_boltzmann_machine.htm)

#### ④ Gabor filtering

: In image processing is a linear filter used for texture analysis, which essentially means that it analyzes whether any specific frequency content in the image is in specific directions in a localized region around the point or region of analysis.

- It is a filter that detects the outline and is similar to the reaction of the human visual system.

#### ⑤ speeded-up robust(SURF)

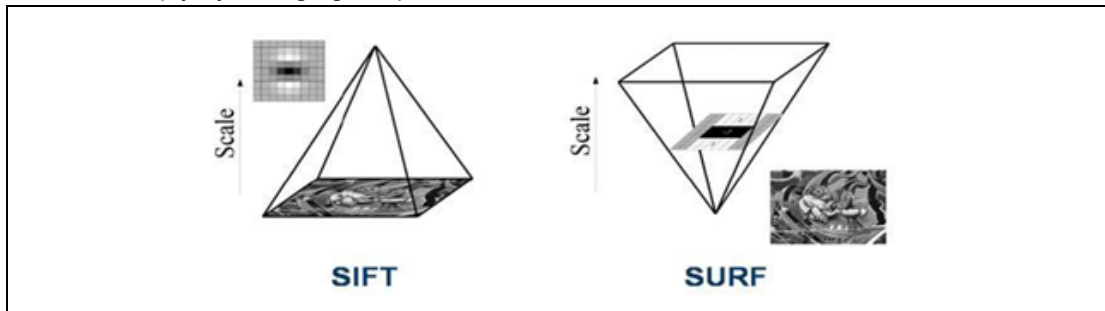
: In computer vision, speeded-up robust(SURF) is a patented local feature detector and descriptor. It can be used for tasks such as object recognition, image registration, classification, or 3D reconstruction. It is partly inspired by the scale-invariant feature transform(SIFT) descriptor. This is a method to solve the computational calculation of SIFT.

- Unlike SIFT, SURF finds an island of interest and a region using an integral image. It can be inferred that the integral of the image is to obtain the width(sum of brightness) of the area of the image. If you create an integral image in this way and then calculate the subtotal of the image, you can calculate it at high speed.

IMAGE					INTEGRAL IMAGE			
0	1	1	1	➔	0	1	2	3
1	2	2	3		1	4	7	11
1	2	1	1		2	7	11	16
1	3	1	0		3	11	16	21

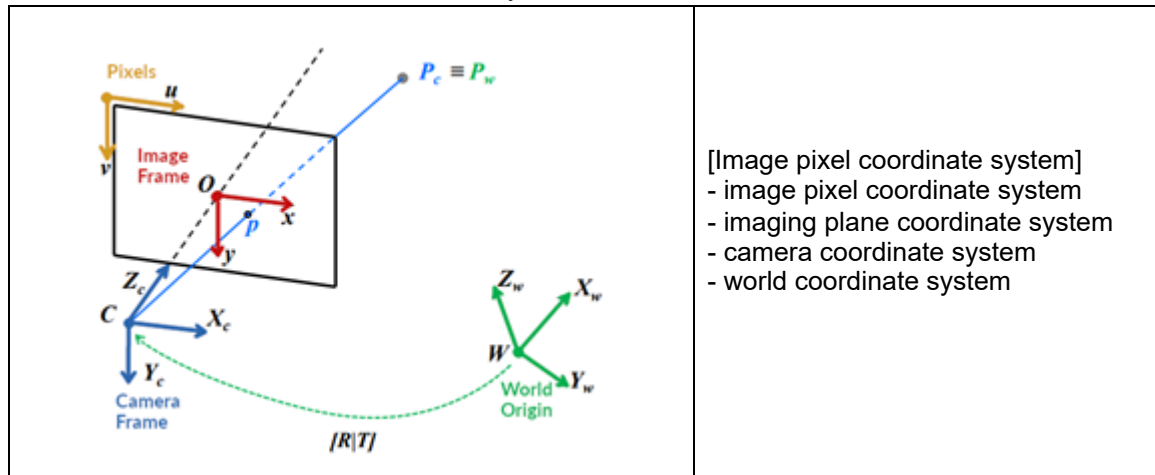
- Fast dimension reduction using reduced detector and descriptor

: SURF extracts points of interest by compacting the image scales and increasing the filter size. Since the integral image is used, it has the efficiency of calculation that can adjust the filter size simply by changing the position.



⑥ SFM (Structure from motion) is to determine the spatial and geometric relationship of the target through the movement of the camera, which is a common method of 3D reconstruction. It only needs an ordinary RGB camera, so the cost is lower, and the environment is less restricted, and it can be used indoors and outdoors.

- Coordinate System: In order to study SFM mathematically, I need a coordinate system. There are mainly two types of coordinate systems in SFM, one is the camera coordinate system, and the other is the world coordinate system.



[Image pixel coordinate system]  
 - image pixel coordinate system  
 - imaging plane coordinate system  
 - camera coordinate system  
 - world coordinate system

<https://towardsdatascience.com/structure-from-motion-311c0cb50e8d>

⑦ A Heuristic method is an approach to finding a solution to a problem that originates from the ancient Greek word 'eurisko', meaning to 'find', 'search', or 'discover'. It is about using a practical method that doesn't necessarily need to be perfect. Heuristic methods speed up the process of reaching a satisfactory solution.

<https://www.toolshero.com/problem-solving/heuristic-method/>

## #10. EFFICIENT DEEP LEARNING-BASED NETWORK FOR CRACK DETECTION IN PIPELINE SYSTEM

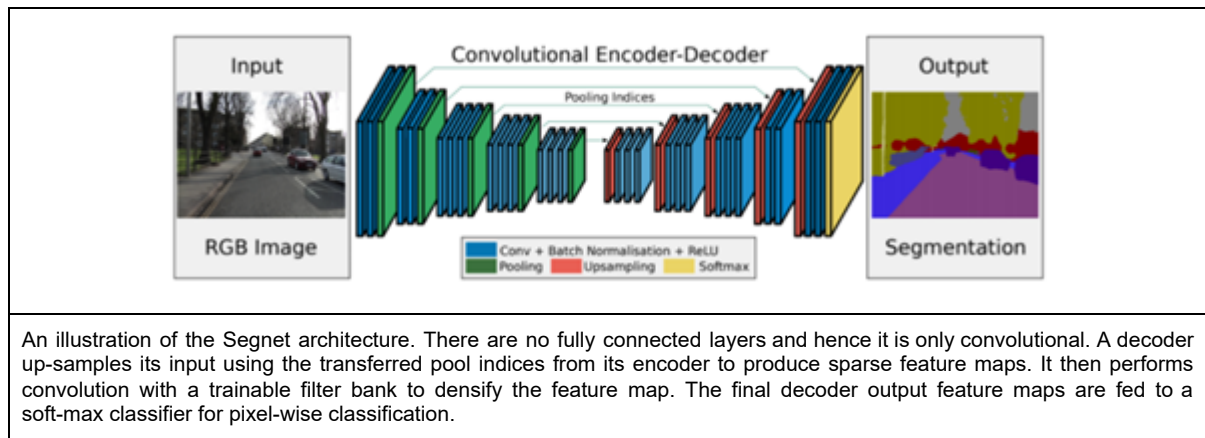
By HOANH NGUYEN, THANH QUYEN NGO

### [Brief Summary]

- This paper proposes a multi-scale deep convolutional network based on SegNet.
- The proposed approach first discards the Softmax layer in original SegNet architecture.
- Then, two enhanced modules are built, which take the convolutional layer before the pooling layer at the first scale and the last scale in the encoder network and last convolutional layer at the corresponding scale in the decoder network as input layers to generate the overall fused layer in the end of the proposed network.
- Furthermore, the focal loss function is adopted to focus on learning the hard examples and down-weighting the numerous easy negatives.
- Experimental results on two public datasets, including CRACK500 and CrackTree, show that our network achieves better results compared to other state-of-the-art methods.

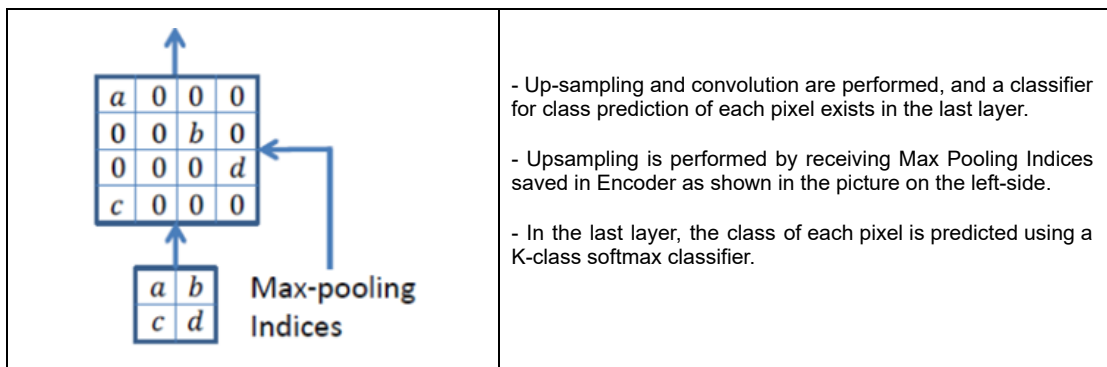
### ① SegNet Network (Semantic segmentation)

: A model was designed for pixel-wise semantic segmentation of structures related to autonomous driving. This has Encoder Network and Decoder Network.



- Encoder Network
  - This network only uses convolution layers of VGG-16 without FCL when it processes convolution and max pooling.
  - Conducting 2 x 2 Max pooling, it gets Max Pooling Indices.

- Decoder Network



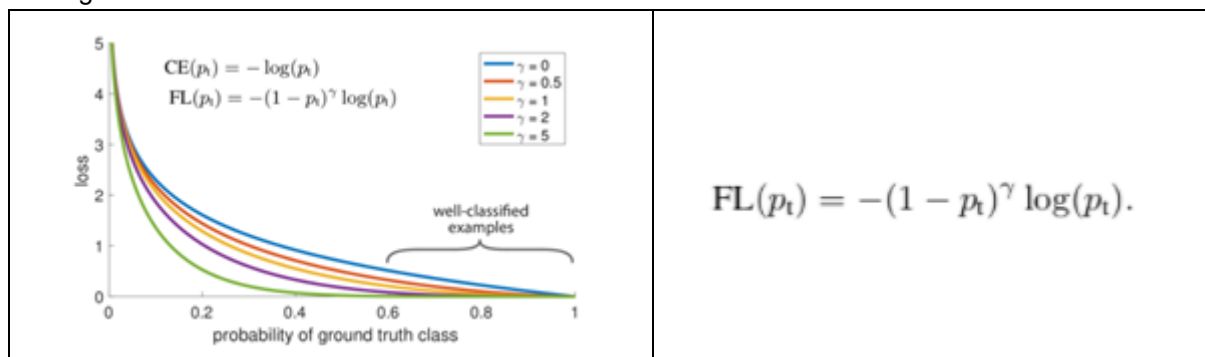
Reusing these Encoder Net's Max Pooling Indexes, that is, like SegNet's Decoder structure, has several advantages.

- I can improve boundary delineation
- The number of parameters that enable end-to-end can be reduced
- This up-sampling scheme can be integrated into any encoder-decoder format

<https://kuklife.tistory.com/120>

## ② Focal loss function

: It improves the performance of simple modified classification in cross-entropy, which was used as the loss function in the classifier in the existing detection algorithm. For classes that are found well(easy negatives), loss is reduced so that loss updates are hardly performed. For classes that are now well found(hard positives), loss is increased by increasing loss. It is gamma( $\gamma$ ) that plays a role in making the difference.





→ Focal loss is the focus of learning more about the class that is now well found.

<https://uk-kim.github.io/2018/12/07/Focal-loss-for-dense-object-detection.html>

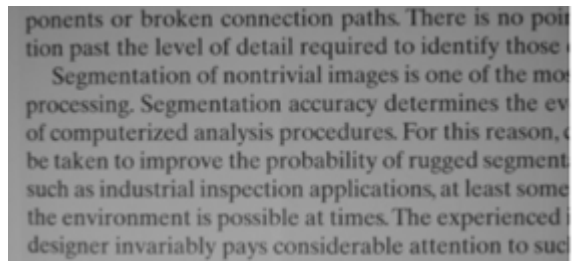
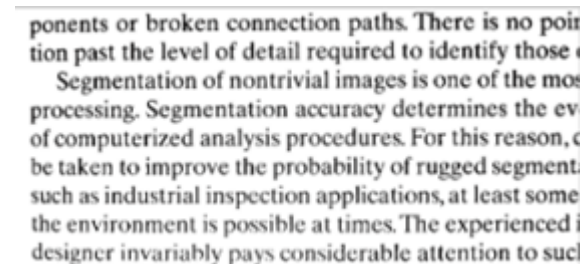
<https://ufris.tistory.com/17>

<https://medium.com/@ayodeleodubela/what-does-focal-loss-mean-for-training-neural-networks-770636f76379>

<https://www.slideshare.net/ssuser06e0c5/focal-loss-detection-classification>

### ③ Otsu thresholding

: In computer vision and image processing, Otsu method, named after Nobuyuki Otsu, is used to perform automatic image thresholding. In the simplest form, **the algorithm returns a single intensity threshold that separates pixels into two classes, foreground and background**. This threshold is determined by minimizing intra-class intensity variance, or equivalently, by maximizing inter-class variance. Otsu's method is a one-dimensional discrete analog of Fisher's Discriminant Analysis, is related to Jenks optimization method, and is equivalent to a globally optimal k-mean performed on the intensity histogram. The extension to multi-level thresholding was described in the original paper, and computationally efficient implementations have since been proposed.

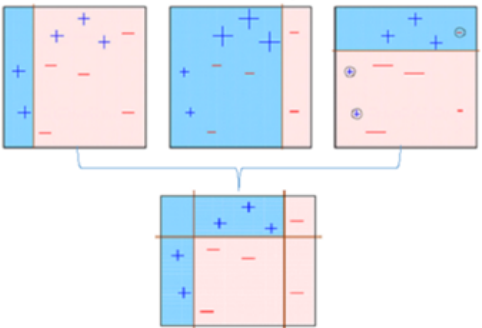
	
original image	image applied Otsu's thresholding

<https://j07051.tistory.com/364>

[https://en.wikipedia.org/wiki/Otsu%27s\\_method](https://en.wikipedia.org/wiki/Otsu%27s_method)

<https://kr.mathworks.com/help/images/ref/graythresh.html?lang=en>

④ Adaboost(Adaptive Boosting) algorithm tries to build a strong learner(predictive model) from the mistakes of several weaker models. Start by creating a model from the training data. Then, creating a second model from the previous one by trying to reduce the errors from the previous model. Models are added sequentially, each correcting its predecessor, until the training data is predicted perfectly to the maximum number of models have been added. It focuses on classification problems and aims to convert a set of weak classifiers into a strong one.

	$F(x) = \text{sign}\left(\sum_{m=1}^M \theta_m f_m(x)\right),$
---	--

→ Although accurate, Random Forest is preferred over AdaBoost, which takes a long time to construct.

<https://dohk.tistory.com/217>

<https://blog.paperspace.com/adaboost-optimizer/>

<https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/>

<https://towardsdatascience.com/boosting-algorithm-adaboost-b6737a9ee60c>



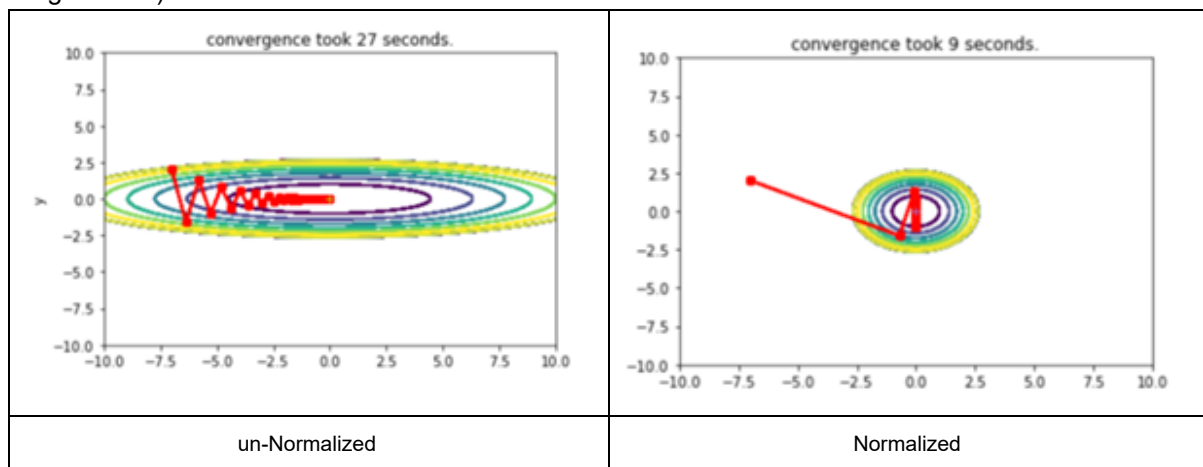
⑤ down-sampling/ up-sampling

: Down-sampling: This is one of sampling methods, when a convolution sample is performed on a 5 x 5 image with 3 x 3, the image size is reduced through convolution as if it is converted to a 3 x 3 image.

- Up-sampling: It means to increase the size of data through sampling. This helps to convert low resolution images into high resolution images generally. Semantic segmentation uses a convolutional layer to extract features from the encoder, and then reconstructs the original image size in the decoder to divide all pixels of the original image.

<http://blog.naver.com/PostView.nhn?blogId=worb1605&logNo=221266339261&parentCategoryNo=&categoryNo=13&viewDate=&isShowPopularPosts=true&from=search>  
<https://zzsza.github.io/data/2018/06/25/upsampling-with-transposed-convolution/>

⑥ Batch Normalization refers to the work of refining(normal distribution) the activation value or the output value. It is said that an experiment is underway to determine whether batch normalization is performed before or after the activation function. This is the task of normalizing the distribution of data in each layer of a neural network. As a method of adding noise, it is normalized for each batch, so that the variance and value of the average for the entire data may vary. Since the activation value/output value is normalized each time it is trained, it is relatively free from the problem of initialization(initial weight value).



<https://sacko.tistory.com/44>  
[https://jsideas.net/batch\\_normalization/](https://jsideas.net/batch_normalization/)  
<https://machinelearningmastery.com/batch-normalization-for-training-of-deep-neural-networks/>  
[https://en.wikipedia.org/wiki/Batch\\_normalization](https://en.wikipedia.org/wiki/Batch_normalization)  
<https://hcnoh.github.io/2018-11-27-batch-normalization>  
<https://m.blog.naver.com/laonple/220808903260>  
<https://shuuki4.wordpress.com/2016/01/13/batch-normalization-%EC%84%A4%EB%AA%85-%EB%B0%8F-%EA%B5%AC%E>  
[D%98%84/](https://shuuki4.wordpress.com/2016/01/13/batch-normalization-%EC%84%A4%EB%AA%85-%EB%B0%8F-%EA%B5%AC%E)

## #11. A Framework for the Segmentation and Classification of 3D Point Clouds using Temporal, Spatial and Semantic Information

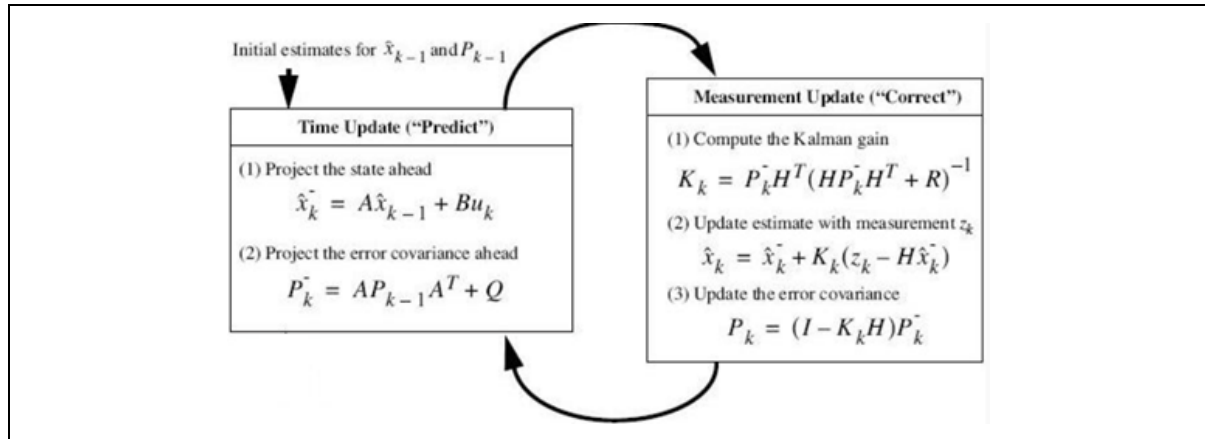
By Mehmet Ali C, ağrı Tuncer and Dirk Schulz

① Kalman filter

: An algorithm that estimates a new function by removing noise contained in the data using past measurement data(a result that is known previously) and new measurement data, which should operate with a linear application. Mathematically predicts the state of the linear system and predicts it while minimizing the error that can be monitored.

There are 2 types of the Kalman filter, which the motion model and measurement model are linear or follow Gaussian distribution.

→ **Motion model** refers to the probability model when the robot moves by receiving motion input from the current position, and the measurement model refers to the model at the position where the robot is located and where it is likely to be using the sensor it is.



The Kalman filter is an algorithm that **repeatedly performs state prediction and measurement updates** and calculates the current position of the robot. **The state prediction step** is a step of predicting the current robot parameter value using the parameters(position, speed, etc) of the previous robot and the robot motion input, and **the measurement update step** is from the current robot parameter value predicted in the state prediction step and the current robot position. This is the step of updating the current robot parameter value using the obtained sensor information.

- State prediction step: The distribution of the current state using the probability distribution calculated from the previous measurement update and the probability distribution of the robot motion input, such as predicting the current robot parameter value using the previous robot parameters(position, speed, etc.) and robot motion input.
- Measurement update step: It is performed by updating the posterior probability distribution using the probability distribution for the predicted current robot position and the probability distribution of observation values measured at the current robot position.

\* **Filter** is a function of filtering out something like a filter of a water purifier, and by filtering out the uncertainty included in the measurement data, it compensates for the insufficient property because the data that can be obtained from the object cannot be perfect. That is, the measurement data or signal accompanies noise, which means an algorithm that selects the desired signal or information.

<http://blog.naver.com/msnayana/80106682874>

simulating kalman filter// <https://kr.mathworks.com/videos/series/understanding-kalman-filters.html>

<https://medium.com/@celinachild/kalman-filter-%EC%86%8C%EA%B0%9C-395c2016b4d6>

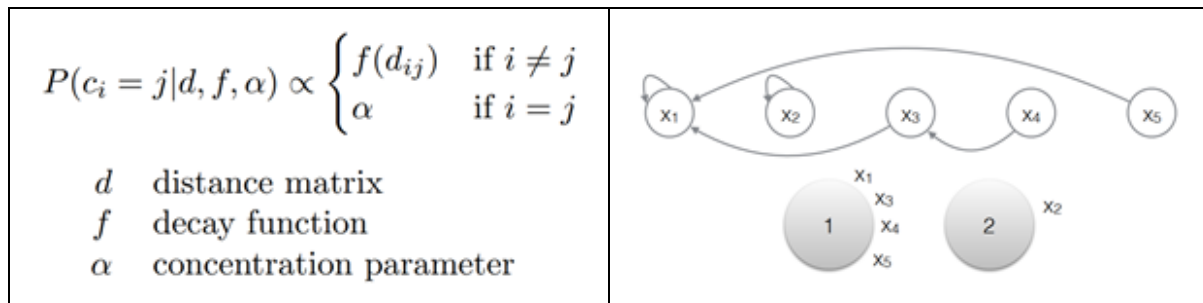
Extended Kalman Filter(for Deeping)// [http://jinyongjeong.github.io/2017/02/14/lec03\\_kalman\\_filter\\_and\\_EKF/](http://jinyongjeong.github.io/2017/02/14/lec03_kalman_filter_and_EKF/)

## ② CRP(Chinese restaurant process)

: The CRP is a distribution on partitions of integers. When used in a mixture model, it provides an alternative representation of a Bayesian nonparametric Dirichlet process mixture.

→ CRP is a stochastic process that generates discrete distribution. Each infinite customer sequence defines a probability distribution over tables. These distributions are infinite dimensional. However, with N data points, only max N tables can be occupied.

③ ddCPR(distance dependent Chinese restaurant process) can naturally obtain spatial structures of the superpixels through a distance measure and a decay function. Also, ddCRP doesn't need to assign the number of superpixels in advance, because the model could infer the parameter from the observed image data.



→ Alternatively, we may want customers to follow other customers they are most similar to. Combine distance and decay into a similarity function. The follower structure defines the seating arrangement. Several follower structures define the same seating arrangement.

→ Sequential ddCRP - all links point backwards.

[http://kodu.ut.ee/~sirts/data/ltg\\_ddcrp.pdf](http://kodu.ut.ee/~sirts/data/ltg_ddcrp.pdf)

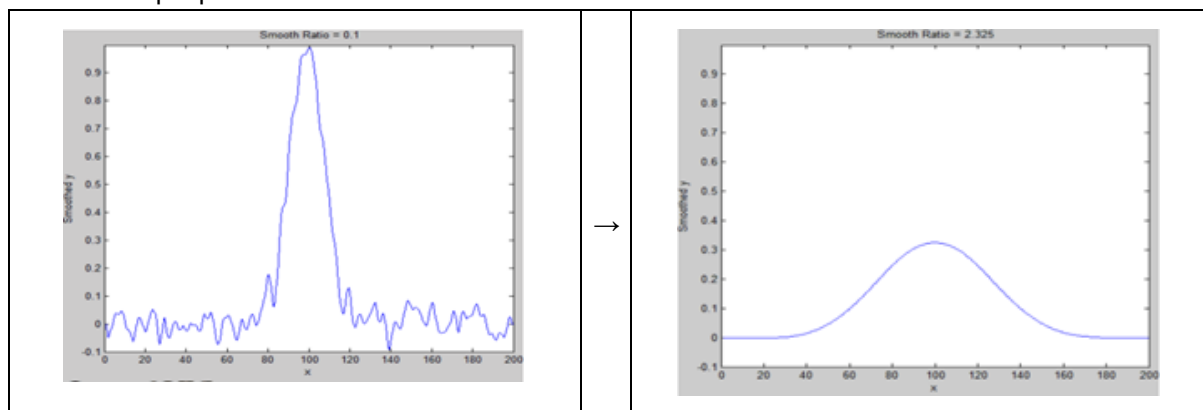
Need to read for CRP//

<http://papers.neurips.cc/paper/4361-spatial-distance-dependent-chinese-restaurant-processes-for-image-segmentation.pdf>

<https://github.com/kristianeschenburg/ddCRP>

#### ④ smoothing process

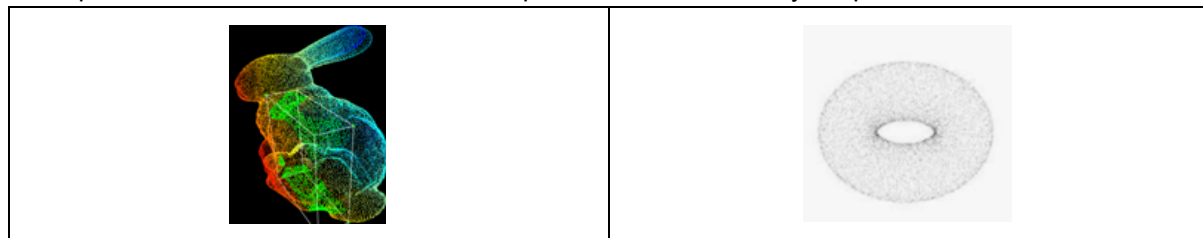
: In statistics and image processing, to smooth a data set is to create an approximating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures/rapid phenomena.



<https://terpconnect.umd.edu/~toh/spectrum/Smoothing.html>

<https://en.wikipedia.org/wiki/Smoothing>

#### ⑤ 3D point cloud data are collections of 3D points located randomly in space.



#### ⑥ Nearest Neighbor(NN) filter

: In pattern recognition, the k-nearest neighbor algorithm is a nonparametric method used for classification or regression. In both cases, the input consists of the k closest training data in the feature classification or as a regression. The output depends on whether k-NN is used as a classification or as a regression.

- In k-NN classification, the output is the belonging item, and the object is the object assigned to the most common item among the k nearest neighbors, and is classified by a majority vote(k is a positive integer and is usually a small number). If k=1, the object is simply assigned to the item of one nearest neighbor.


- In k-NN regression, the output is the property value of the object. This value is the average of the values of k nearest neighbors.

[https://ko.wikipedia.org/wiki/K-%EC%B5%9C%EA%B7%BC%EC%A0%91\\_%EC%9D%B4%EC%9B%83\\_%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98](https://ko.wikipedia.org/wiki/K-%EC%B5%9C%EA%B7%BC%EC%A0%91_%EC%9D%B4%EC%9B%83_%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98)  
[https://ko.wikipedia.org/wiki/%EC%B5%9C%EA%B7%BC%EC%A0%91\\_%EC%9D%B4%EC%9B%83\\_%ED%83%90%EC%83%89#k-%EC%B5%9C%EA%B7%BC%EC%A0%91\\_%EC%9D%B4%EC%9B%83](https://ko.wikipedia.org/wiki/%EC%B5%9C%EA%B7%BC%EC%A0%91_%EC%9D%B4%EC%9B%83_%ED%83%90%EC%83%89#k-%EC%B5%9C%EA%B7%BC%EC%A0%91_%EC%9D%B4%EC%9B%83)

⑦ kD-tree(k dimensional tree) is a spatially divided data structure that structures points in k-dimensional space, and the k-d tree is a useful data structure for application such as search related to multidimensional search keys.

[https://ko.wikipedia.org/wiki/K-d\\_%ED%8A%B8%EB%A6%AC](https://ko.wikipedia.org/wiki/K-d_%ED%8A%B8%EB%A6%AC)

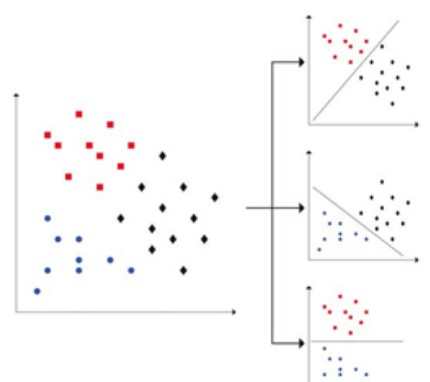
⑧ KITTI dataset takes advantage of their autonomous driving platform Annieway to develop novel challenging real-world computer vision benchmarks.

 <p>360° Velodyne Laserscanner Stereo Camera Rig Monochrome Color GPS</p>	<p>Their tasks of interest are: stereo, optical flow, visual odometry, 3D object detection and 3D tracking. For this purpose, they equipped a standard station wagon with two high-resolution color and grayscale video cameras. Accurate ground truth is provided by a Velodyne laser scanner and a GPS localization system. Their datasets are captured by driving around the mid-size city of Karlsruhe, in rural areas and on highways. Up to 15 cars and 30 pedestrians are visible per image. Besides providing all data in raw format, they extract benchmarks for each task. For each of their benchmarks, they also provide an evaluation metric and this evaluation website. Preliminary experiments show that methods ranking high on established benchmarks such as Middlebury perform below average when being moved outside the laboratory to the real work. Their goal is to reduce this bias and complement existing benchmarks by providing the real world benchmarks with novel difficulties to the community.</p>
--	--

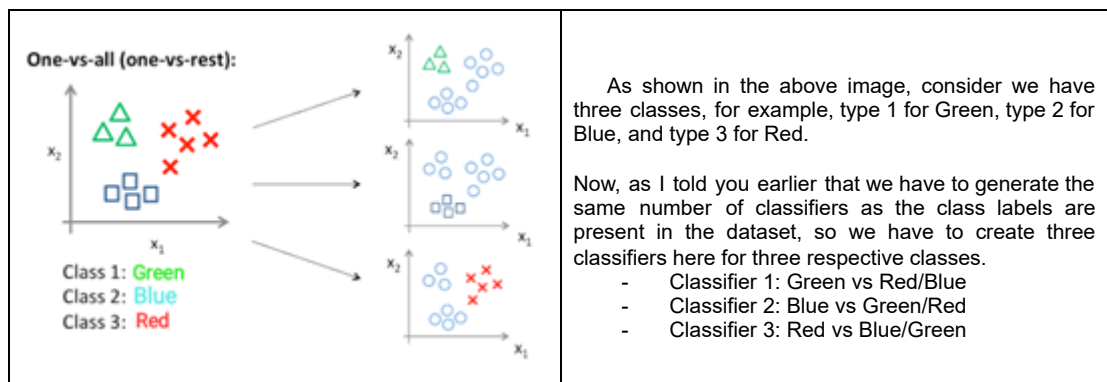
<http://www.cvlibs.net/datasets/kitti/>

⑨ For Multi-Classification

- one-vs-one(OvO) approach is another heuristic method for using binary classification algorithms for multi-class classification. Unlike one-vs-rest that splits it into one binary dataset for each class, the one-vs-one approach splits the dataset into one dataset for each class versus every other class.

 <p>The diagram shows a 2D scatter plot with three classes of data points: red squares, blue circles, and black dots. Three arrows point from this main plot to three smaller subplots. Each subplot shows a binary classification problem between two classes, with a linear decision boundary separating them. The subplots represent: 1) Red vs. Blue, 2) Red vs. Black, and 3) Blue vs. Black.</p>	<p>Taking the example, we have a classification problem having three types: Green, Blue, and Red (N=3).</p> <p>We divide this problem into <math>N(N-1)/2=3</math> binary classifier problem:</p> <ul style="list-style-type: none"> <li>- Classifier 1: Green vs Blue</li> <li>- Classifier 2: Green vs Red</li> <li>- Classifier 3: Blue vs Red</li> </ul> <p>Each binary classifier predicts one class label. When we input the test data to the classifier, then the model with the majority counts is concluded as a result.</p>
---	---

- one-vs-all(OvA, OvR, One-vs-Rest) approach: In one-vs-all classification, for the N-class instances dataset, we have to generate the N-binary classifier models. The number of class labels present in the dataset and the number of generated binary classifiers must be the same.



<https://machinelearningmastery.com/one-vs-rest-and-one-vs-one-for-multi-class-classification/>  
<https://daeson.tistory.com/183>  
<https://towardsdatascience.com/multi-class-classification-one-vs-all-one-vs-one-94daed32a87b>

## #12. A MultiPath Network for Object Detection

**By Sergey Zagoruyko, Adam Lerer, Tsung-Yi Lin, Pedro O. Pinheiro, Sam Gross, Soumith Chintala, Piotr Dollár**

### ① R-CNN(Region proposals with CNN)

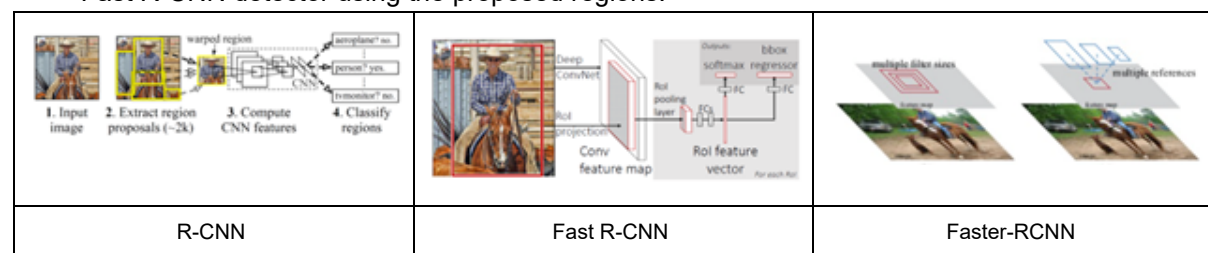
: The purpose is to find out what objects are in the image. First, about 2000 subimages are extracted from the input image and the part where objects are likely to exist is cut out. And after converting each image into an appropriate size to pass through ConvNet, classification is performed through CNN. After that, classify each object using SVM. In addition, it uses a selective search algorithm to create a bounding box at the location where an object exists, and adjusts the box location using a linear regression model.

→ Except for the selection of region proposals, simple classification is very slow, mAP(mean Average Precision) is not high, and performance is poor.

② Fast R-CNN: As an improvement of R-CNN, training is performed at once using mAP and multi-task loss higher than R-CNN. In addition, in training, all network layers are updated and feature caching disk storage space is not required.

③ Faster-RCNN's goal is to create a model with a structure that learns RPN(Region Proposal Network) itself without selective search.

- Deep fully convolutional network suggesting region
- Fast R-CNN detector using the proposed regions.



<https://leechamin.tistory.com/221>  
<https://curt-park.github.io/2017-03-17/faster-rcnn/>  
<https://zszs.github.io/data/2018/05/09/Faster-RCNN-review/>

### #13. Deep Learning with Spatial Constraint for Tunnel Crack Detection

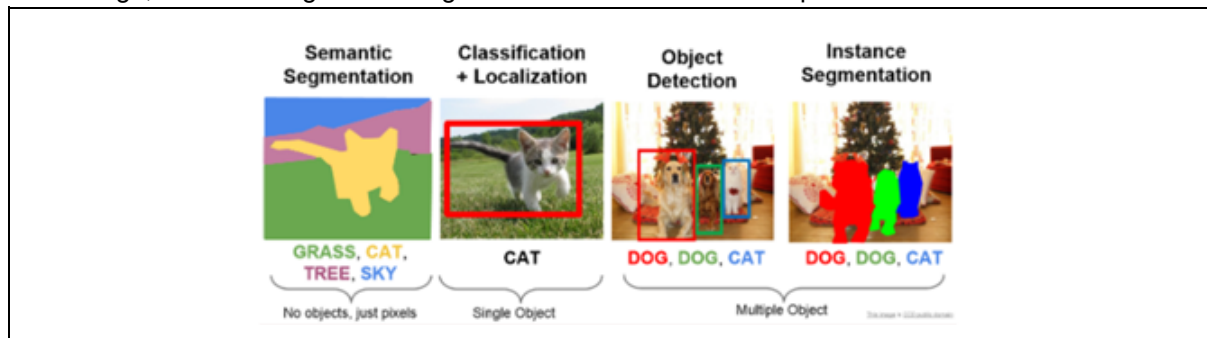
By Qingquan Li, Ph.D., Qin Zou, Ph.D., Jianghai Liao, Yuanhao Yue and Song Wang, Ph.D.

→ How spatial constraints ensure crack continuity.

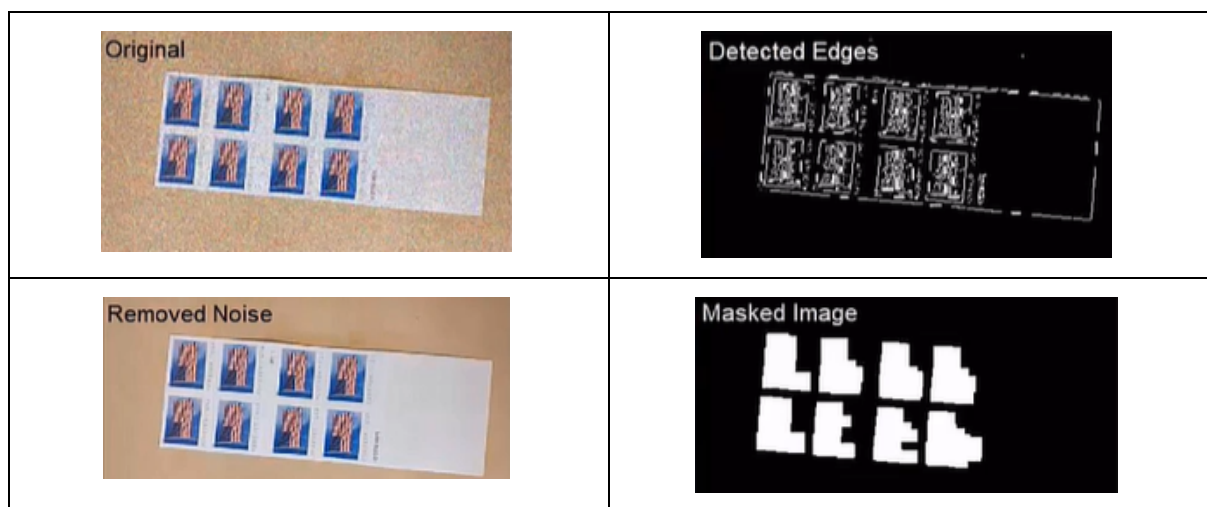
#### ① spatial constraint

- Introduce to overcome the problem that line structures are not well modeled in the traditional deep models
- Even if cracks in the original image have low continuity, the detection output is made into a continuous line structure

② threshold segmentation technique is the process of classifying each pixel belonging to a particular label. It doesn't differ across different instances of the same object. For example if there are 2 cats in an image, semantic segmentation gives the same label to all the pixels of both cats.



<https://nanonets.com/blog/semantic-image-segmentation-2020/>



- Pre-process images using enhancement and filtering techniques
- Separate object of interest using segmentation techniques
- Test your algorithm on large sets of images

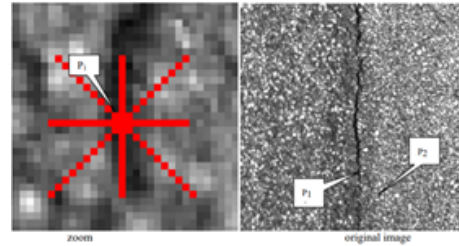
[http://homes.di.unimi.it/ferrari/lmgProc2011\\_12/EI2011\\_12\\_16\\_segmentation\\_double.pdf](http://homes.di.unimi.it/ferrari/lmgProc2011_12/EI2011_12_16_segmentation_double.pdf)

<https://kr.mathworks.com/videos/image-processing-made-easy-81718.html>

#### ③ Conditional Texture Anisotropy(CTA)

: The main idea is to find out a measure which has a different value in one direction(along the orientation of crack) between other directions. This method takes into account the three characteristics: gray intensities, connectivity and orientation of cracks during segmentation steps.

$$CTA(X^I) = \frac{\max_j \{p(x_j^I / l \in w_l)\} - \min_j \{p(x_j^I / l \in w_l)\}}{\max_j \{p(x_j^I / l \in w_l)\}}$$



- CTA takes values close to 0 for the free defect region. In this case, feature values extracted according to each orientation are close together. Around defects, CTA will take values close to 1. For example, along crack orientation, feature values will be very different from others directions.

<https://pdfs.semanticscholar.org/2a18/f417e32b2690e67ef3f0b40deb679d40a288.pdf>

#### ④ low-level feature/ high-level feature

- If one feature can extract the local specifications of the object, we call it “low-level feature”, and if it can extract the global specifications of the object, it is called a “high-level feature”.
- low-level feature: edges, circles, etc
- high-level features: eyes, nose, mouth, etc. High-level features are composed of low-level features.
- object-level: head, etc.

[http://cbl.eng.cam.ac.uk/pub/Intranet/MLG/ReadingGroup/cnn\\_basics.pdf](http://cbl.eng.cam.ac.uk/pub/Intranet/MLG/ReadingGroup/cnn_basics.pdf)

[https://www.researchgate.net/figure/1-Characteristics-and-examples-of-the-low-level-and-high-level-features-that-represent\\_tbl1\\_249862831](https://www.researchgate.net/figure/1-Characteristics-and-examples-of-the-low-level-and-high-level-features-that-represent_tbl1_249862831)