CHULA ΣNGINEERING
Innovation toward Sustainability | ACTNOW

DEEP LEARNING FOR REMOTE SENSING

Geospatial Programming

Modern Integrated Surveying Technologies 2024

Thepchai Srinoi

Master Degree Student and Teaching Assistant,

Department of Survey Engineering Chulalongkorn University

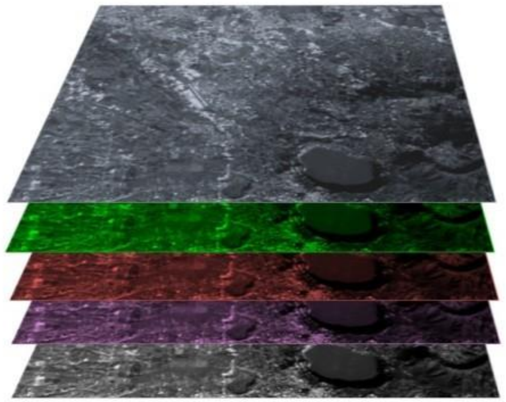
Artificial Neural Network

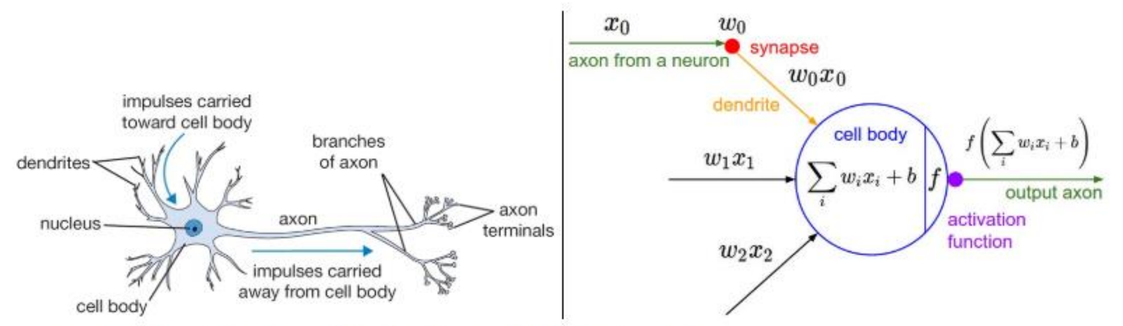
Band Selection

Convert to one vector

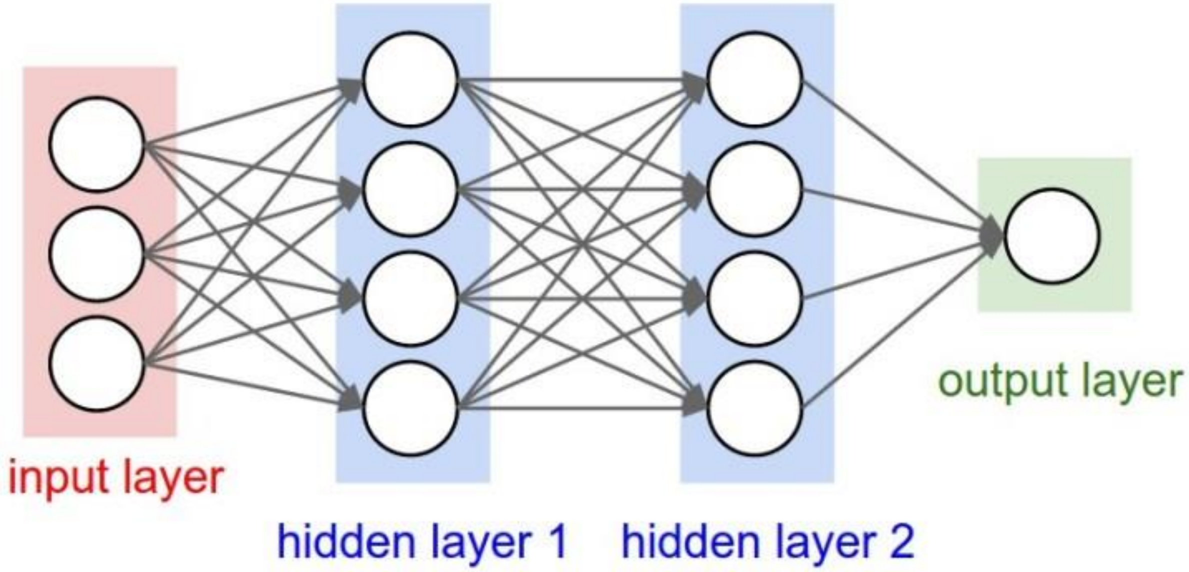
Field Data

Satellite Imagery





A cartoon drawing of a biological neuron (left) and its mathematical model (right).

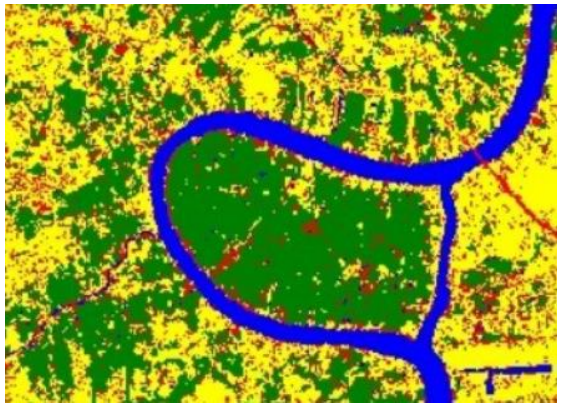


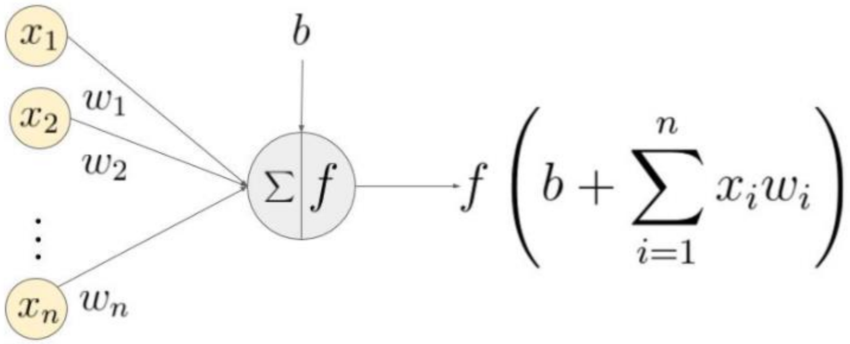
Classification Result

Estimation Result

Single Value

Raster of Result Value

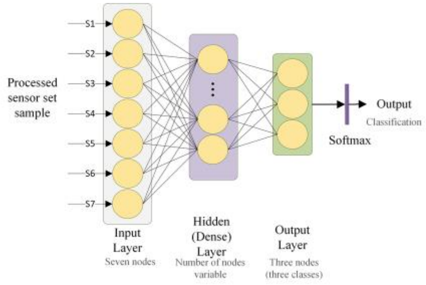


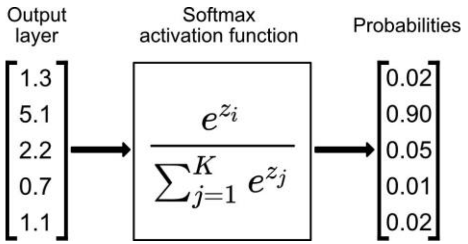


ACTIVATE FUNCTION

An example of a neuron showing the input (x_1 - x_n), their corresponding weights (w_1 - w_n), a bias (b) and the activation function f applied to the weighted sum of the inputs.

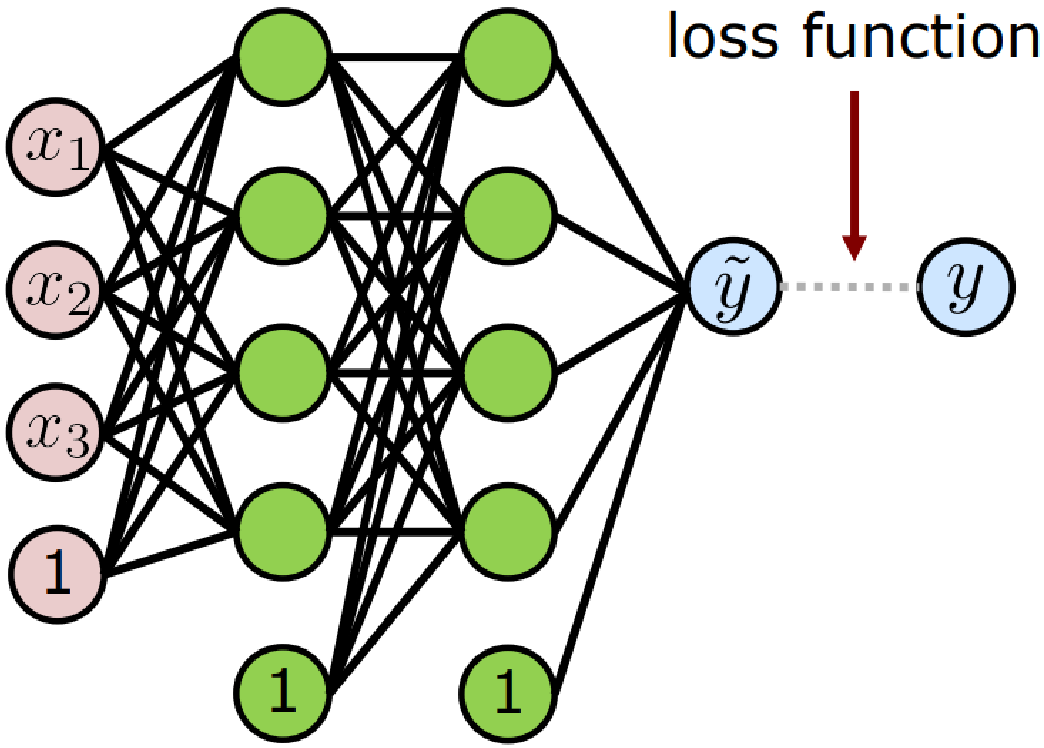
Problem Type	Output Type	Final Activation Function	Loss Function
Regression	Numerical value	Linear	Mean Squared Error (MSE)
Classification	Binary outcome	Sigmoid	Binary Cross Entropy
Classification	Single label, multiple classes	Softmax	Cross Entropy
Classification	Multiple labels, multiple classes	Sigmoid	Binary Cross Entropy





Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus	$\phi(z) = \ln(1 + e^z)$	Multi-layer Neural Networks	

Artificial Neural Network : Loss Function



$$\tilde{y} = f(\boldsymbol{x}, \boldsymbol{w}) = \mathcal{F}_1\left(W_1^{\mathrm{T}} \cdot \mathcal{F}_0\left(W_0^{\mathrm{T}} \cdot \boldsymbol{x}\right)\right)$$

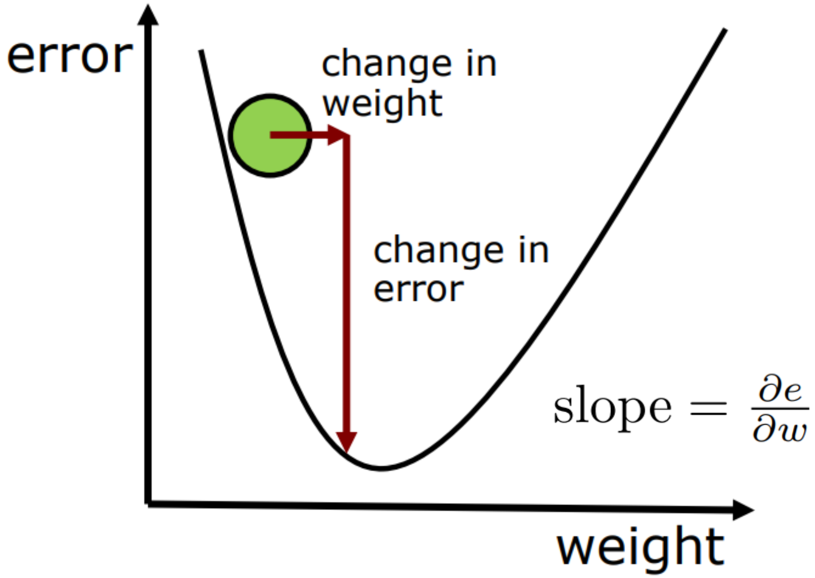
DailyDoseofDS.com

10 Most Common Loss Functions in Machine Learning

by Avi Chawla

Loss Function Name	Description	Function
Regression Losses		
Mean Bias Error	Captures average bias in prediction. But is rarely used for training.	$\mathcal{L}_{MBE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i))$
Mean Absolute Error	Measures absolute average bias in prediction. Also called L1 Loss.	$\mathcal{L}_{MAE} = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - f(x_i)\rvert$
Mean Squared Error	Average squared distance between actual and predicted. Also called L2 Loss.	$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i))^2$
Root Mean Squared Error	Square root of MSE. Loss and dependent variable have same units.	$\mathcal{L}_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i))^2}$
Huber Loss	A combination of MSE and MAE. It is parametric loss function.	$\mathcal{L}_{\text{Huberloss}} = \begin{cases} \frac{1}{2}(y_i - f(x_i))^2 & : \lvert y_i - f(x_i)\rvert \leq \delta \\ \delta(\lvert y_i - f(x_i)\rvert - \frac{1}{2}\delta) & : otherwise \end{cases}$
Log Cosh Loss	Similar to Huber Loss + non-parametric. But computationally expensive.	$\mathcal{L}_{\text{LogCosh}} = \frac{1}{N}\sum_{i=1}^{N} log(cosh(f(x_i) - y_i))$
Classification Losses (Binary + Multi-class)		
Binary Cross Entropy (BCE)	Loss function for binary classification tasks.	$\mathcal{L}_{BCE} = \frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(x_i)) + (1 - y_i) \cdot log(1 - p(x_i))$
Hinge Loss	Penalizes wrong and right (but less confident) predictions. Commonly used in SVMs.	$\mathcal{L}_{\text{Hinge}} = max(0, 1 - (f(x) \cdot y))$
Cross Entropy Loss	Extension of BCE loss to multi-class classification.	$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}.log(f(x_{ij}))$ $N : samples; M : classes$
KL Divergence	Minimizes the divergence between predicted and true probability distribution	$\mathcal{L}_{KL} = \sum_{i=1}^{N} y_i \cdot log(\frac{y_i}{f(x_i)})$

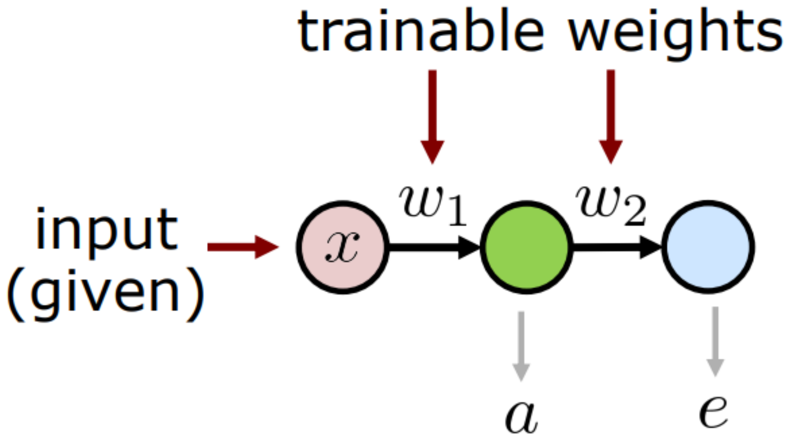
Gradient Decent and Backpropagation



$$\boldsymbol{w}_{\mathrm{up}} = \boldsymbol{w} - \nu \cdot \nabla e(\boldsymbol{w})$$

LEARNING RATE

Backpropagation



forward pass

$$a = x \cdot w_1$$

$$e = x \cdot w_1 \cdot w_2 \longrightarrow \text{Min.}$$

backward pass

$$\frac{\partial e}{\partial w_2} = x \cdot w_1$$

$$\boxed{\frac{\partial e}{\partial a} = w_2}$$

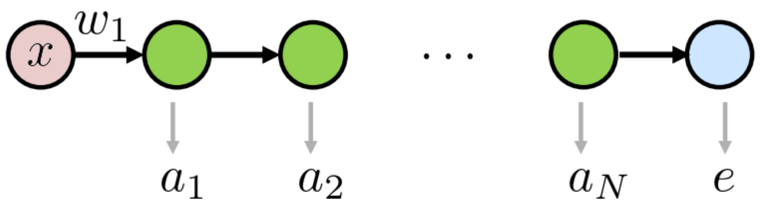
$$\boxed{\frac{\partial a}{\partial w_1} = x}$$

$$\frac{\partial e}{\partial w_1} = \boxed{x} \cdot \boxed{w_2}$$

$$\frac{\partial e}{\partial w_1} = \boxed{\frac{\partial a}{\partial w_1}} \cdot \boxed{\frac{\partial e}{\partial a}}$$

Backpropagation Neural Network

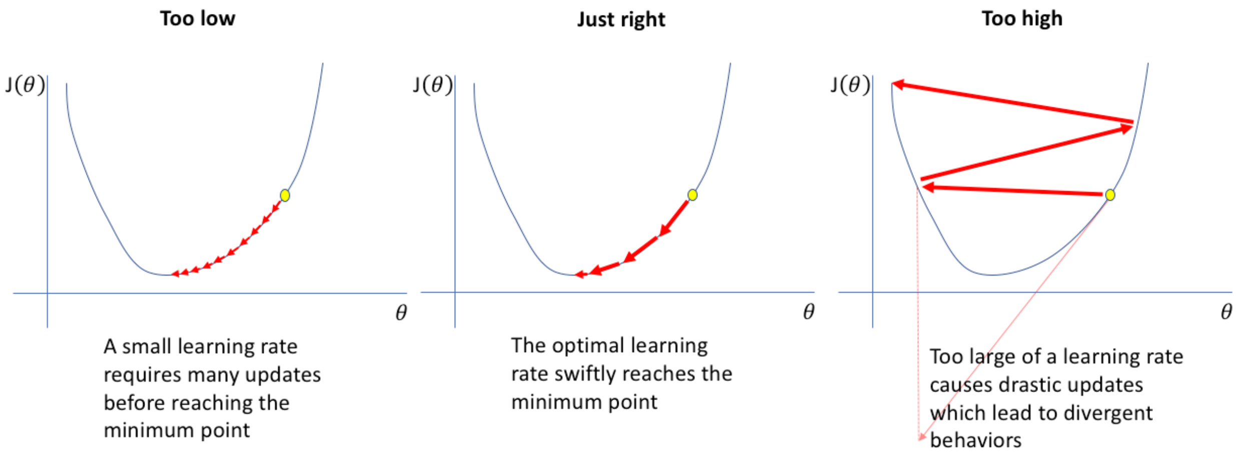
Chaining

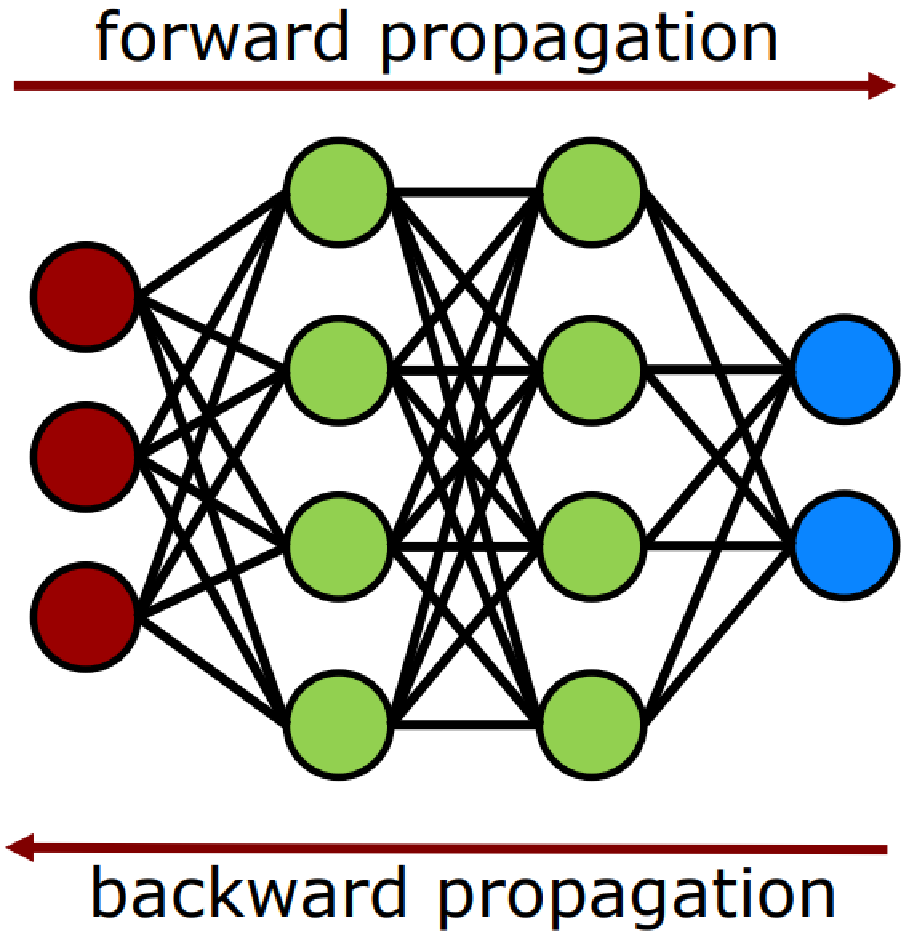


$$\frac{\partial e}{\partial w_1} = \frac{\partial a_1}{\partial w_1} \cdot \frac{\partial a_2}{\partial a_1} \cdot \ldots \cdot \frac{\partial e}{\partial a_N}$$

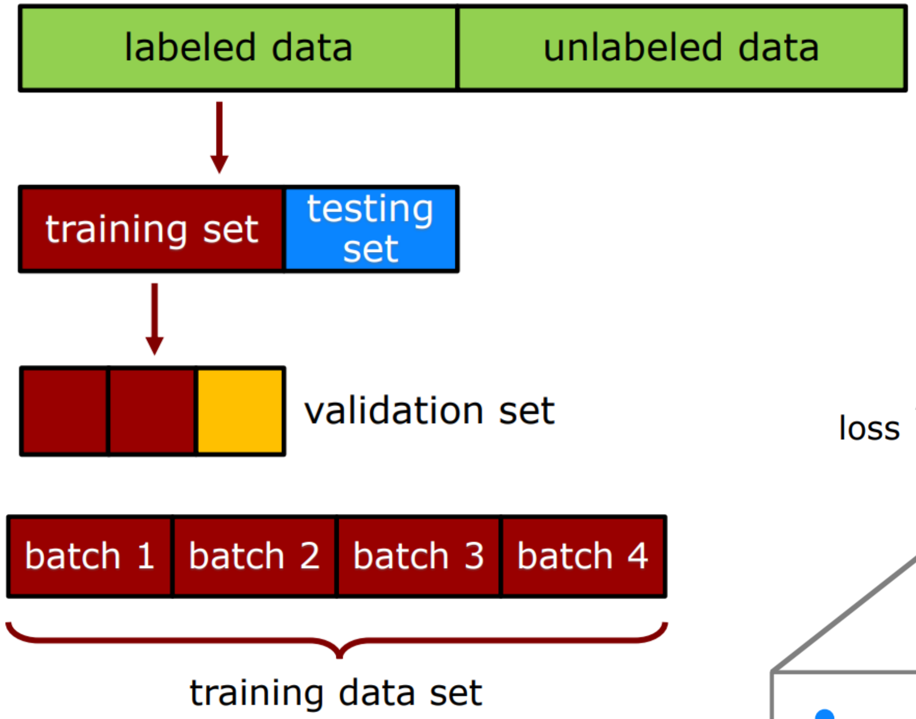
$$\boxed{\boldsymbol{w}_{\text{up}} = \boldsymbol{w} - \nu \cdot \nabla e(\boldsymbol{w})}$$

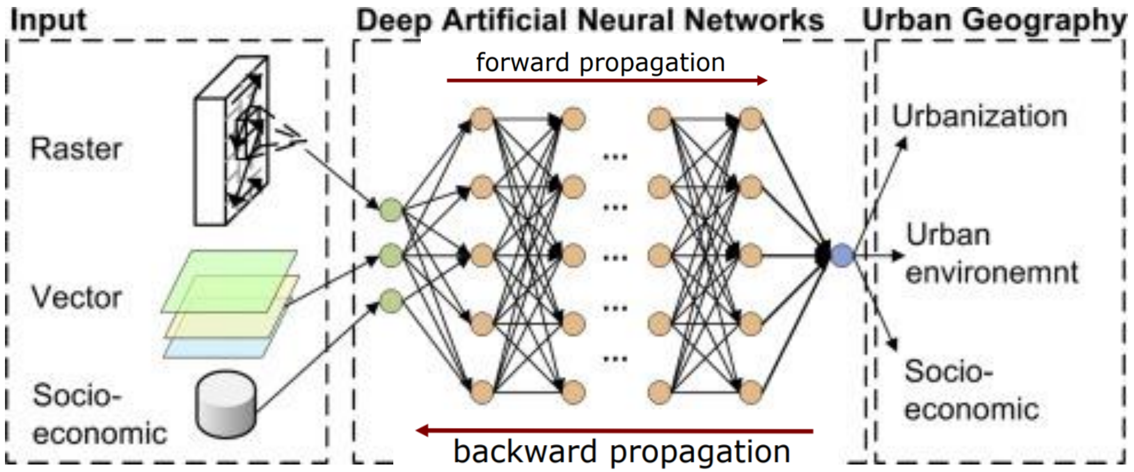
LEARNING RATE

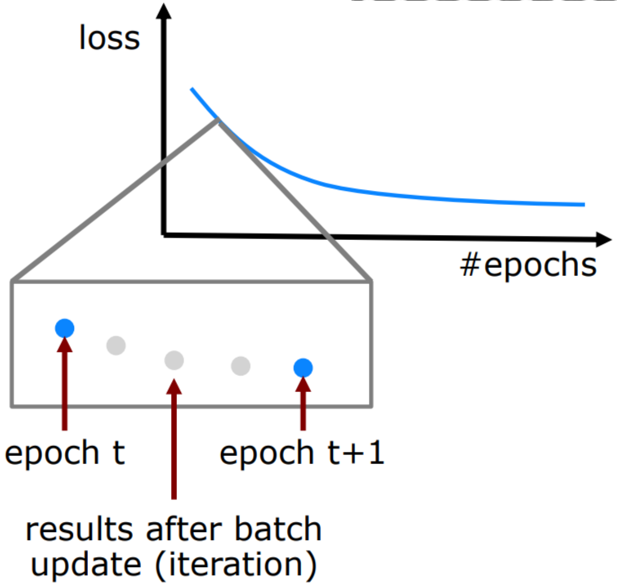


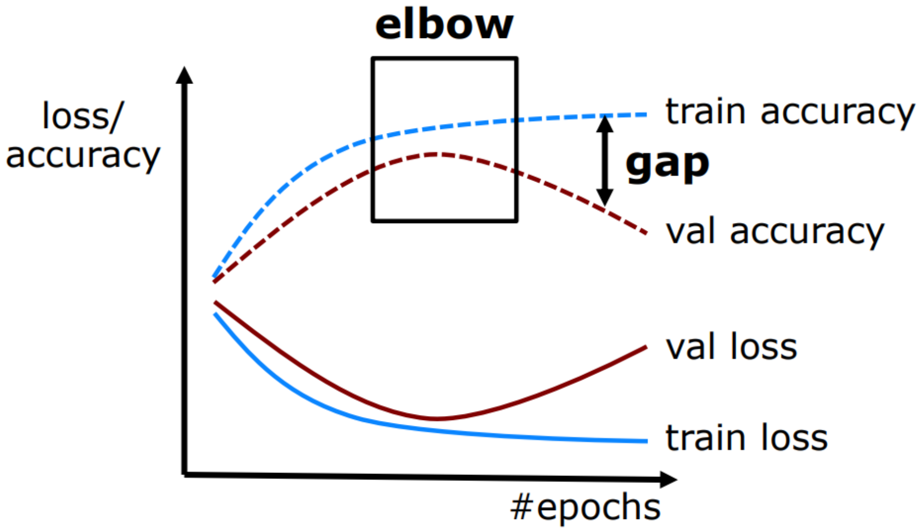


Neural Network Training

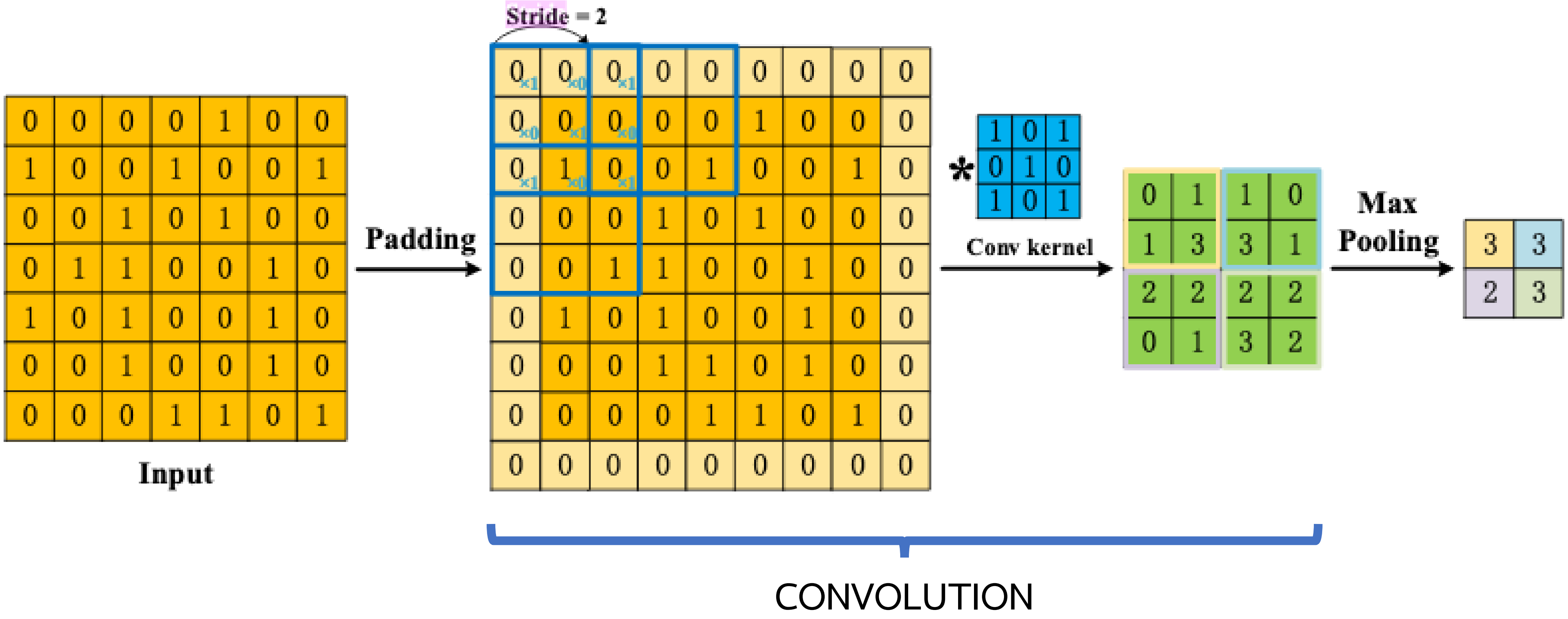




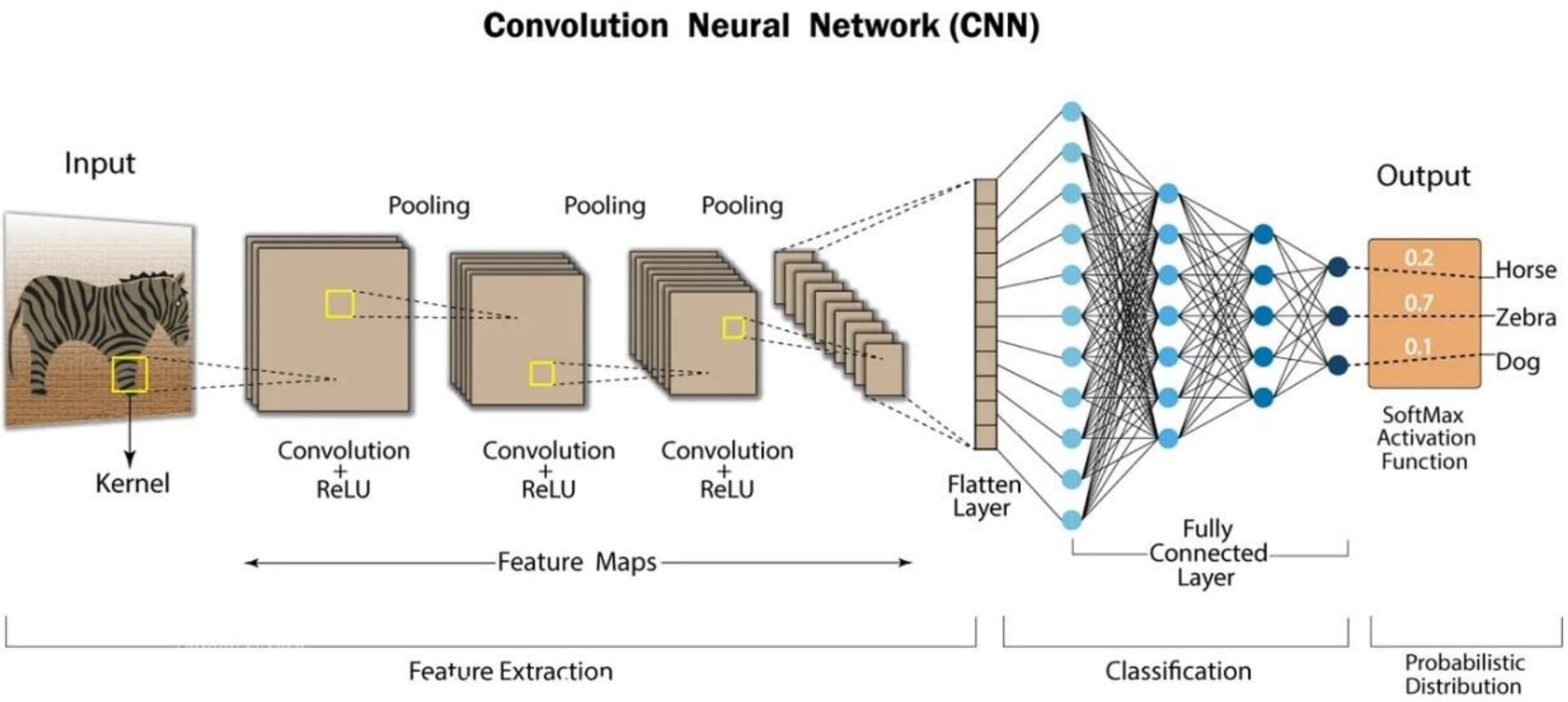




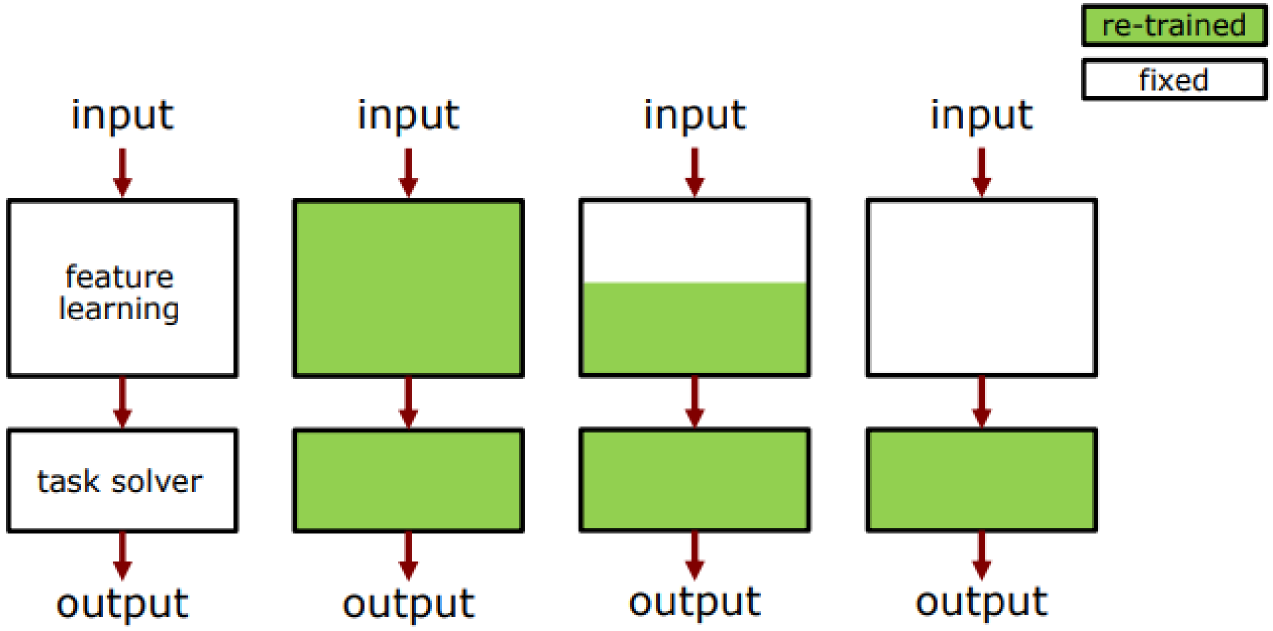
เราไม่ถอดทีละแถว/หลัก เป็นเวกเตอร์เดี่ยวยาวๆ โยนเข้า Deep Neural Network (DNN)
แต่เราจะเอาภาพมาปรับก่อน ด้วย Kernel Filter แล้วโยนเข้า DNN Fully Connected Layer



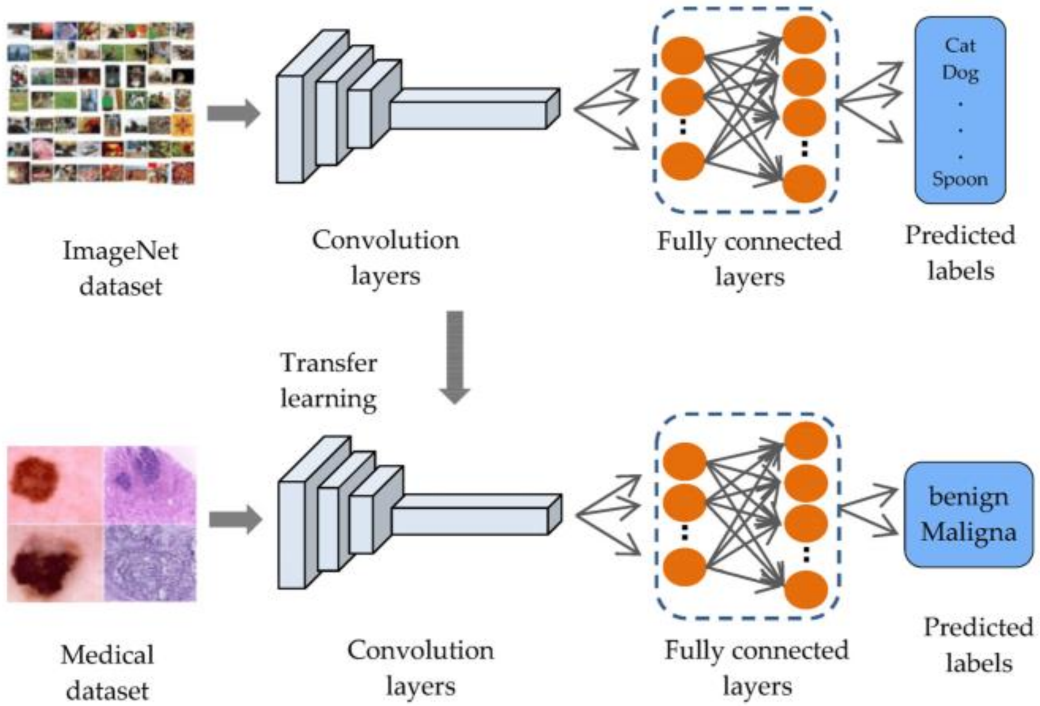
Convolution Neural Network



หยิบ Deep Learning Architecture ที่เค้าทำไว้แล้ว มาใช้ในงาน หรือมีการเทรนใหม่ ปรับให้เข้ากับงานเรา



Use a **pre-trained** network (opposite: training from scratch)

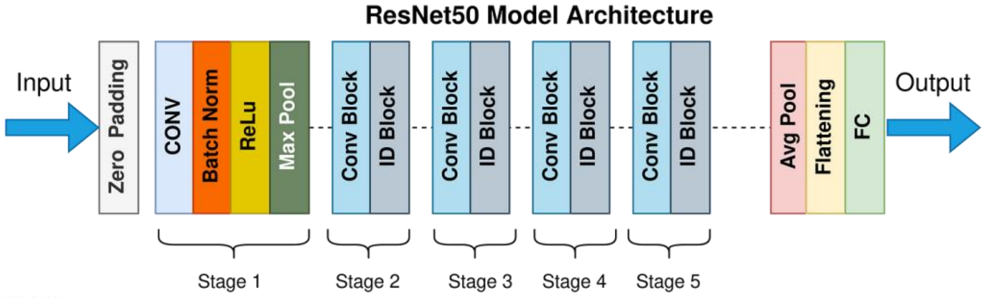


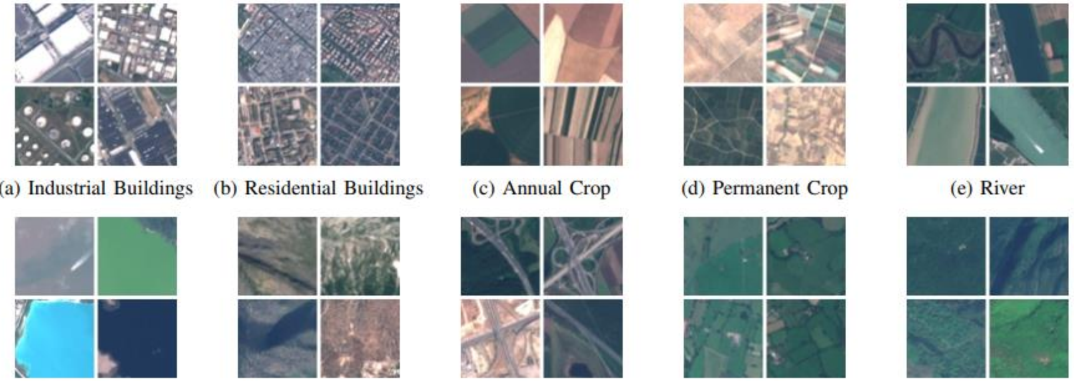
Application in Remote Sensing

EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification

Article in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing · August 2017
DOI: 10.1109/JSTARS.2019.2918242

Landcover Classification





(a) Industrial Buildings (b) Residential Buildings (c) Annual Crop (d) Permanent Crop (e) River

(f) Sea & Lake (g) Herbaceous Vegetation (h) Highway (i) Pasture (j) Forest

Fig. 4: This overview shows sample image patches of all 10 classes covered in the proposed EuroSAT dataset. The images measure 64x64 pixels. Each class contains 2,000 to 3,000 image. In total, the dataset has 27,000 geo-referenced images.

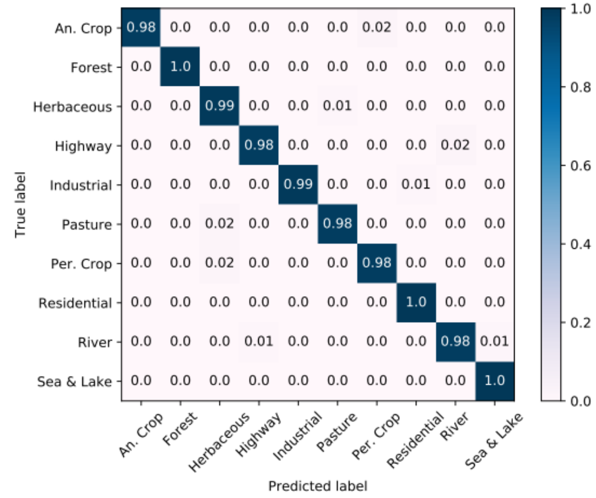


Fig. 8: Confusion matrix of a fine-tuned ResNet-50 CNN on the proposed EuroSAT dataset using satellite images in the RGB color space.

Method	UCM	AID	SAT-6	BCS	EuroSAT
ResNet-50	96.42	94.38	99.56	93.57	98.57
GoogLeNet	97.32	93.99	98.29	92.70	98.18

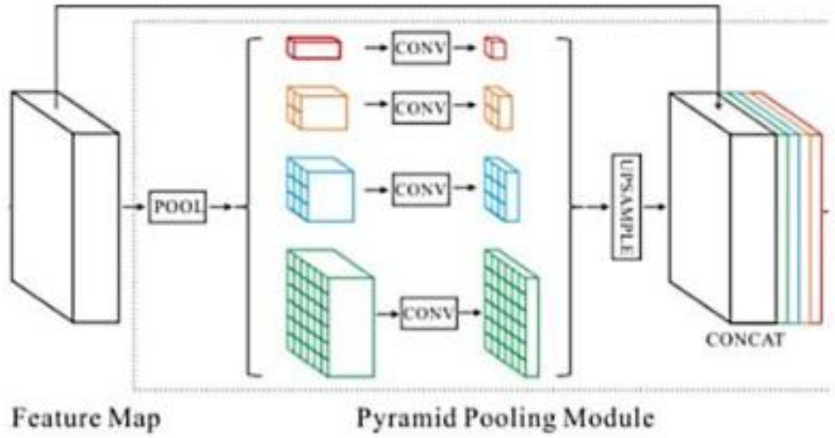
Application in Remote Sensing

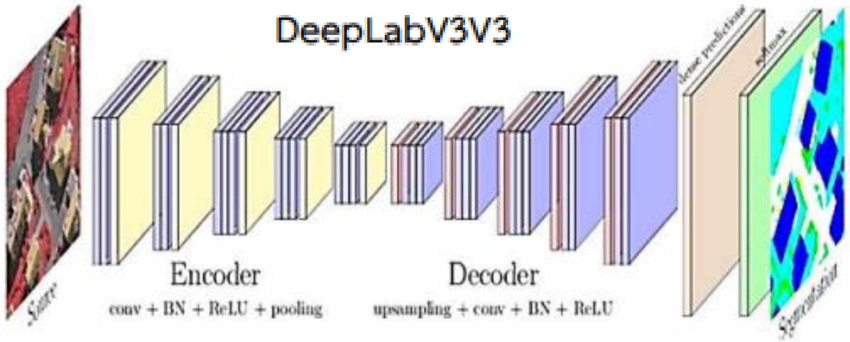
Land Cover Segmentation Using Convolutional Neural Network and Transfer Learning

Wachirawit Tangsirivichaikul
Kritchayan Intarat
Faculty of Liberal Arts, Thammasat University, Thailand
Corresponding author's email: wachirawit.srd@gmail.com

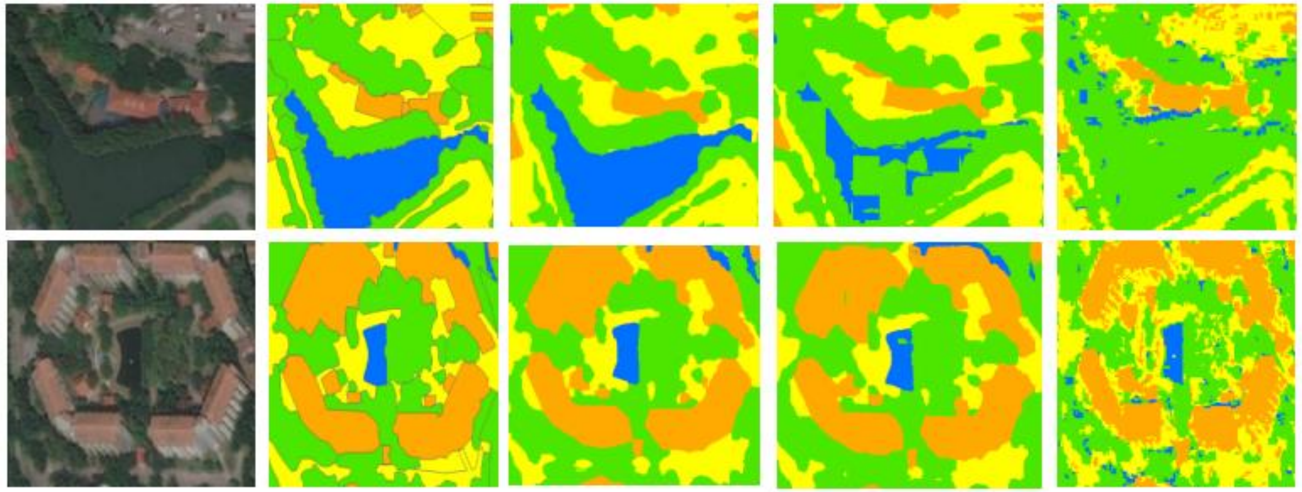
Image Segmentation

PSPNet

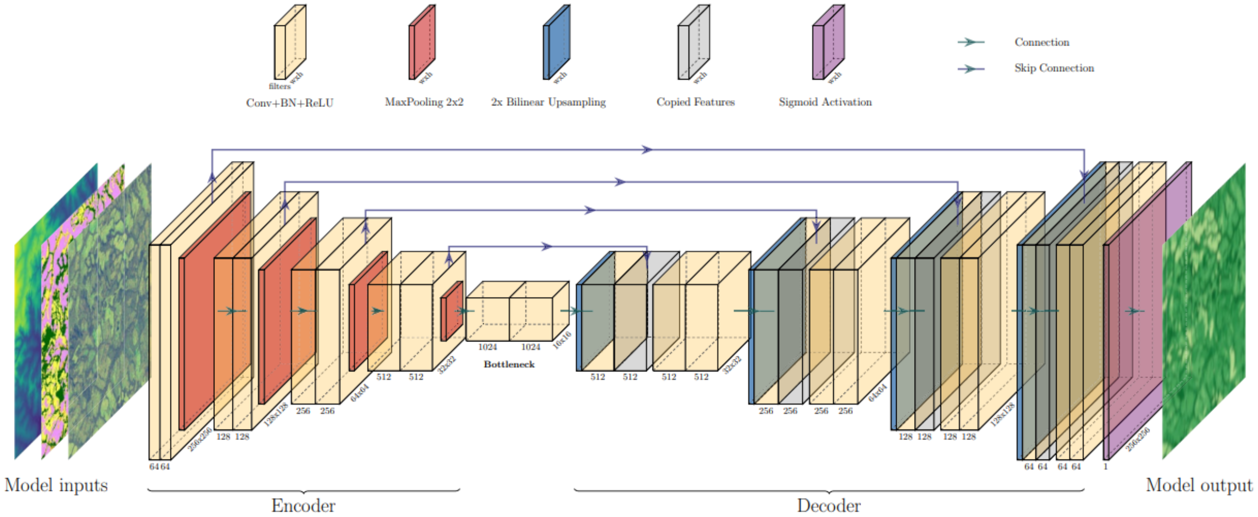




Input Image | Label Image | Segmentation: DeepLabV3V3 | PSPNet | Object-Based



Model	Precision	Recall	F1-Socre
DeepLabV3	0.906	0.940	0.921
PSPNet	0.904	0.864	0.867
object-based	0.620	0.668	0.607



A Globally Applicable Method for NDVI Estimation from Sentinel-1 SAR Backscatter Using a Deep Neural Network and the SEN12TP Dataset

Thomas Roßberg[1] · Michael Schmitt[1]

Fig. 6 U-Net architecture based on Ronneberger et al. (2015). Features are extracted from the model inputs in the Encoder and transformed by the Decoder to the model output. To extract abstract features, the data size is reduced in the encoder. Spatial details are added in the decoder by the skip connections to end up with a spatially accurate prediction. Unlike the original paper, the feature maps are enlarged by a bilinear upsampling. Conv, BN, and ReLU denote convolutional layer, batch normalisation, and ReLU activation function, respectively. The number of filters a convolutional layer has is shown below each layer, next to it are the width and height wxh

Table 7 Mean absolute error (MAE) per ESA WorldCover class on the test set

Worldcover class	Mean MAE
Trees	0.104186
Shrubland	0.106922
Grassland	0.104518
Cropland	0.103508
Built-up	0.106082
Barren/sparse vegetation	0.115451
Snow and ice	0.108554
Open water	0.099051
Herbaceous wetland	0.104841
Mangroves	0.105674
Moss and lichen	0.105813

The performance is averaged over five models, each using $\sigma°$ backscatter data, the WorldCover, and the DEM as input. For each model, first, the MAE per land-cover class is calculated and then the mean calculated. The class Unknown is discarded, because only 0.01% of the test pixels belong to that class

Building Height Estimation

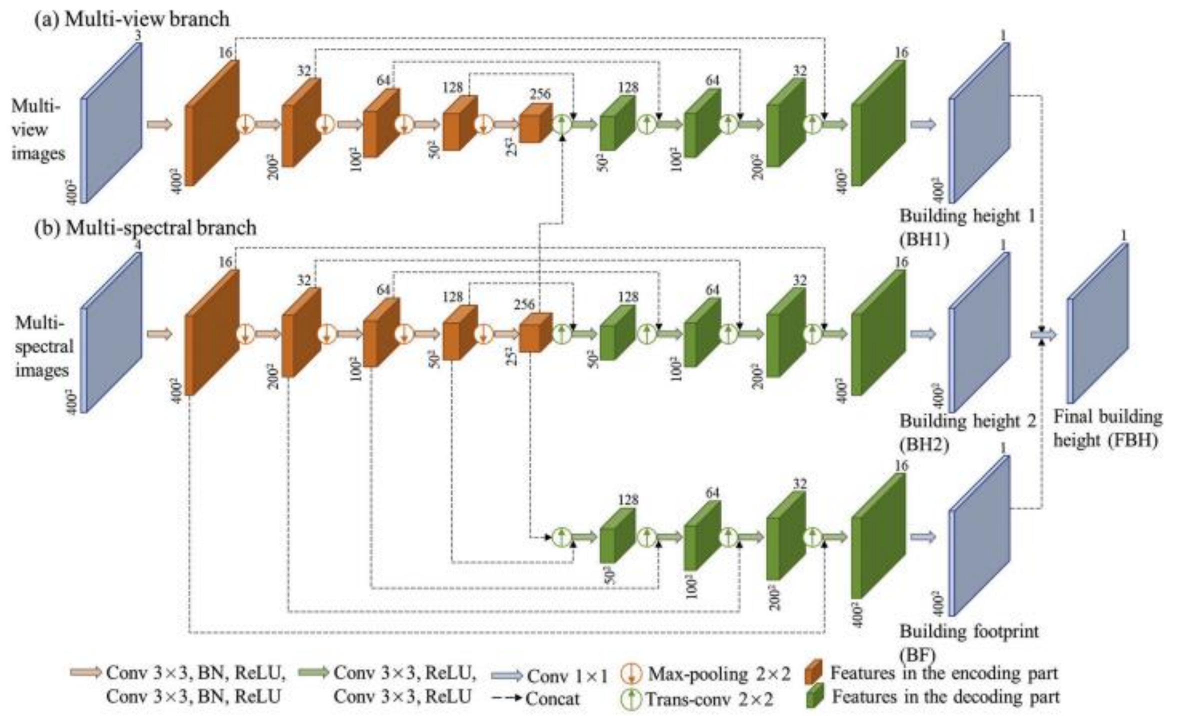
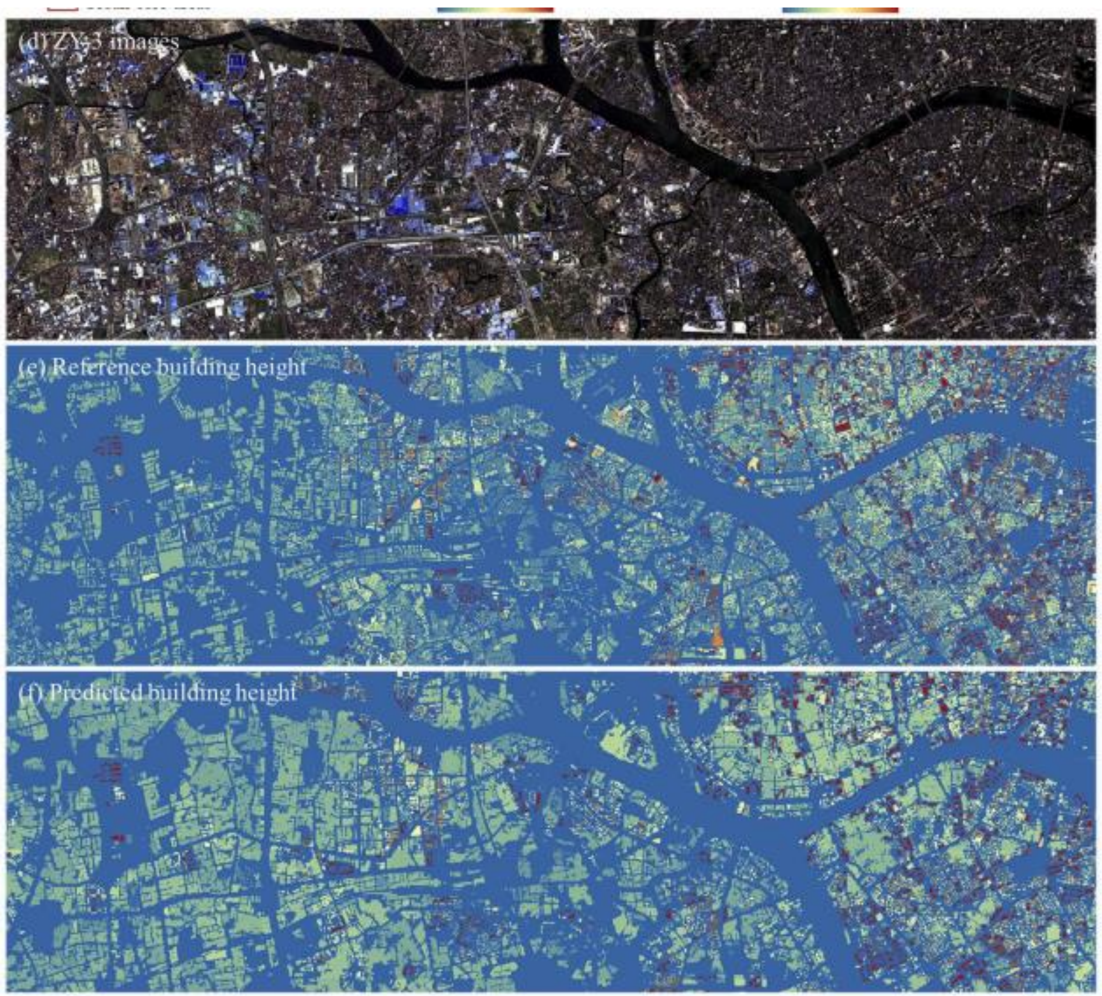


Fig. 4. Structure of the proposed M^3Net consisting of (a) the multi-view branch and (b) the multi-spectral branch. The dimension of features is described by the number of channels and resolution. For instance, 400^2 (pixels) refers to the width and the height of features, and 1 denotes the number of feature channels. Conv: convolution layer; BN: batch normalization; Trans-conv: transposed convolutional layer; ReLU: rectified linear unit activation function. 3×3 or 2×2 corresponds to the kernel size.



GOOD LUCK :)

Geospatial Programming

Modern Integrated Surveying Technologies

31st January 2024

Thepchai Srinoi

Master Degree Student and Teaching Assistant,

Department of Survey Engineering Chulalongkorn University