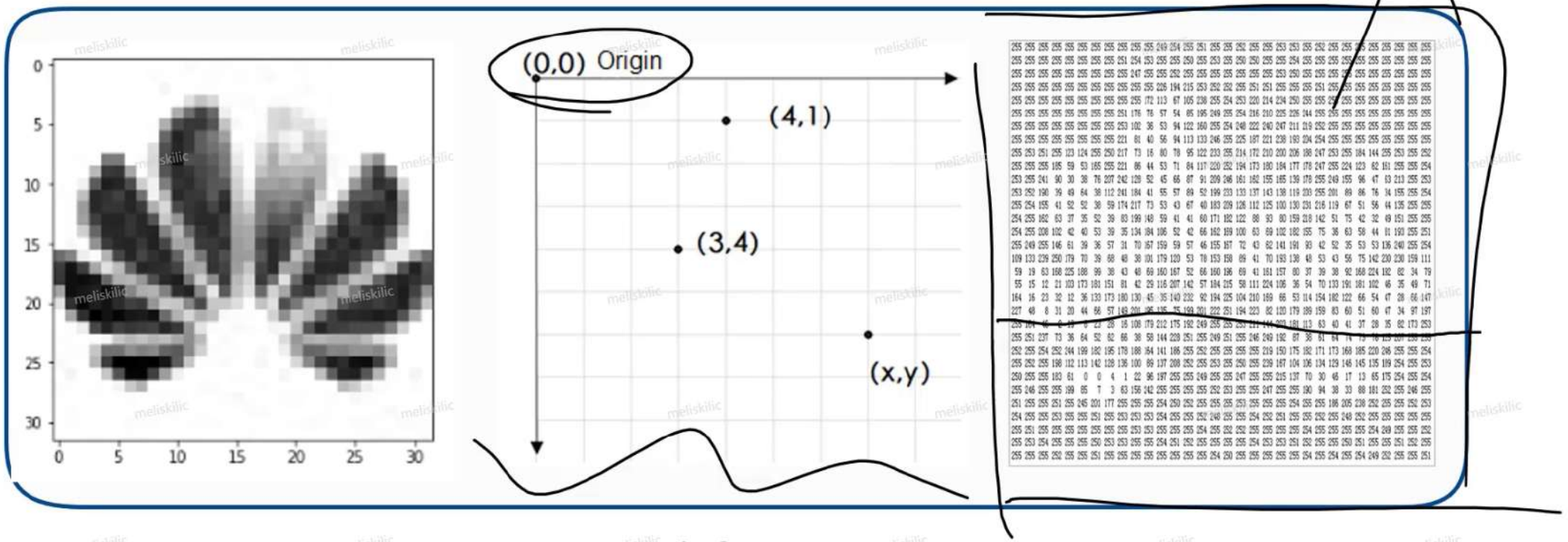
Effect of Different Sampling Resolutions and Quantization Depth

Sampling resolutions

Quantization Depth

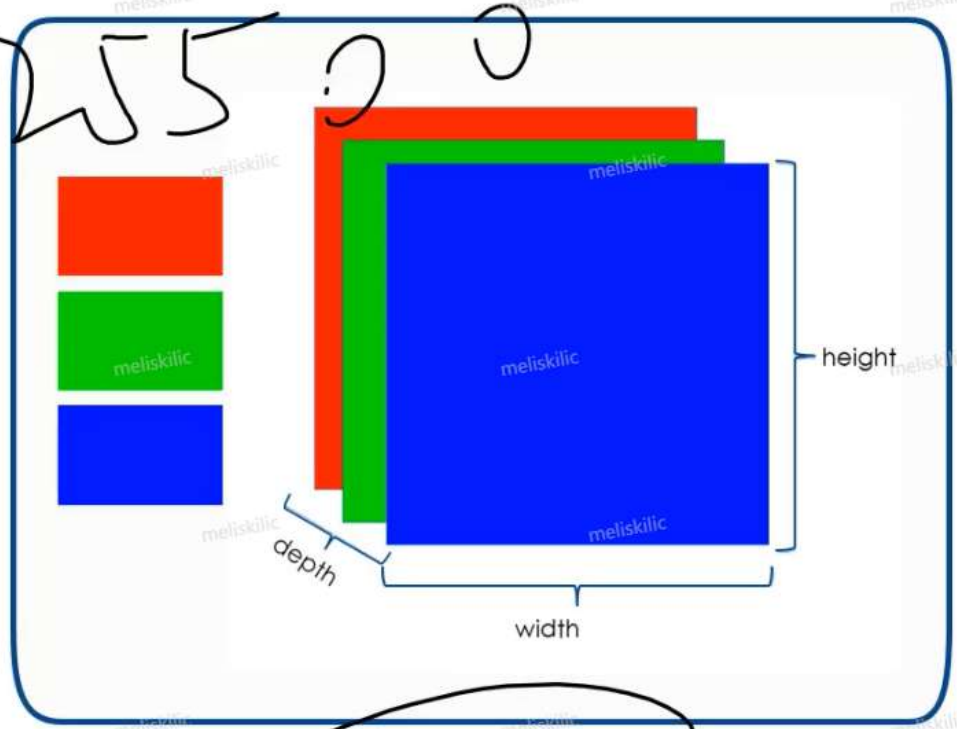
Grayscale Image



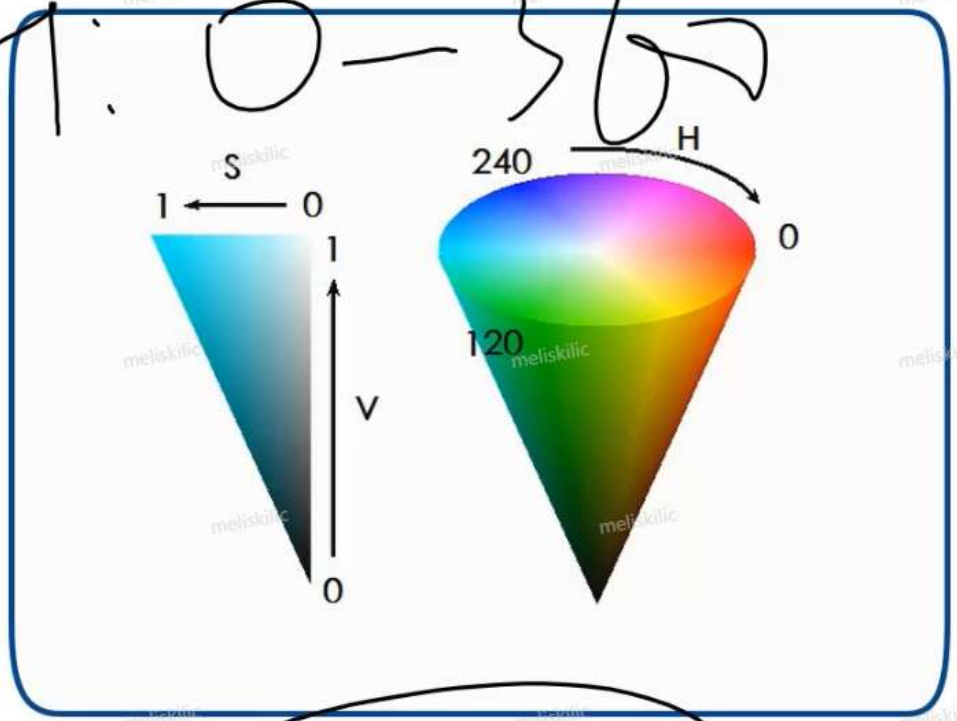
Detail of an Image

Color Space

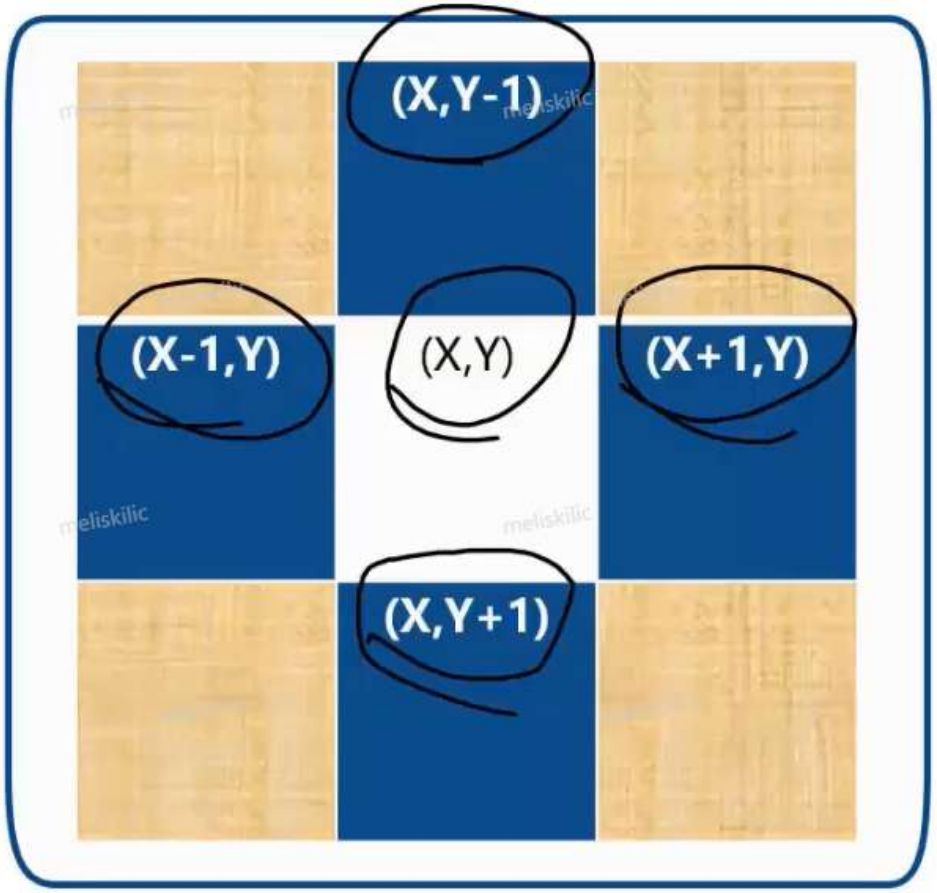
There are some other color space, such as: YUV, CMYK, Lab .



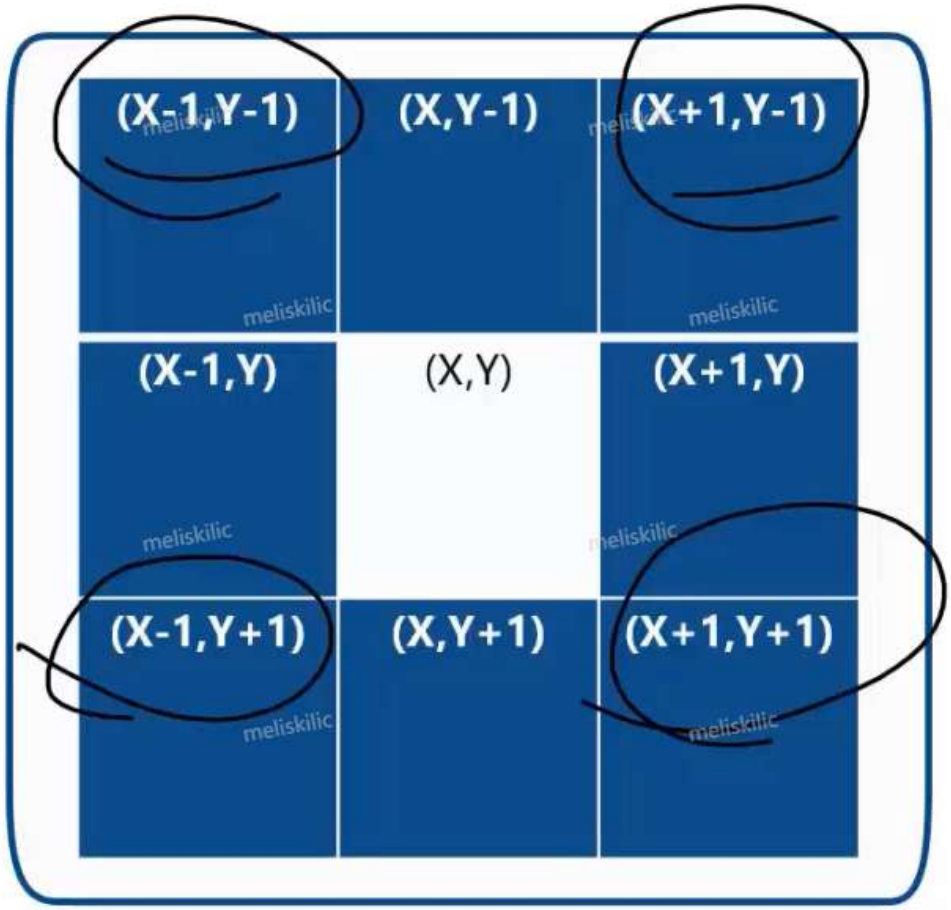
RGB Color space



HSV Color space



4-neighborhood



8-neighborhood

$$A = \begin{bmatrix} -1 & 0 & w-1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & h-1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Original image

Horizontal mirroring

Vertical mirroring

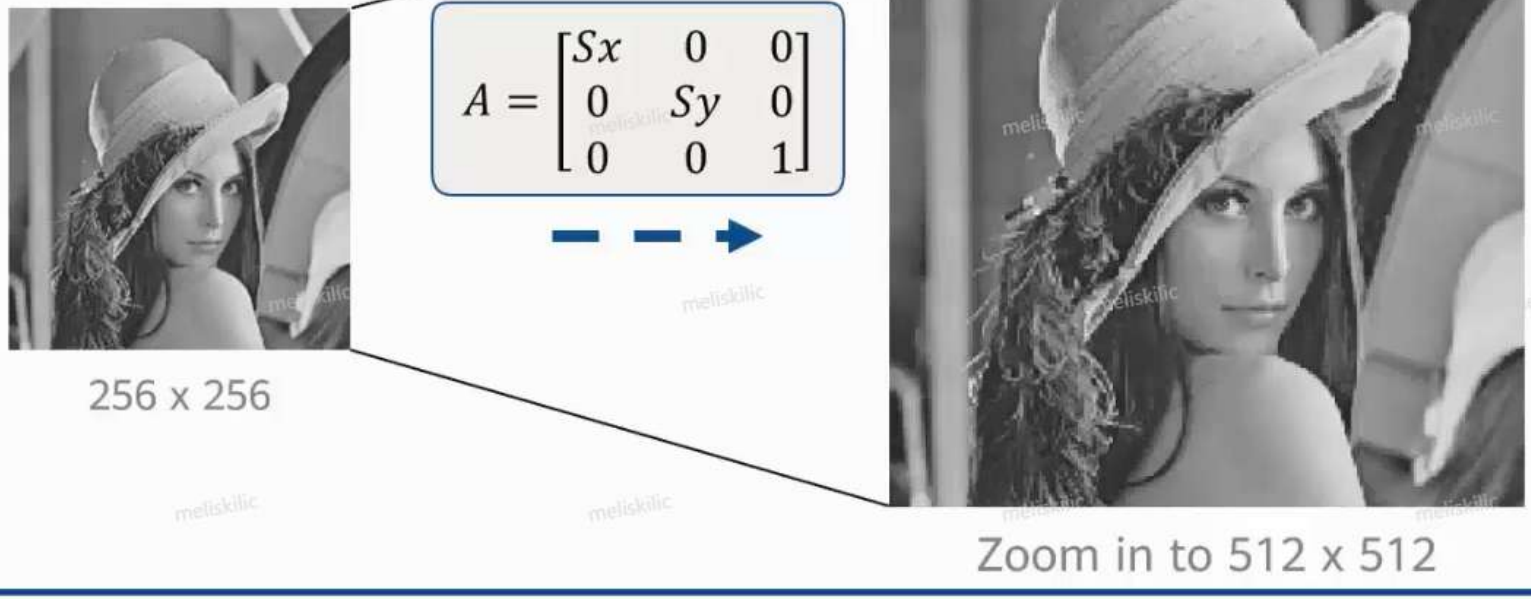
Rotation

The rotation matrix is $A = \begin{bmatrix} cos\theta & sin\theta & T_x \\ -sin\theta & cos\theta & T_y \\ 0 & 0 & 1 \end{bmatrix}$.

$$A = \begin{bmatrix} cos\theta & sin\theta & T_x \\ -sin\theta & cos\theta & T_y \\ 0 & 0 & 1 \end{bmatrix}$$

Zooming

The transformation matrix of zooming is $A = \begin{bmatrix} Sx & 0 & 0 \\ 0 & Sy & 0 \\ 0 & 0 & 1 \end{bmatrix}$



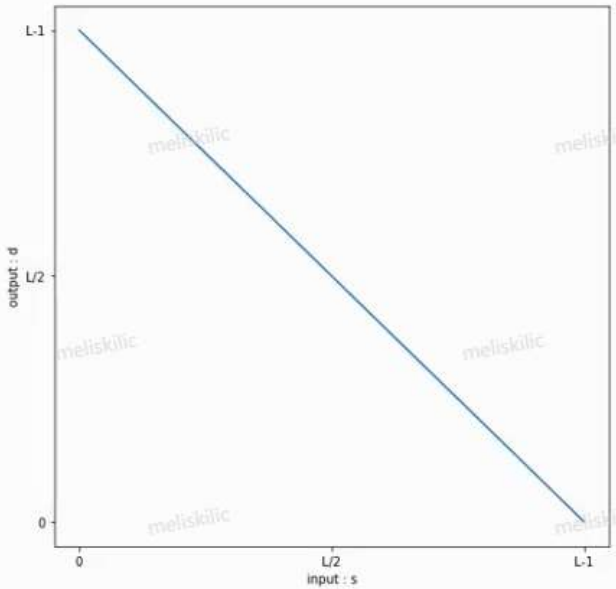
Interpolation

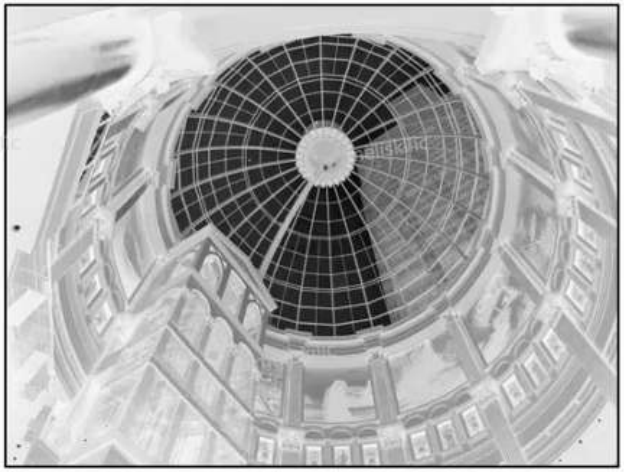
The values of (x,y) in a digital image should be integers, but the coordinates (x',y') obtained after coordinate transformation may not be integers. The pixel values at non-integer coordinates need to be calculated using the pixel values of the neighboring integer coordinates. This process is called interpolation.

Zooming using nearest neighbor interpolation

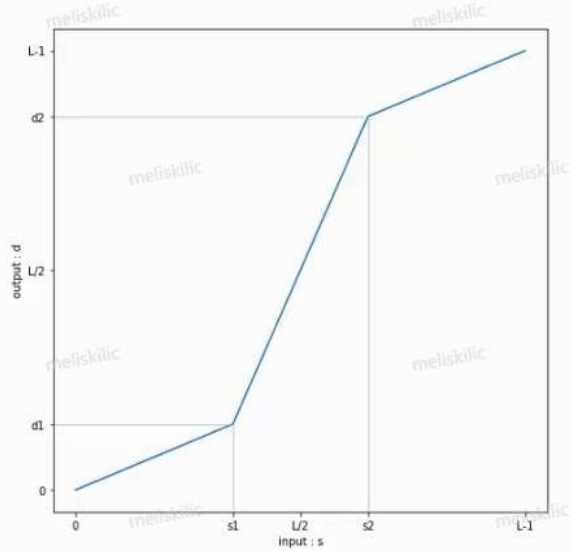
Zooming using bilinear interpolation

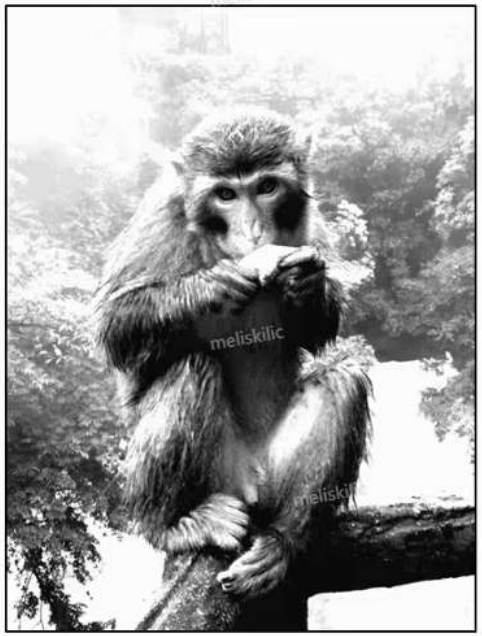
Grayscale Transformation-Inversion



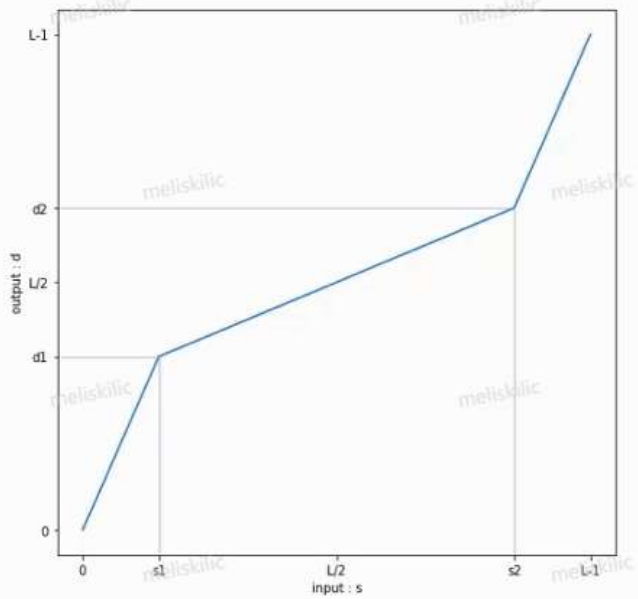


Grayscale Transformation-Contrast Enhancement

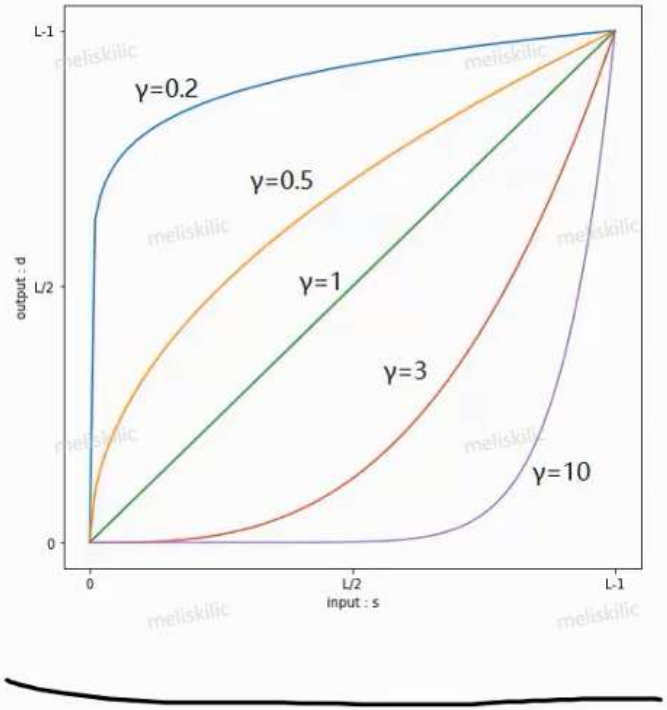




Contrast Compression



Gamma Correction



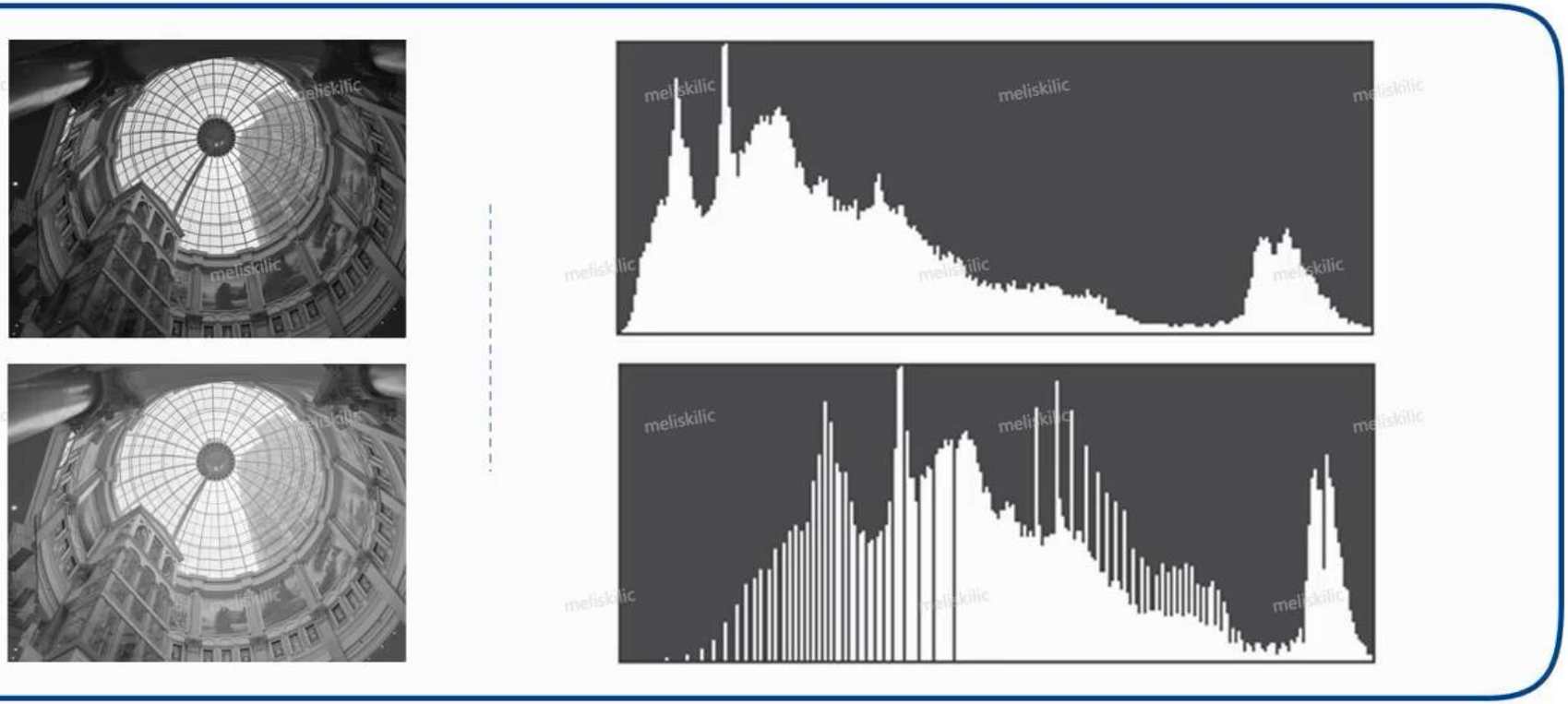
γ = 0.5

Original image

γ = 3

Histogram of a Grayscale Image

The following are histograms of two grayscale images. The shapes of the histograms reflect the visual effect of the images.

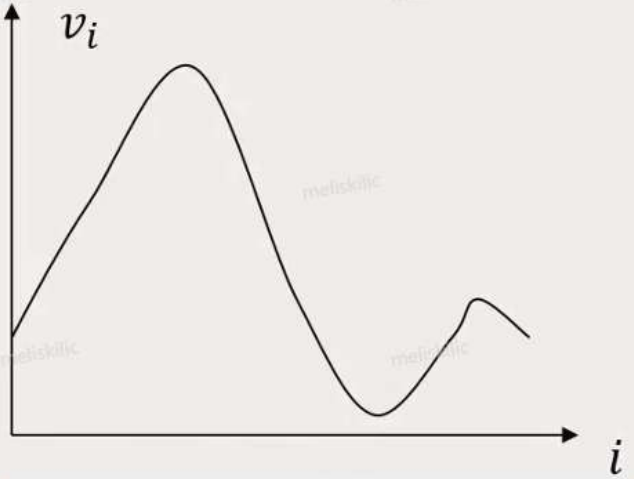


Histogram Calculation

For the normalized histogram, $v_i = \frac{n_i}{n}$,and v_i indicates the frequency of the ith gray level.

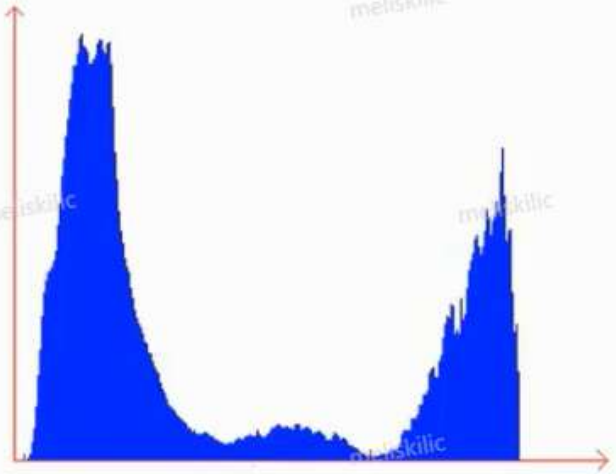
1	2	3	1	2	3	1	0
7	6	5	2	6	7	5	0
2	1	5	3	6	0	6	1
0	5	6	3	5	7	6	2
6	1	4	2	7	2	2	3
0	6	7	2	6	5	2	2
2	1	2	1	2	3	2	1
1	2	2	1	3	2	1	3

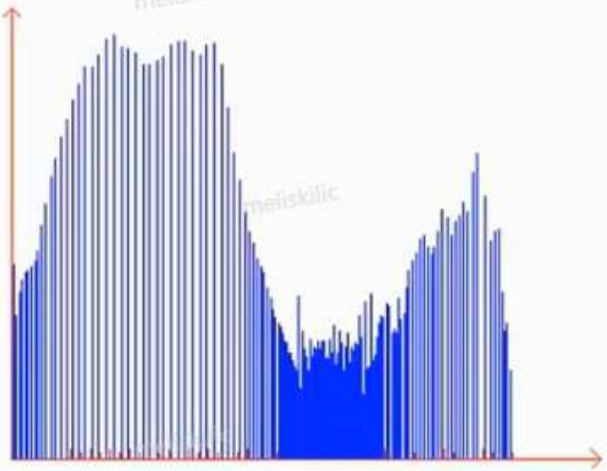
$$v_0 = 5/64$$
$$v_1 = 12/64$$
$$v_2 = 18/64$$
$$v_3 = 8/64$$
$$v_4 = 1/64$$
$$v_5 = 5/64$$
$$v_6 = 8/64$$
$$v_7 = 5/64$$



Histogram Equalization

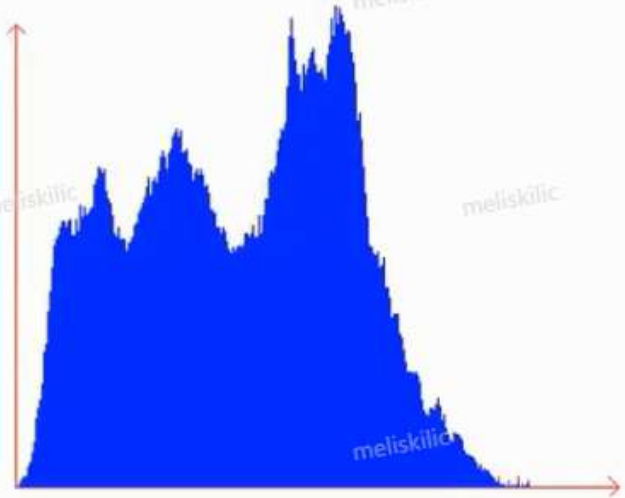
Histogram Equalization

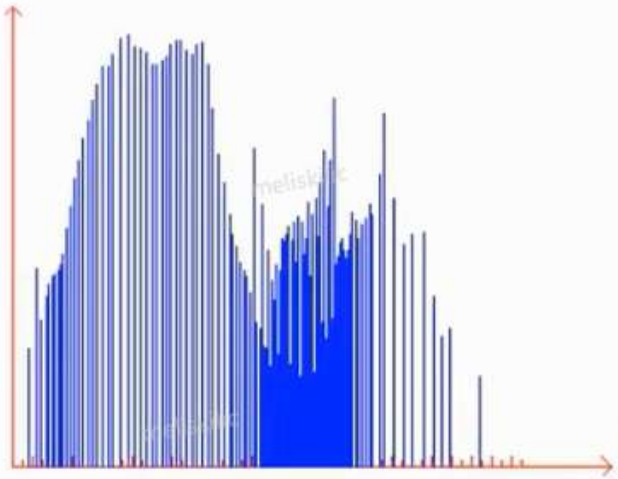




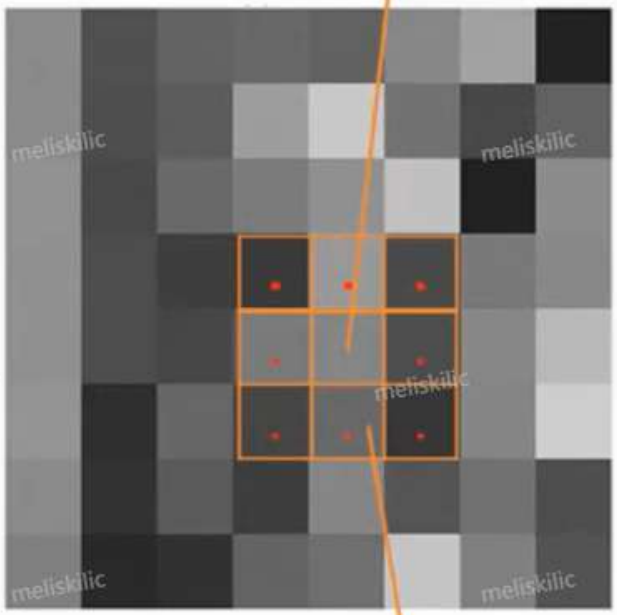
Histogram Specification

Histogram Specification





Template Operation-Spatial Filtering



Central pixel

Filter subimage

1	1	1
1	1	1
1	1	1

Coefficients rather than pixels

N neighbors

Mean Filtering and Median Filtering

1	1	1
1	1	1
1	1	1

Original image

3 x 3 mean filtering

5 x 5 mean filtering

Original image

3 x 3 mean filtering

3 x 3 median filtering

Gaussian Filtering

1	4	7	4	1
4	16	26	16	4
7	26	41	26	7
4	16	26	16	4
1	4	7	4	1

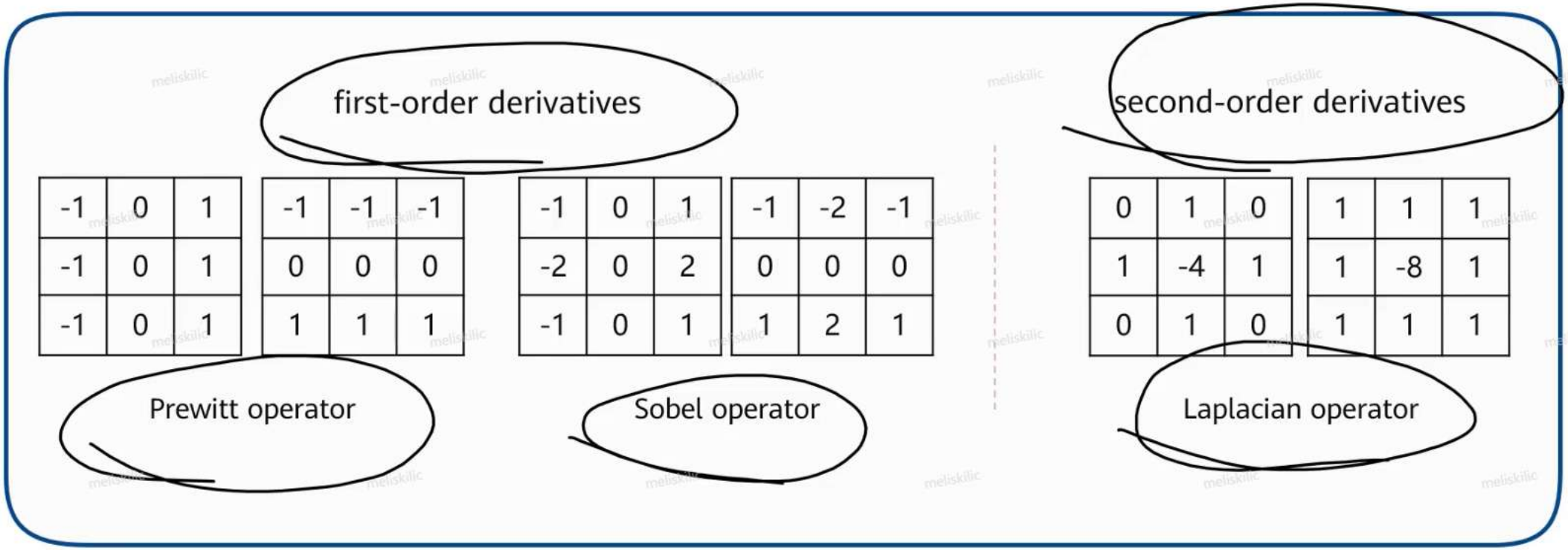
Original image

5 x 5 mean filtering

5 x 5 Gaussian filtering

Sharpening

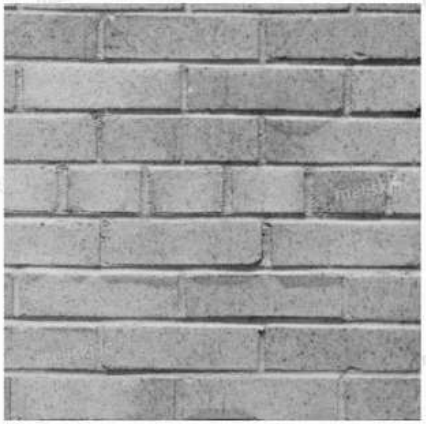
The main purpose of sharpening is to highlight details in images and enhance blurred details in images.



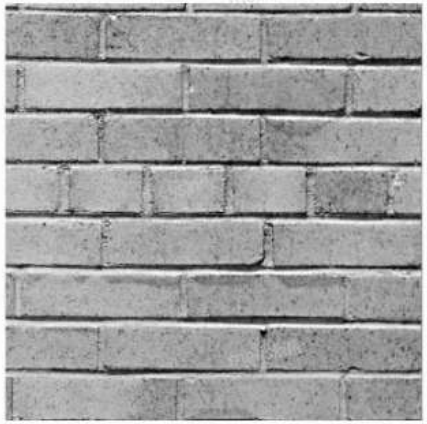
Sharpening Implementation

The Laplacian gradient operator is used as the sharpening operation template, and A is a coefficient greater than or equal to 1.

0	**-1**	**0**
-1	A+ 4	-1
0	-1	0



Original image



Effect after sharpening using the Laplacian operator

Affine Transformation

Example of correcting QR code distortion: $\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Perspective Transformation

Example of correcting document distortion:

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Color Image Processing

Color image

R-channel

G-channel

B-channel

Brightness Enhancement

Color image

Small V

Big V

Saturation Enhancement

Color image

Small S

Big S

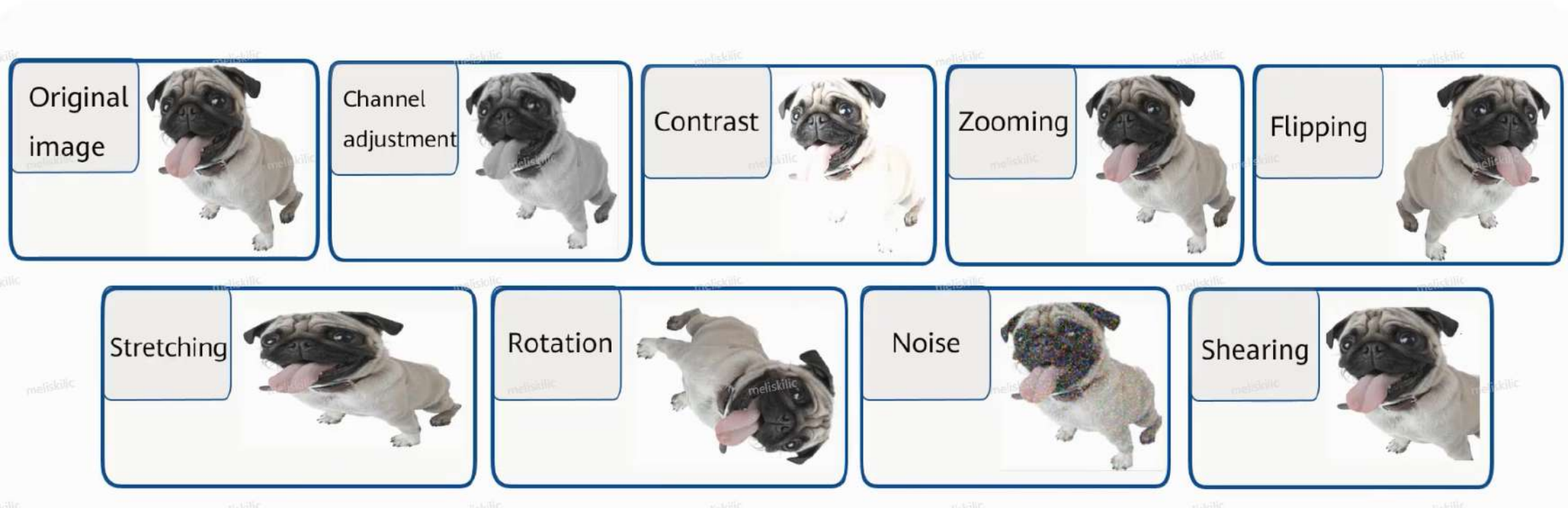
Hue Enhancement

Color image

Small H

Big H

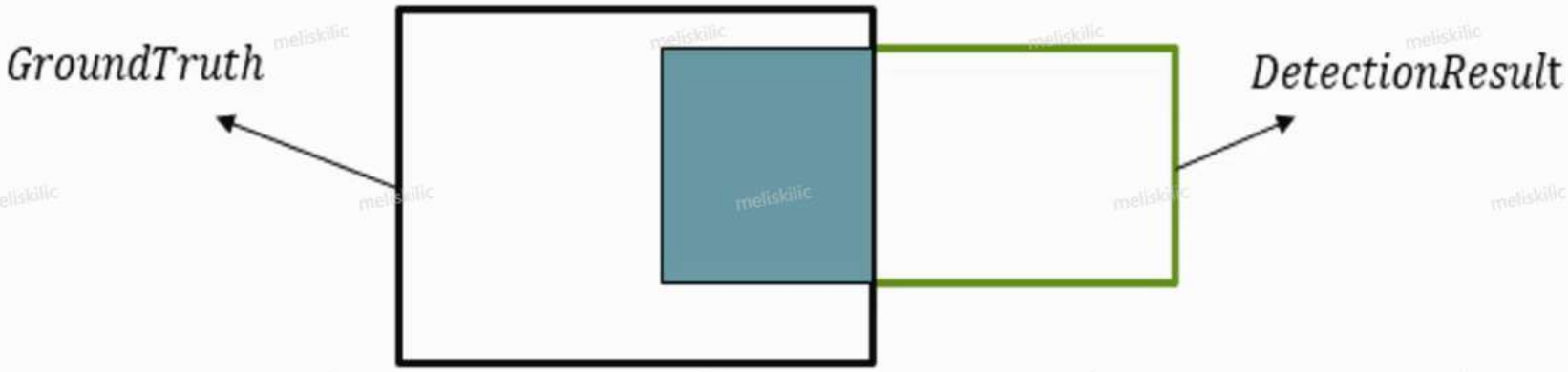
Image Data Augmentation for Deep Learning



Object Detection Performance Measurement

The value of IoU can reflect the accuracy of the detection result.

$$IoU = \frac{DetectionResult \cap GroundTruth}{DetectionResult \cup GroundTruth}$$



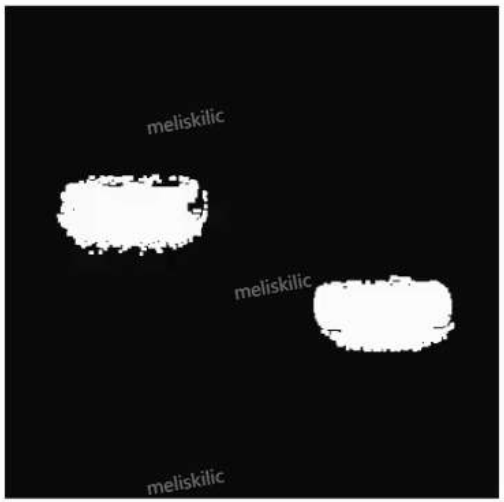
IOU is used to evaluate the output of object detection

Connected - Region Segmentation

Connected-region segmentation is the process of partitioning the region of interest from an image based on the connected domain. It is commonly used in character recognition technologies.

Motion Segmentation

Motion segmentation is the process of partitioning an image into the moving objects and background based on consecutive frames. Different application requirements can be met by separating the moving objects and background.



Object Segmentation

Object segmentation is the process of detecting object boundaries and performing pixel-level object segmentation from an image.

DAVIS-2016 (left) and DAVIS-2017 (right) multi-object segmentation

Image Segmentation Performance Measurement

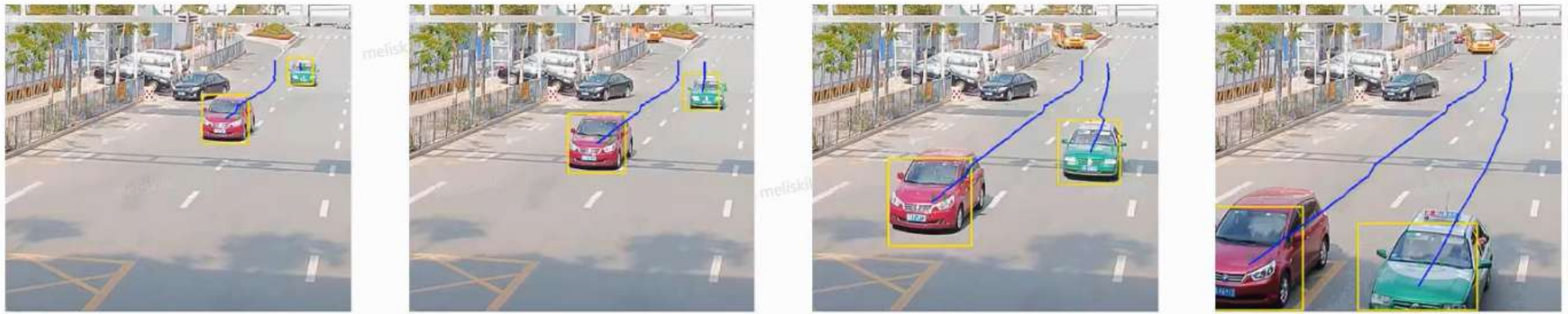
By comparing the area segmented by the algorithm and ground-truth area, you can evaluate the segmentation result. The Dice coefficient can be used to intuitively measure the segmentation performance.

$$Dice = \frac{2 * (V_{gt}\ and\ V_{seg})}{V_{gt} + V_{seg}}$$

Object Tracking

Difference between object detection:

- Object detection is to locate objects in a static single-frame image
- Object tracking is to locate objects in consecutive frames.

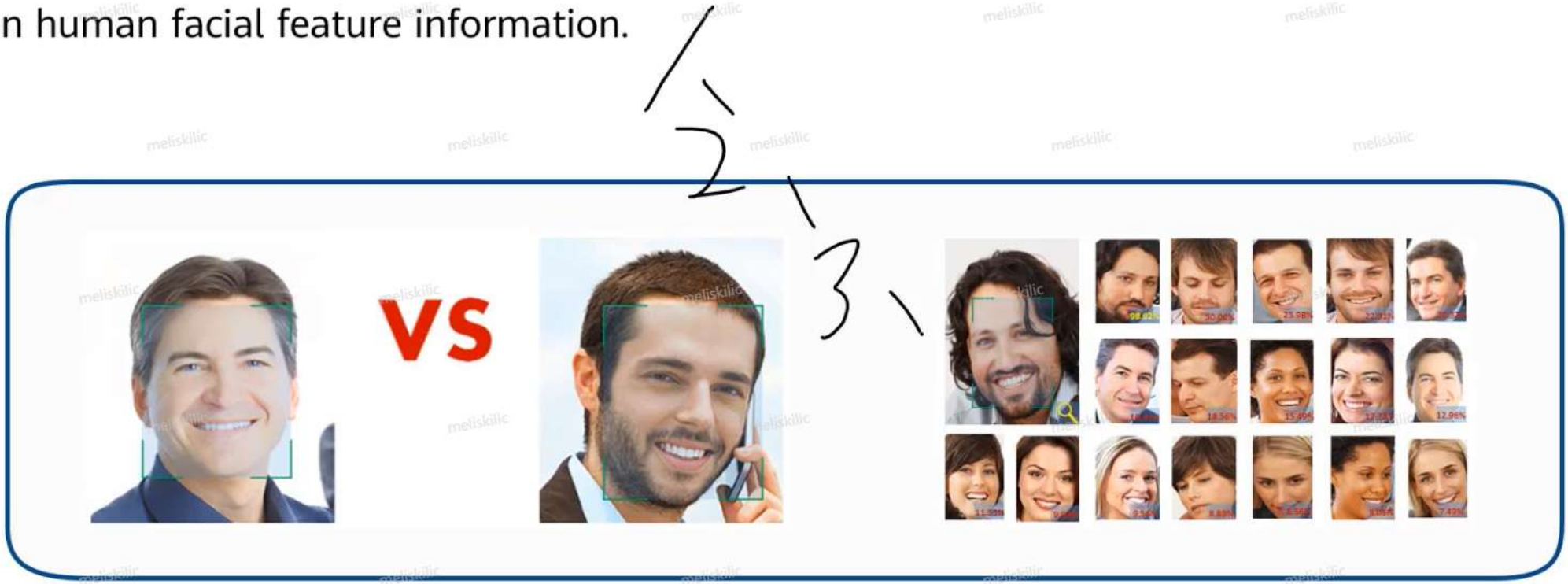


Character Recognition

Optical Character Recognition (OCR) is the process of recognizing characters in images or scanned copies and converting them into editable ones, using image processing technologies such as character object detection and object classification.

Facial Recognition

Facial recognition is the process of processing, analyzing, and understanding face images based on human facial feature information.



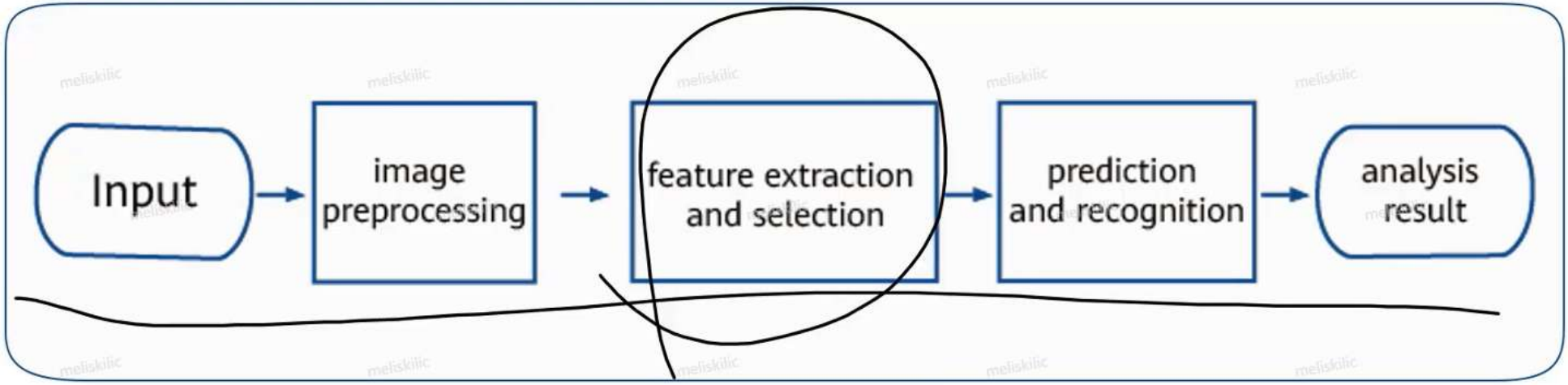
Similarity: 1.21%

Content Detection

Content Detection is a technology that uses computers to process, analyze, and understand images, so that targets and objects in different modes can be recognized.

Traditional Image Processing Algorithms

Data of digital images is unstructured and cannot be processed using pattern recognition algorithms. To convert digital images into structured data effectively, the feature extraction technology is used to reduce data dimensions.



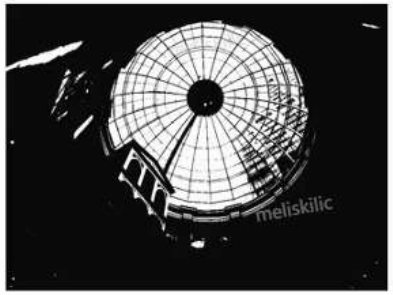
Thresholding

Thresholding is the process of replacing each pixel with the intensity less (greater) than the threshold in an image with a white pixel (or a black pixel). The threshold is represented by T.

The image segmentation effect varies depending on the threshold selected.


Original image


T = 60


T = 120

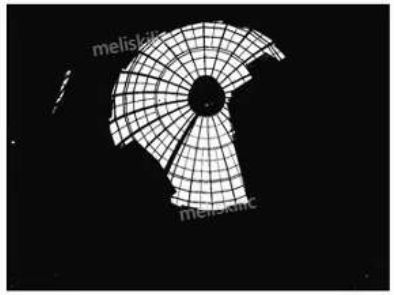
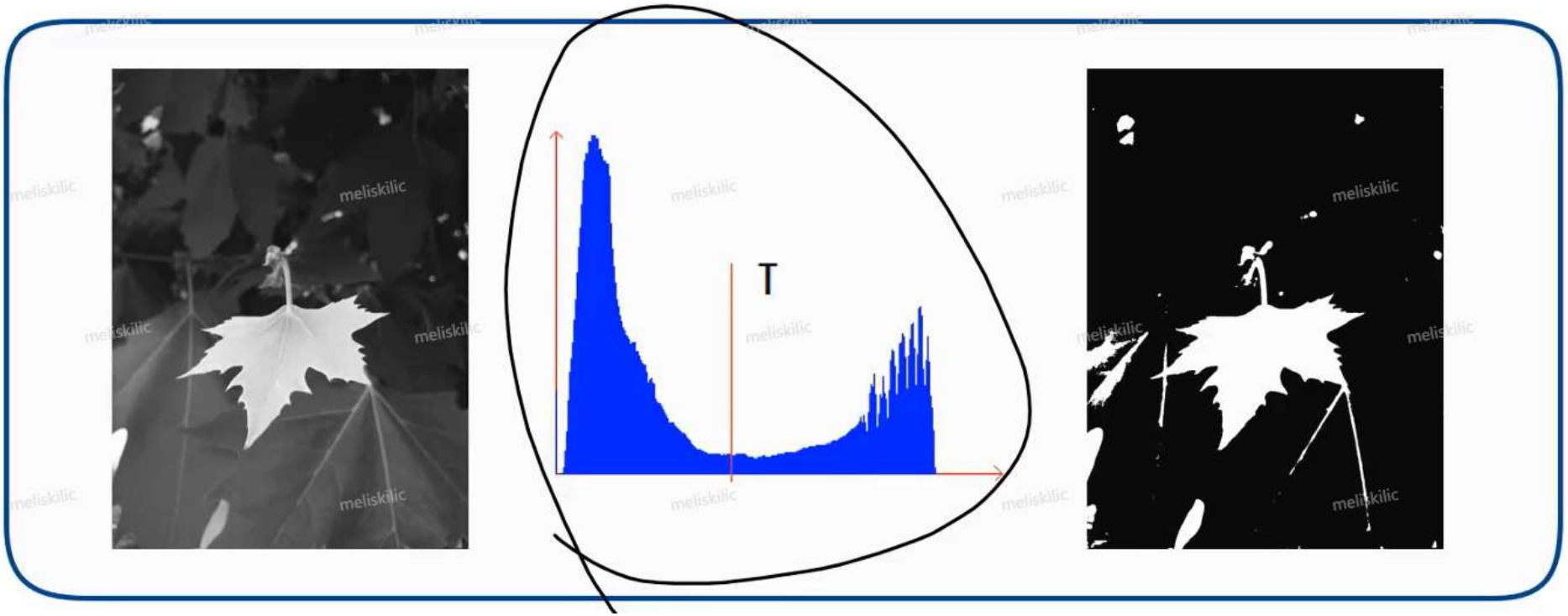

T = 180

Image Binarization

Image binarization is the process of converting an image to a black and white image.

Bimodal Histogram

A threshold value can be calculated based on the image histogram for a bimodal image. For a bimodal image (whose histogram has two peaks), the intensity of the valley between the two peaks is used as the threshold for image segmentation.



Morphological Image Processing

Morphological Operations

- erosion and dilation
- opening and closing

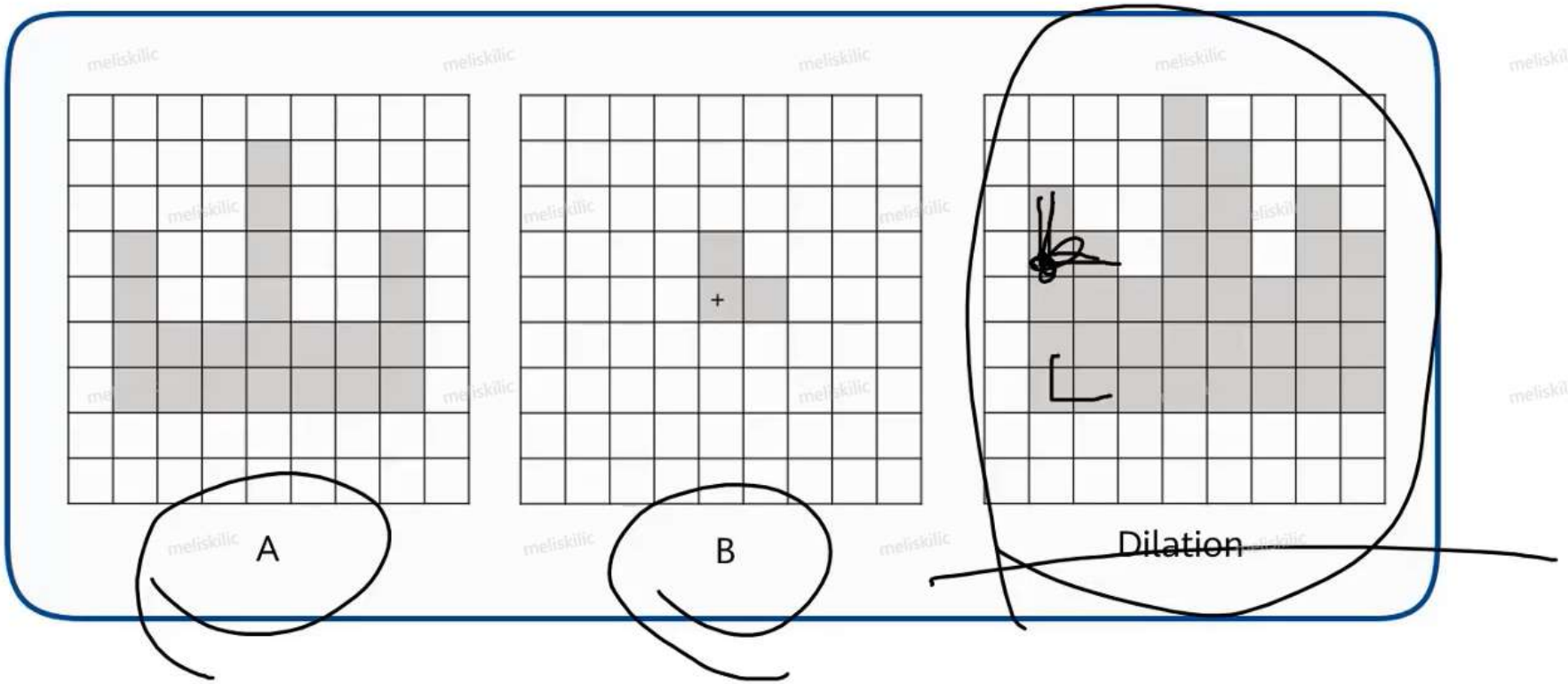
Advantages:
simplify image data, enhance shape characteristics, and eliminate interference noise.

Applications:
noise reduction, boundary extraction, object locating, and region filling.

Note: Morphological operations can only apply to binary images or grayscale images.

Morphological Image Processing

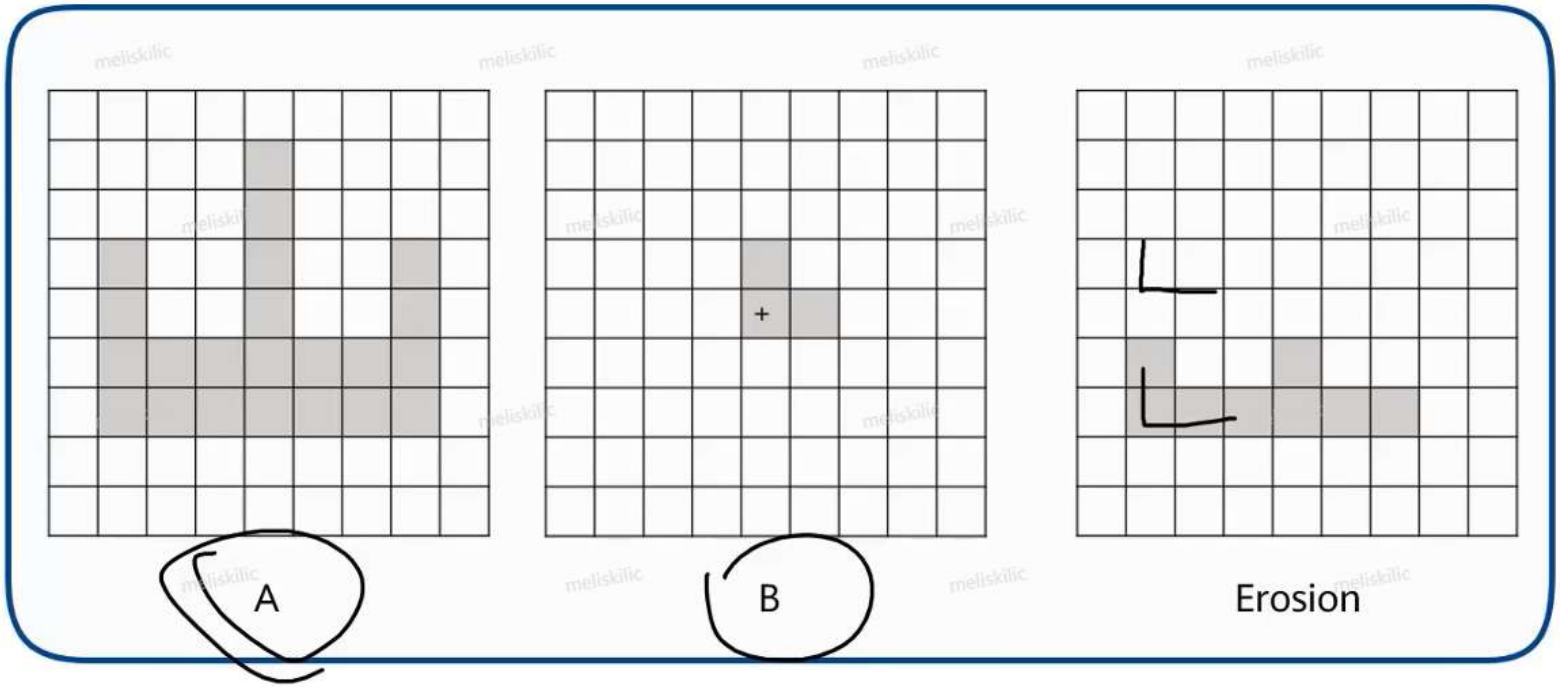
Dilation



The dilation of a set (A) by structuring element (B) is the process of expanding the foreground region in (A) based on the shape of (B).

Morphological Image Processing

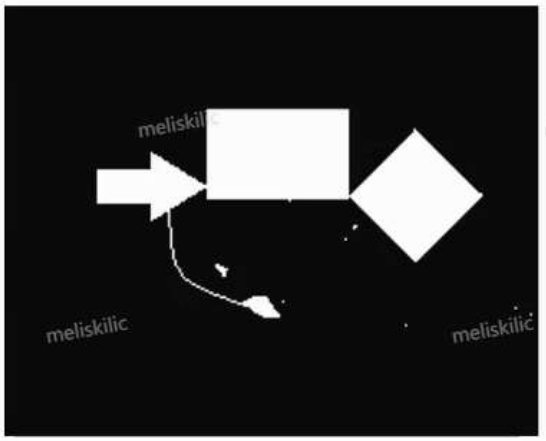
Erosion



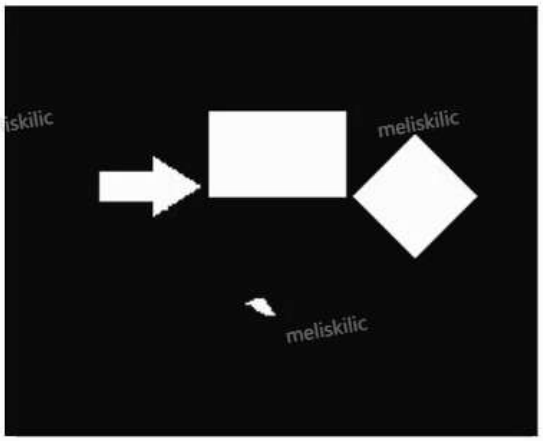
The erosion of a set (A) by structuring element (B) is the process of shrinking the foreground region based on the shape of (B).

Morphological Image Processing

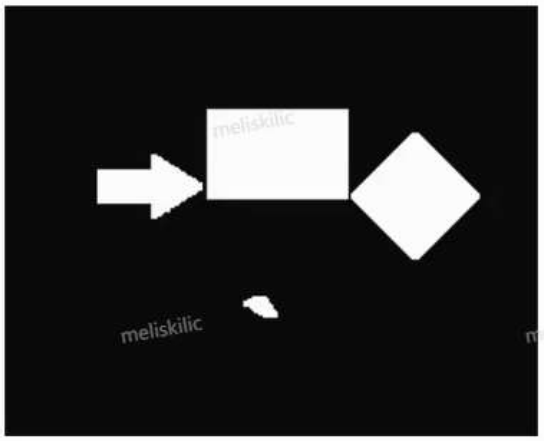
Opening



Original image



Erosion

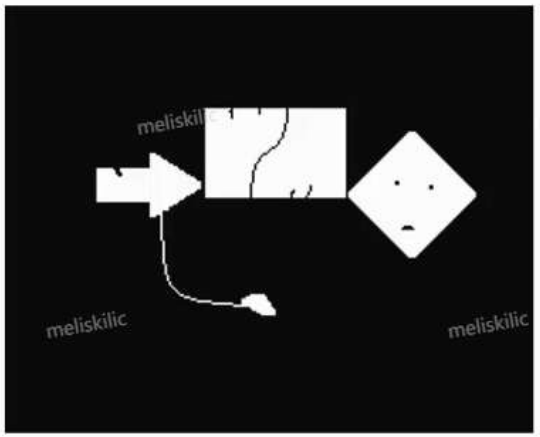


Opening

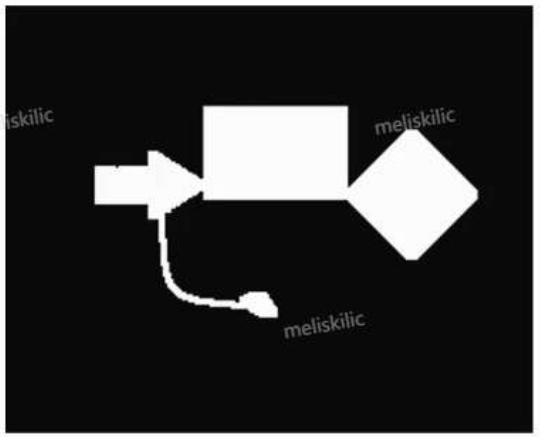
The opening is defined as an erosion followed by a dilation of a set by a structuring element.

Morphological Image Processing

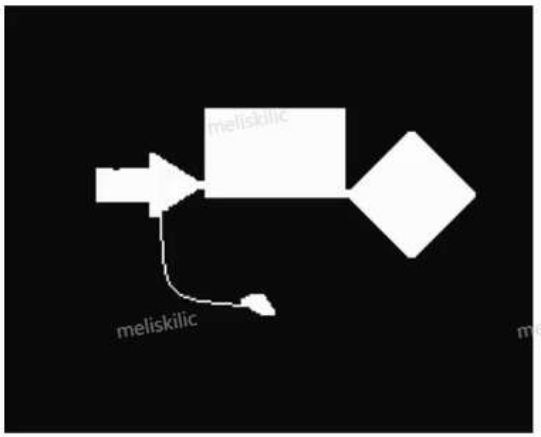
Closing



Original image



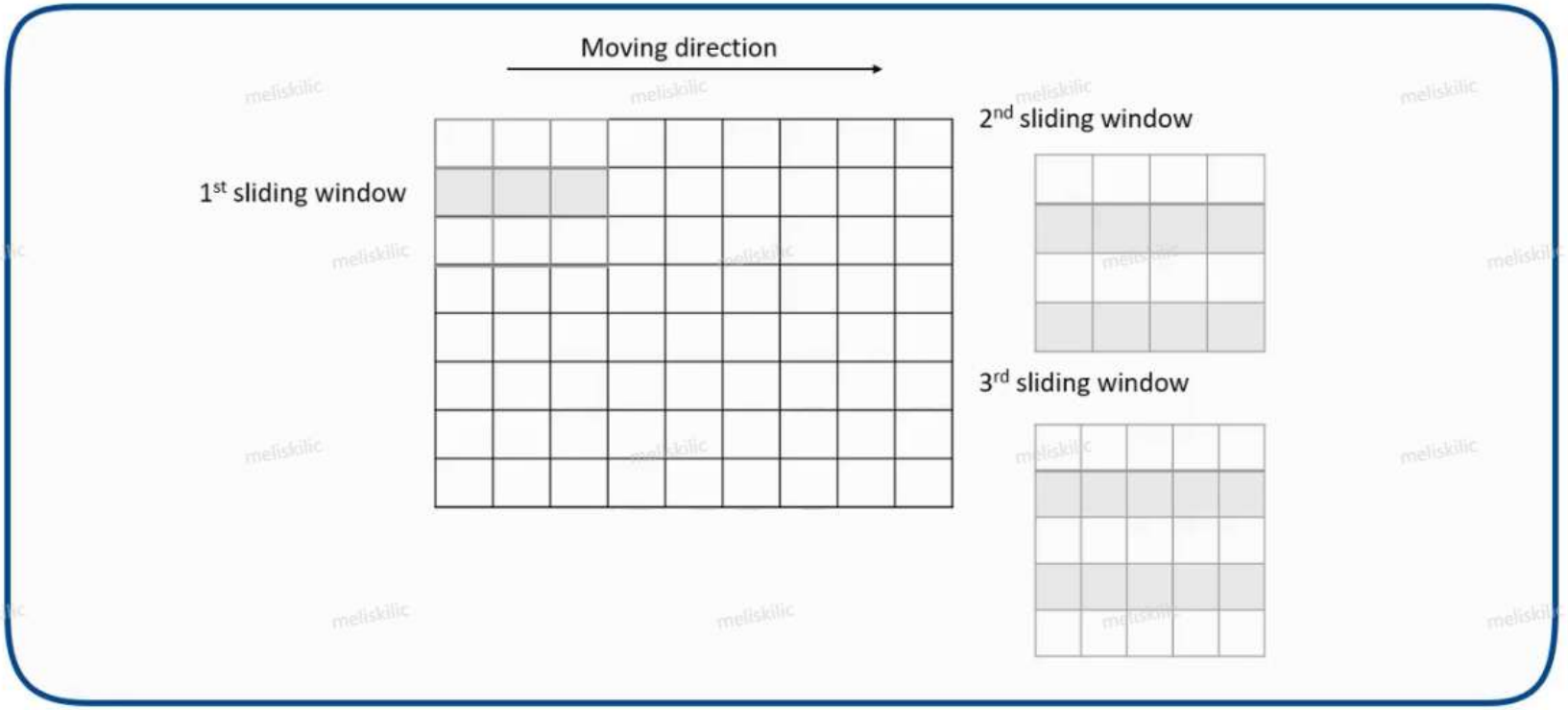
Dilation



Closing

The closing is defined as a dilation followed by an erosion of a set by a structuring element.

Sliding Window



The sliding window method is commonly used in image processing.

Multi-scale sliding window: multiple windows of different sizes are used in actual operations.

Template Matching

Template matching is a pattern recognition method in image processing and is a basic object detection algorithm.

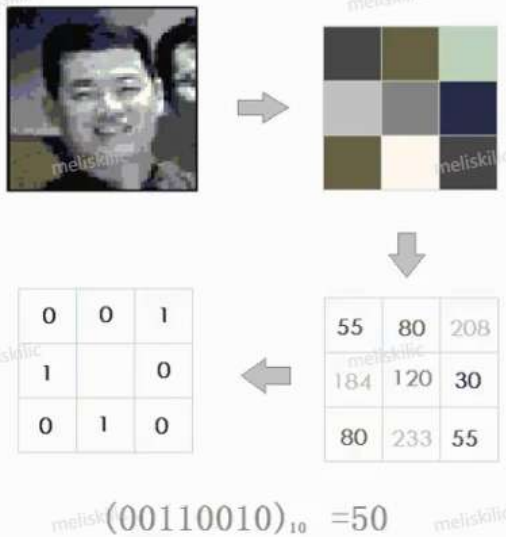
Feature Descriptor

Input

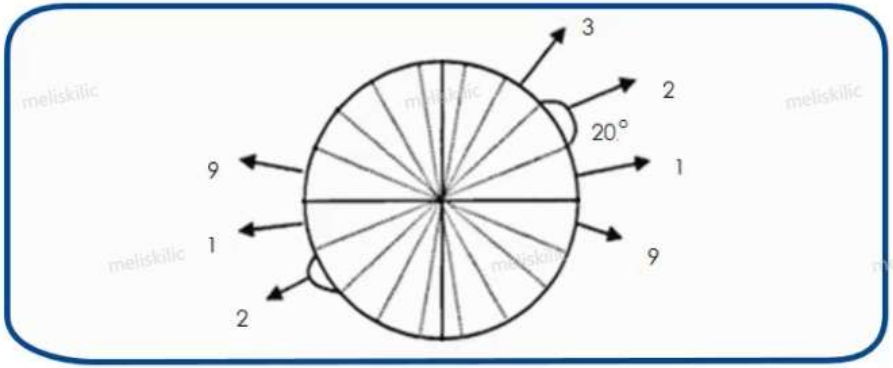
An image

Output

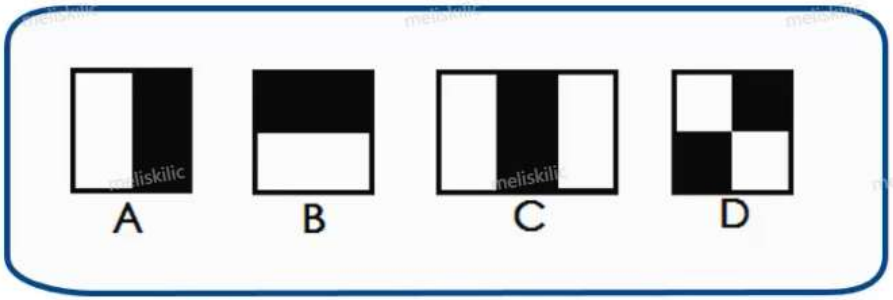
fixed-length feature vectors



HOG



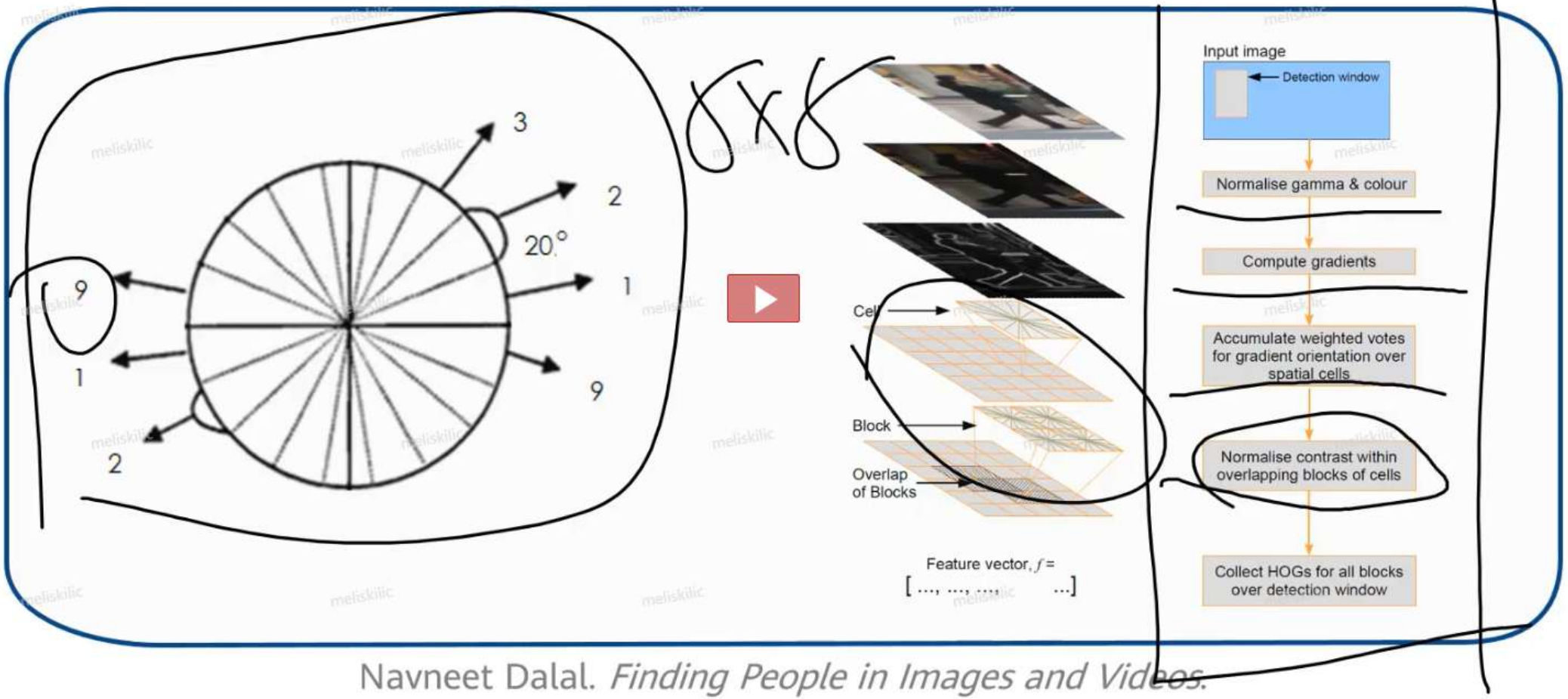
LBP



Haar

HOG

An overview of static HOG feature extraction is shown as follows:



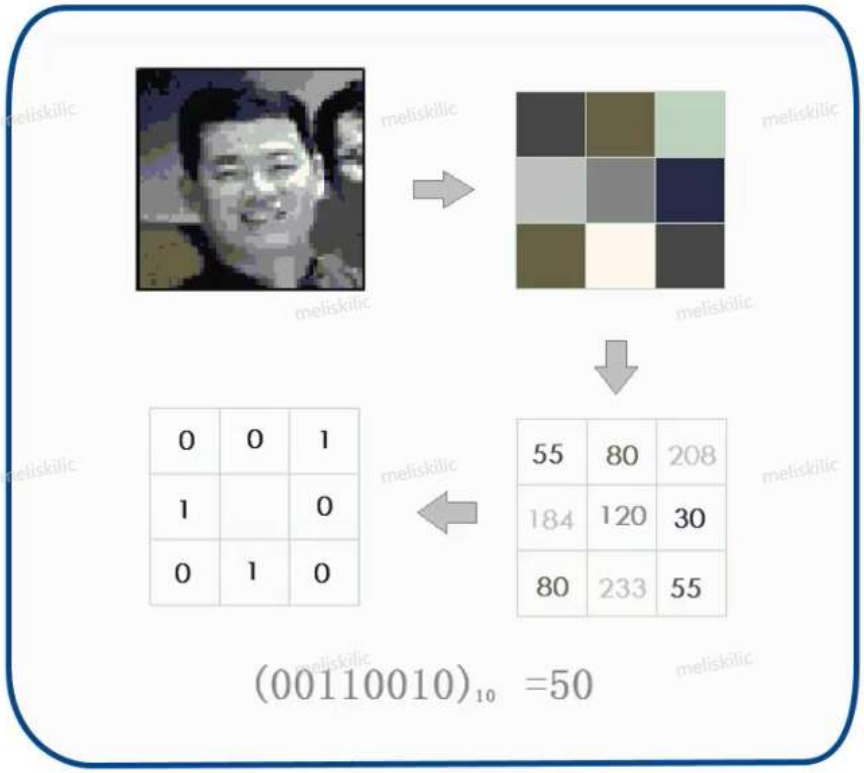
Navneet Dalal. *Finding People in Images and Videos.*

Pedestrian Detection

Navneet Dalal. *Finding People in Images and Videos.*

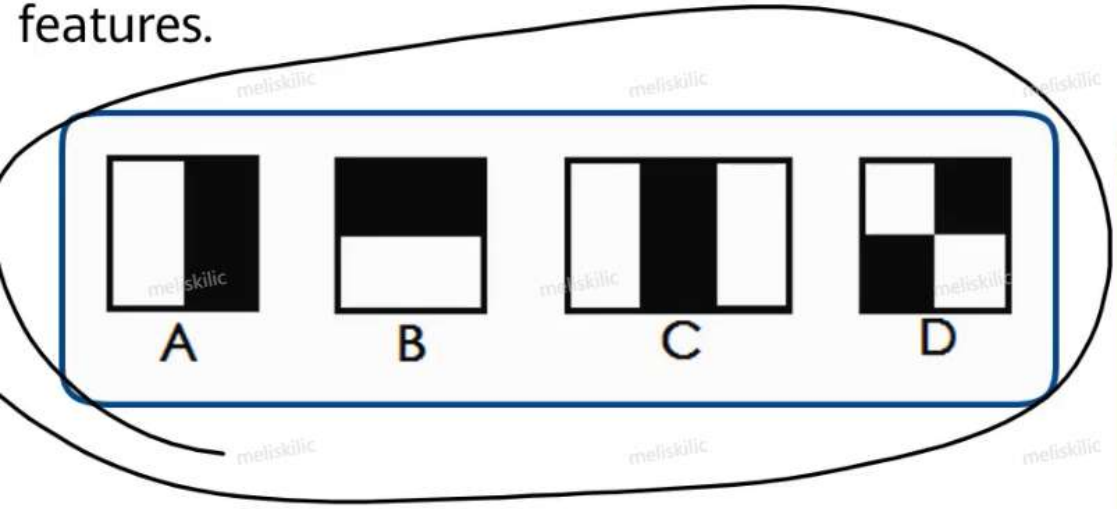
LBP

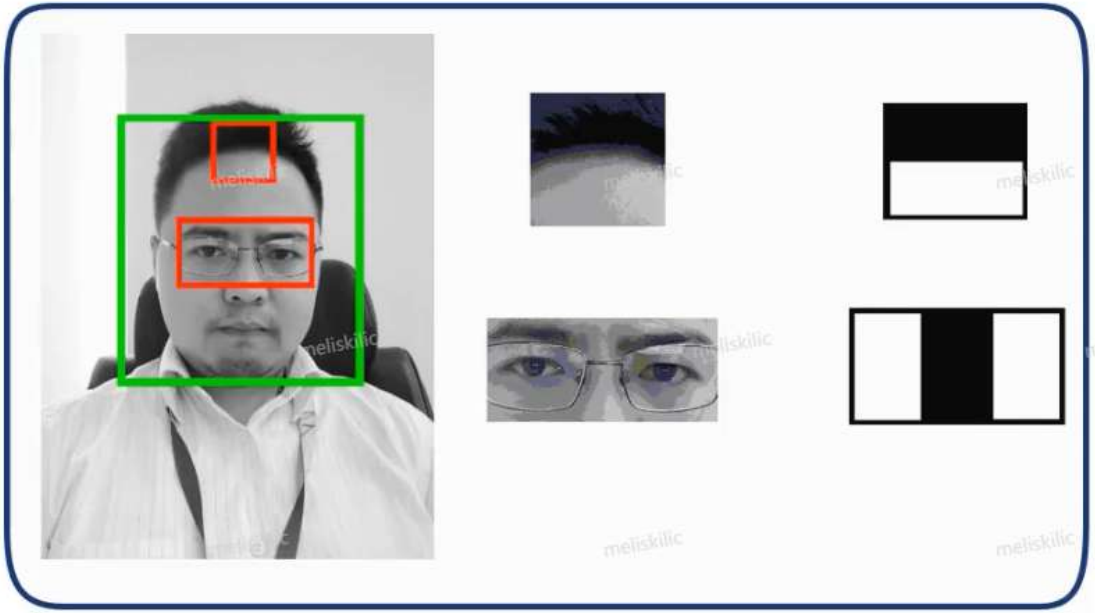
The LBP is a descriptor that can be used to describe local texture features of images. The LBP has obvious advantages such as rotation invariance and intensity invariance.



Haar

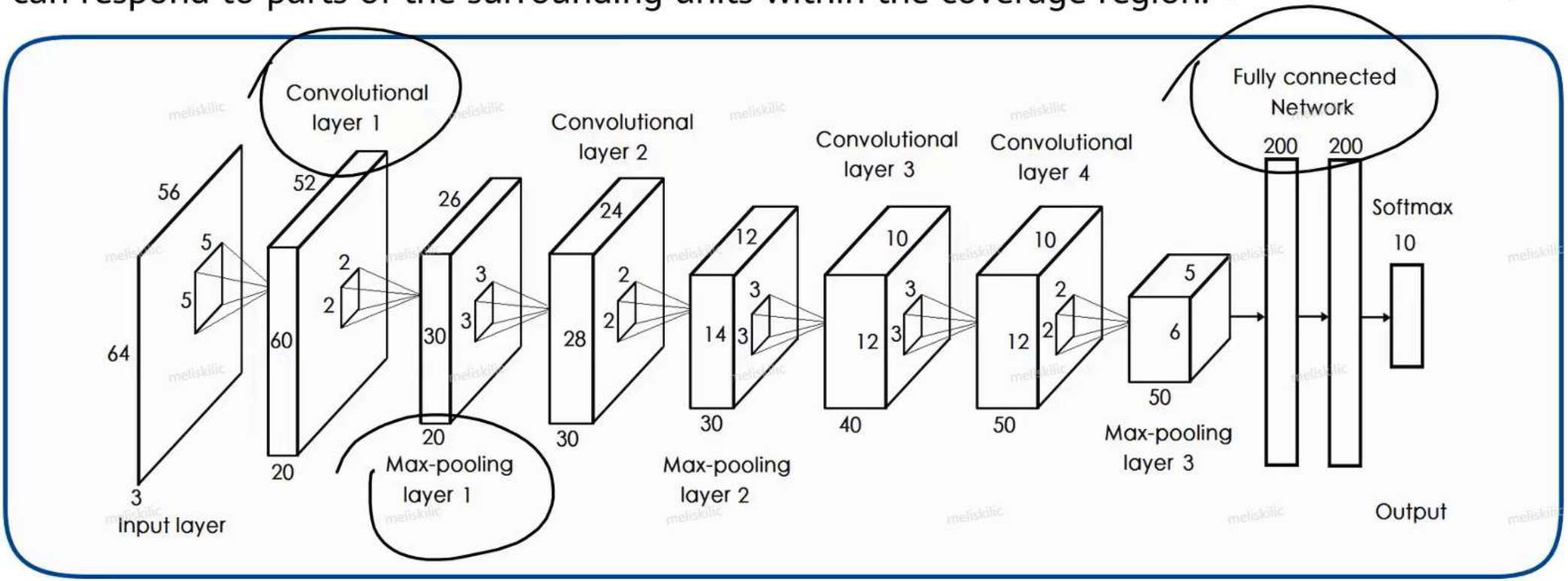
The Haar-like feature is a feature extraction descriptor. The Haar-like feature defines the following basic structures, which can be used to extract edge, linear, center, and diagonal features.





Introduction to Convolutional Neural Networks

A convolutional neural network (CNN) is a feedforward neural network. Its artificial neurons can respond to parts of the surrounding units within the coverage region.



Convolution

Information theory

$$y(n) = h(n) * x(n) = \int_{-\infty}^{\infty} h(\alpha)x(n-\alpha)d\alpha$$

Matrix theory

$$G = H * F$$

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v]F[i-u, j-v]$$

Deep learning

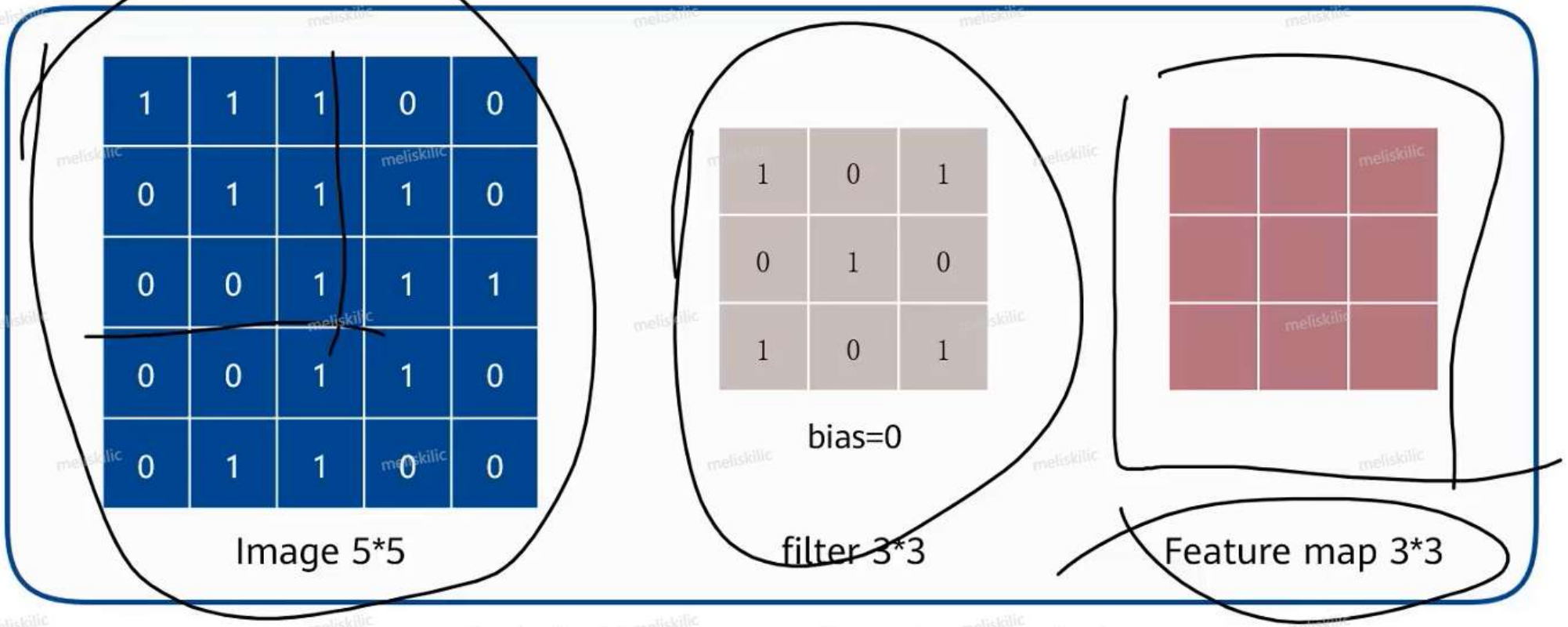
$$FeatureMap = Kernel * Input$$

$$FM[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} K[u,v]I\ [i+u, j+v]$$

$$FM[i,j] = \sum_{d} \sum_{u=-k}^{k} \sum_{v=-k}^{k} K[u,v,d]I\ [i+u, j+v, d]$$

Single - Kernel Convolution Calculation (1)

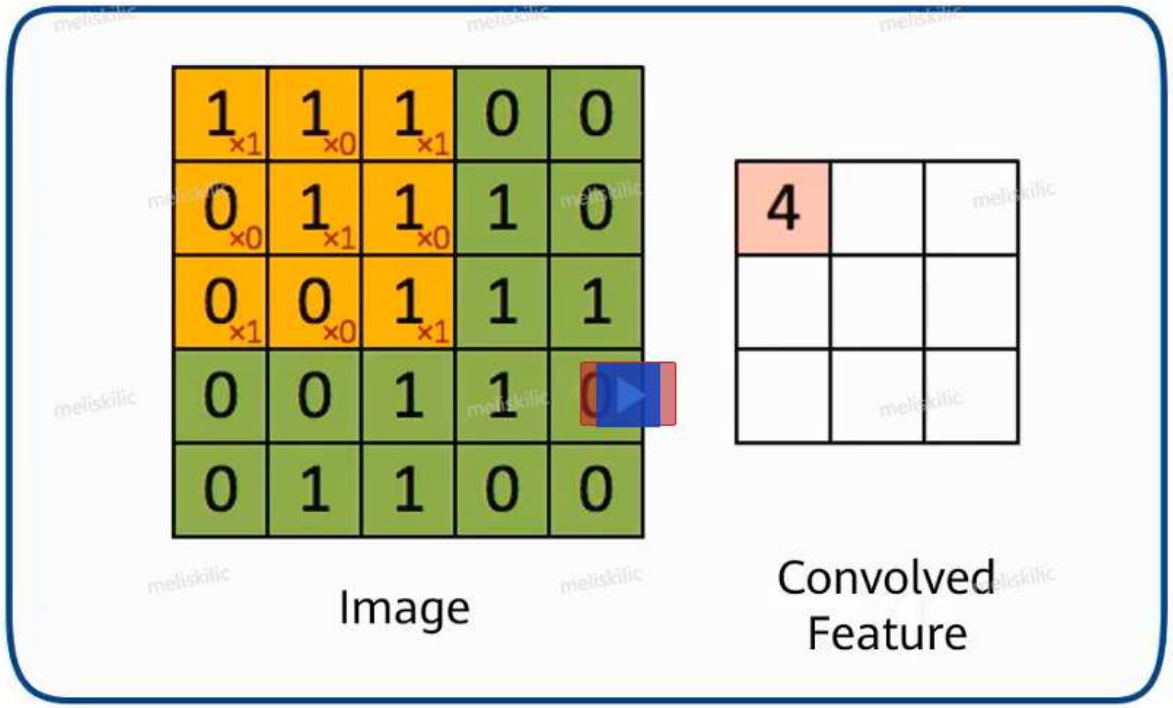
Convolution calculation description



stanford.edu, Feature extraction using convolution

Single - Kernel Convolution Calculation (2)

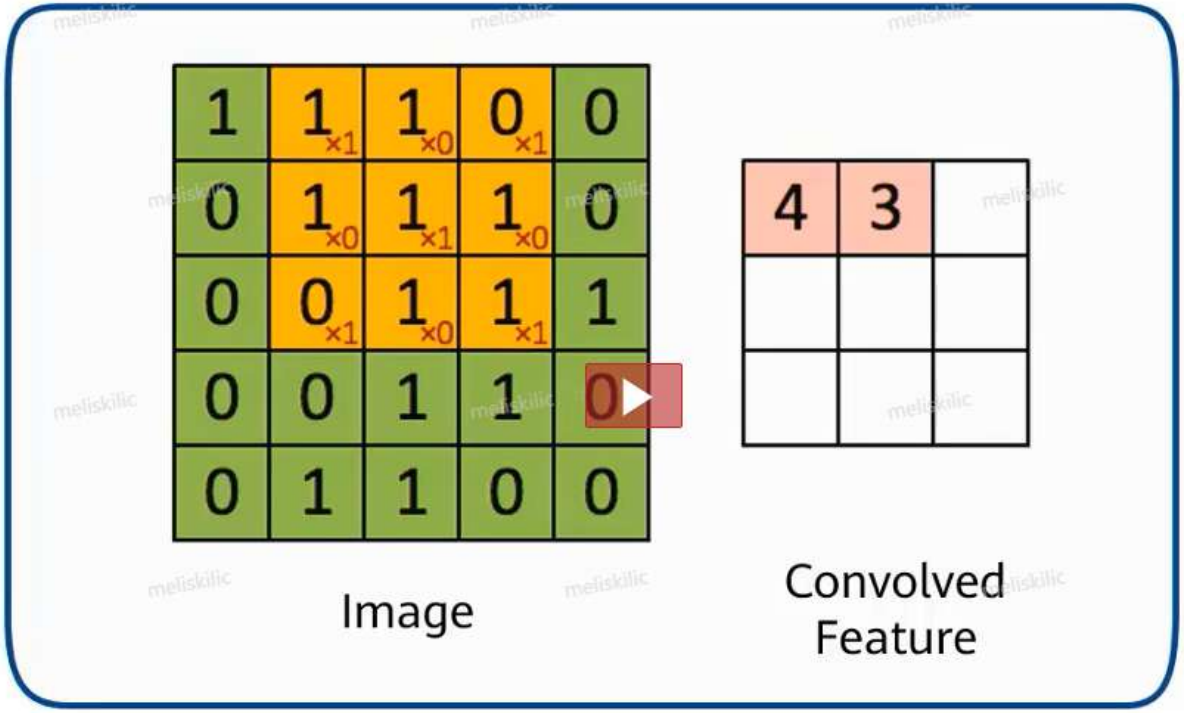
Demonstration of the convolution calculation



stanford.edu, Feature extraction using convolution

Single - Kernel Convolution Calculation (2)

Demonstration of the convolution calculation



stanford.edu, Feature extraction using convolution

Multi - Kernel Convolution Calculation

Input image

5 x 5 x 3

Padding = 1

Two convolution

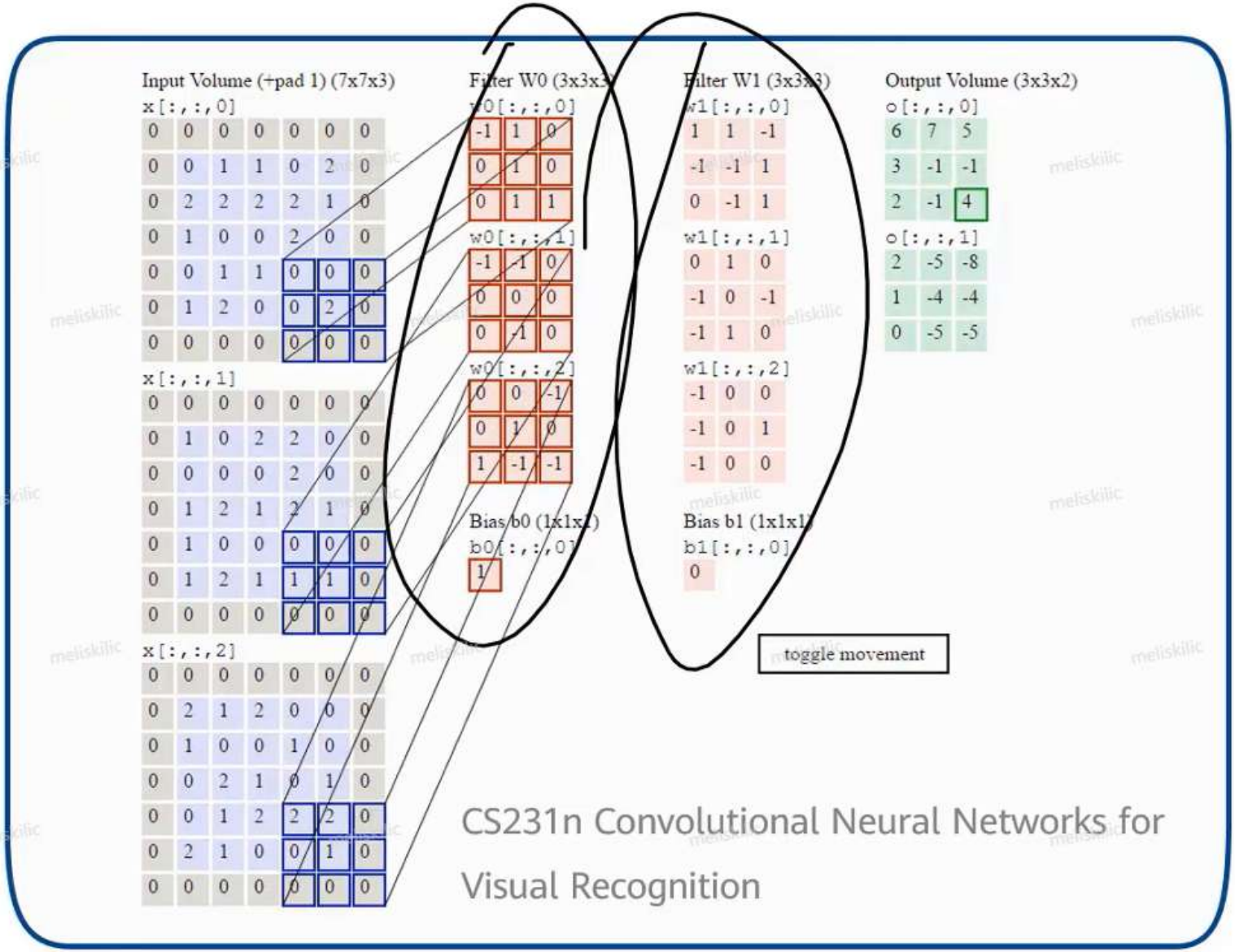
kernels

3 x 3 x 3

Stride = 2

Feature map

3 x 3 x 2



Important Convolutional Concepts

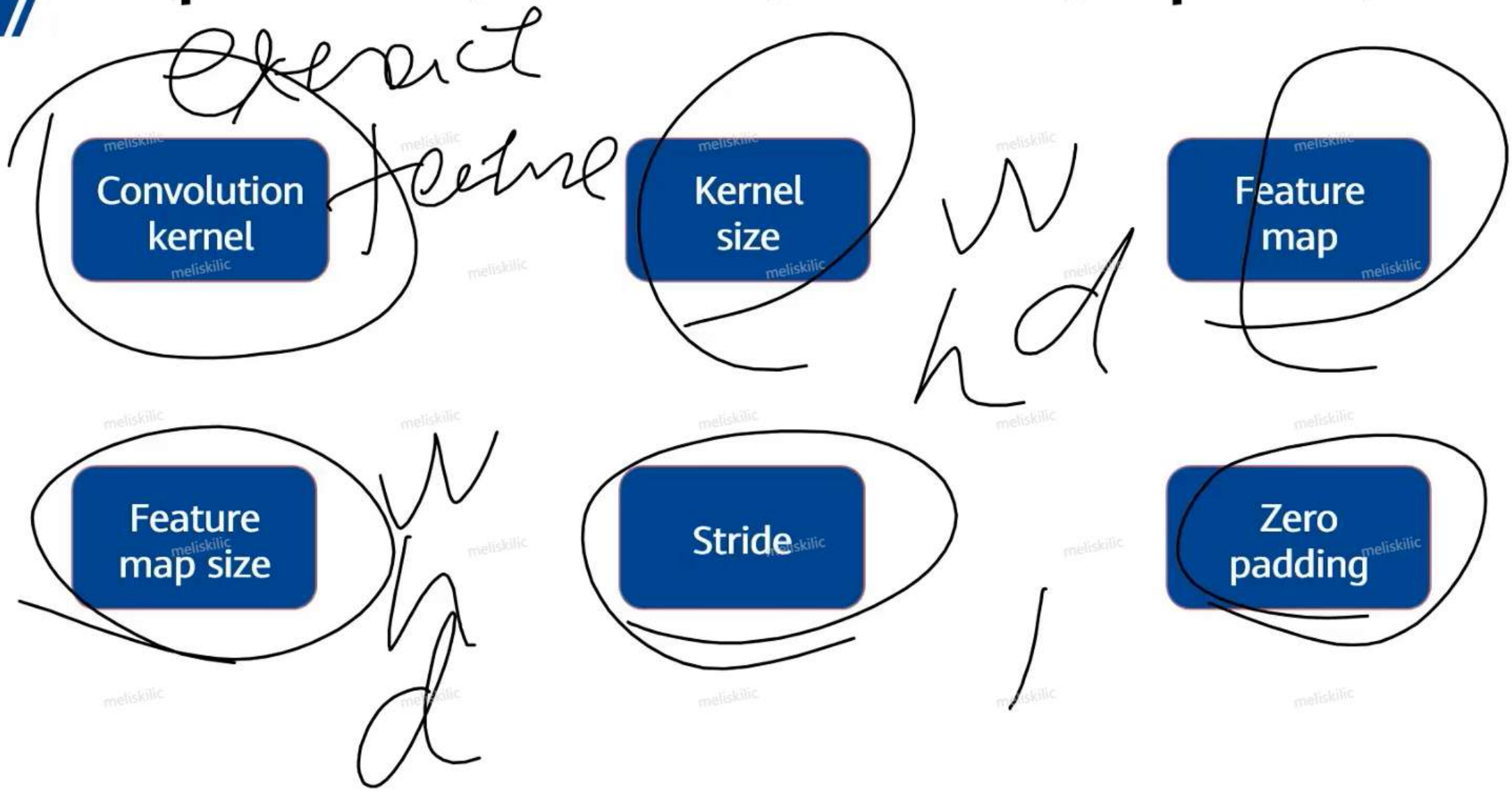


Image Invariance

Invariance means that an object can be recognized even when its appearance varies in some way.

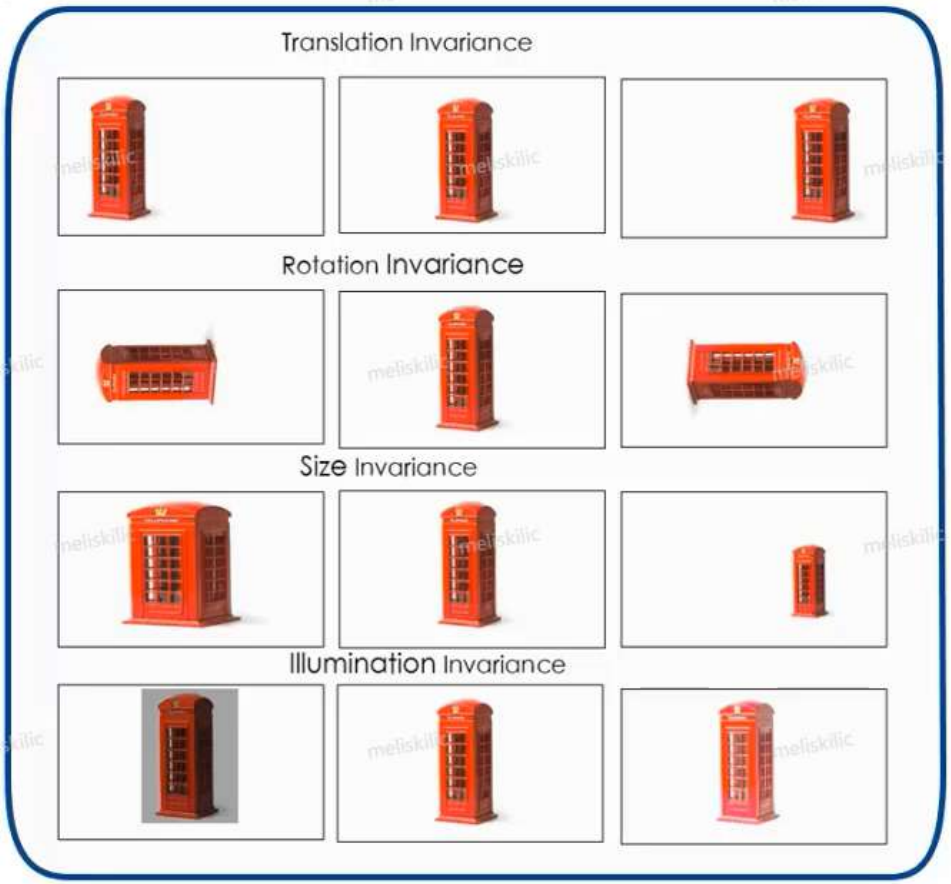
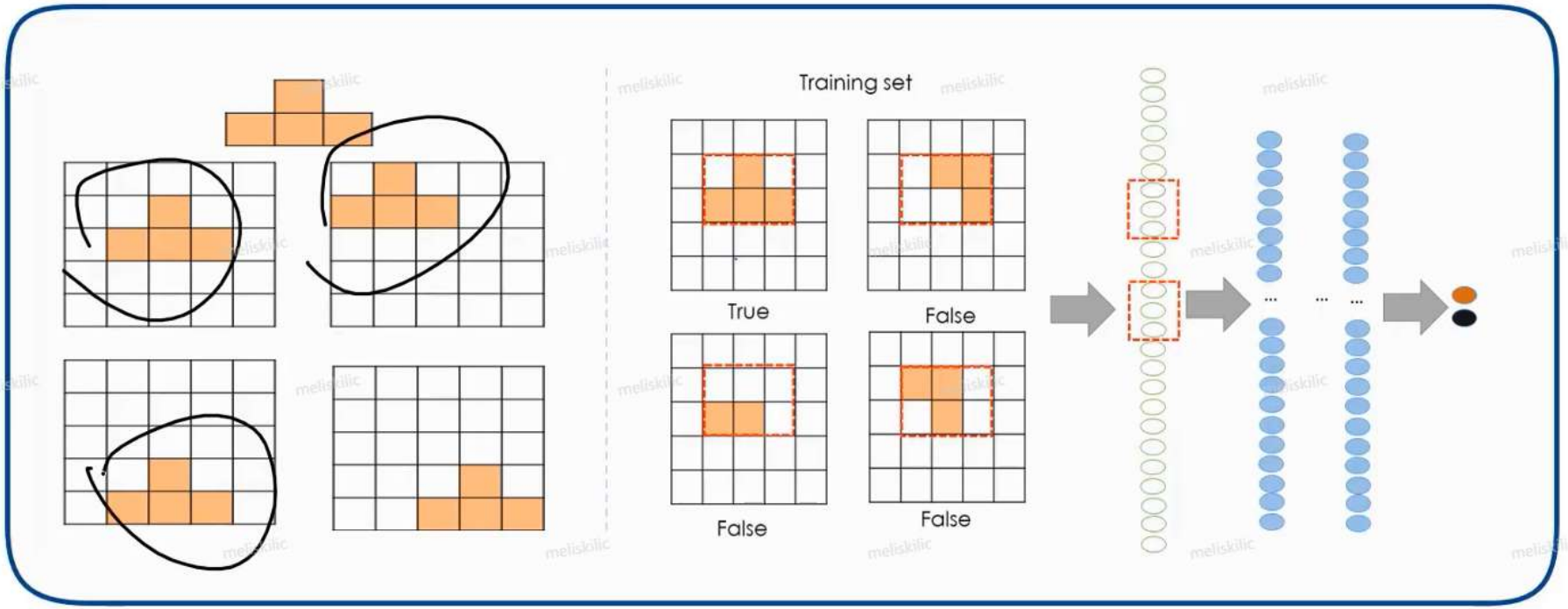
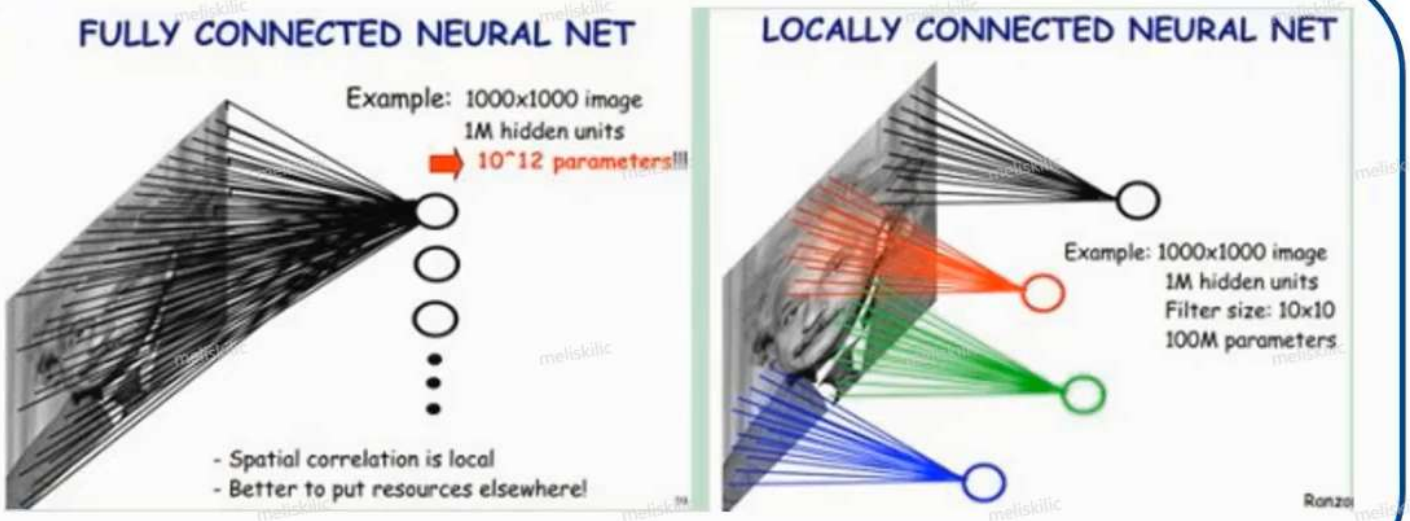


Image Invariance

Invariance means that an object can be recognized even when its appearance varies in some way.



CNN Bright Spot-Local Receptive Field



R. Fergus. *M-A Ranzatto Deep Learning Tutorial*. CVPR.

Implementation:

make the kernel much smaller than the input image.

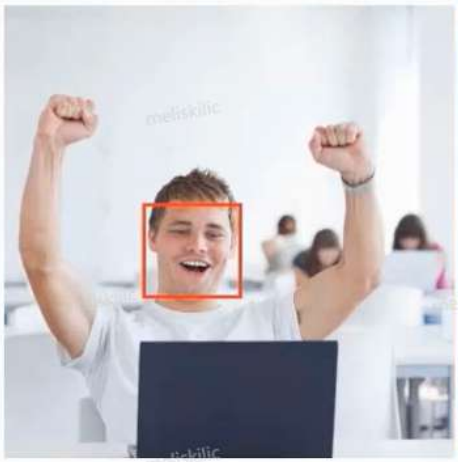
Merits:

- Reduce memory requirement
- Simulate the neurons in the visual cortex that collect information locally

The local receptive field is used to extract local details.

In CNNs, each neuron only needs to collect local information.

CNN Bright Spot-Parameter Sharing


Implementation:

Use the kernel with the same parameters to scan the input.

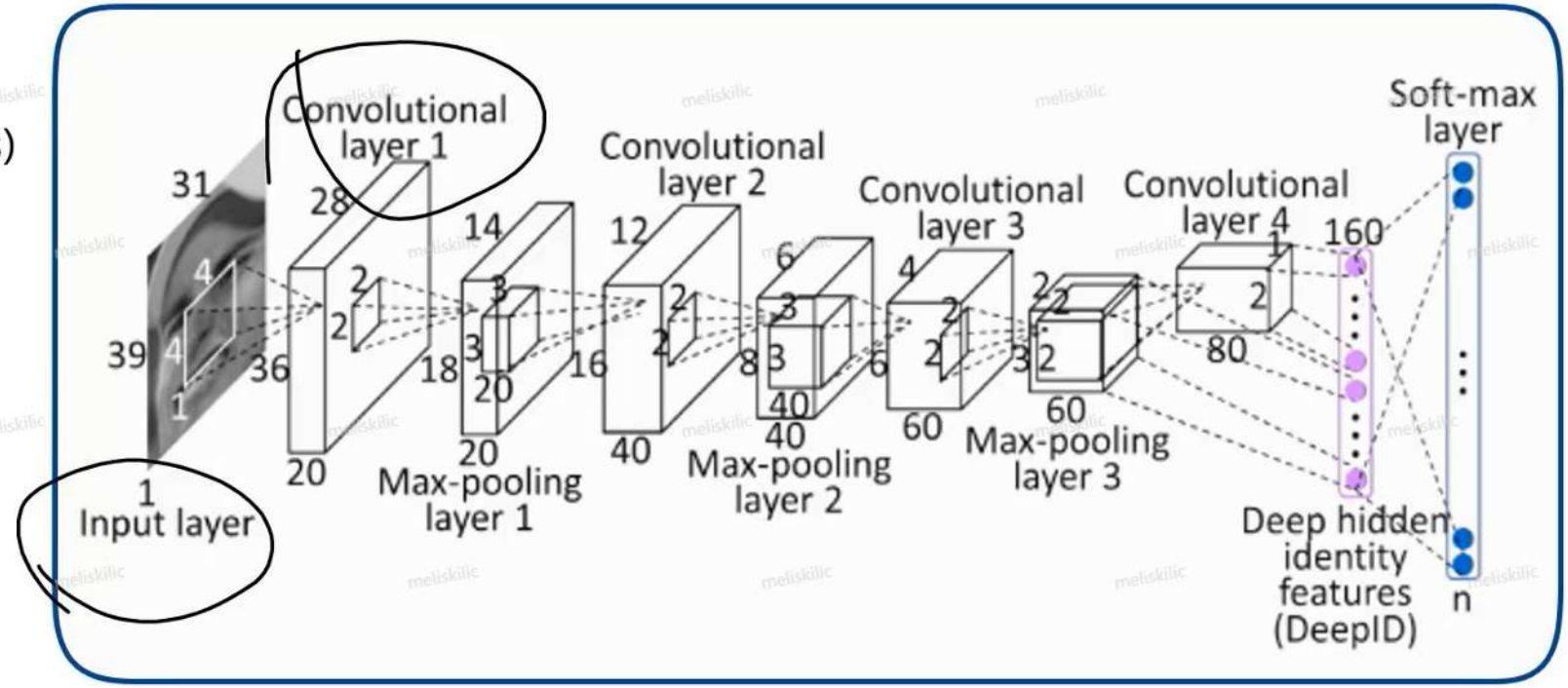
Merits:

- Implement position invariance.
- Reduce memory requirement.

The parameter sharing feature is used to implement position invariance.

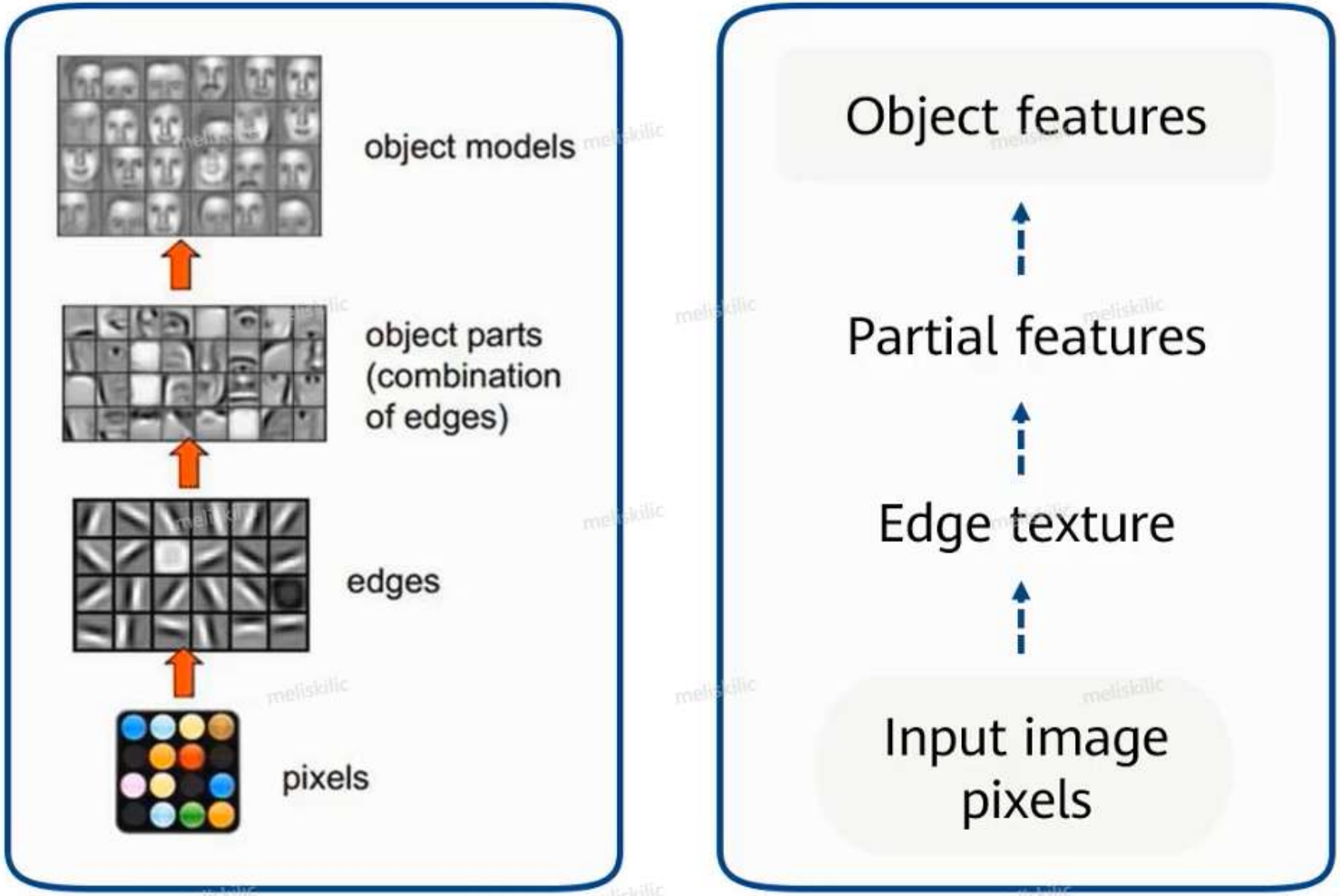
CNN Architecture

- Input layer
- Convolutional layer (4 layers)
- Max pooling layer (3 layers)
- Fully connected layer
- Output layer



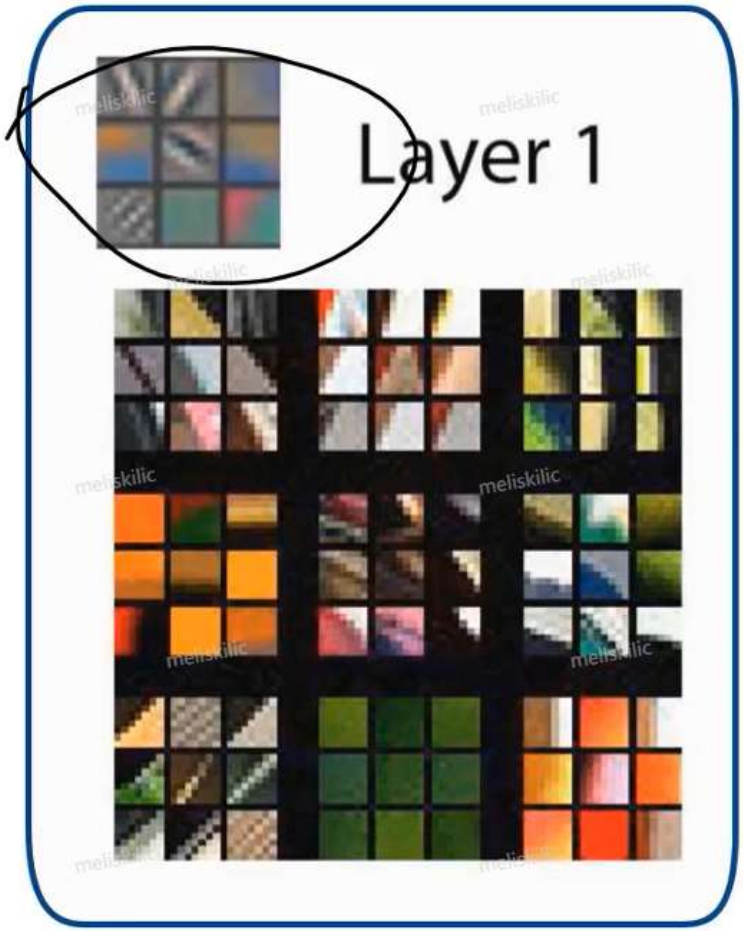
Sun Y, Deep learning face representation from predicting 10,000 classes

Convolutional Layer Function



Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations

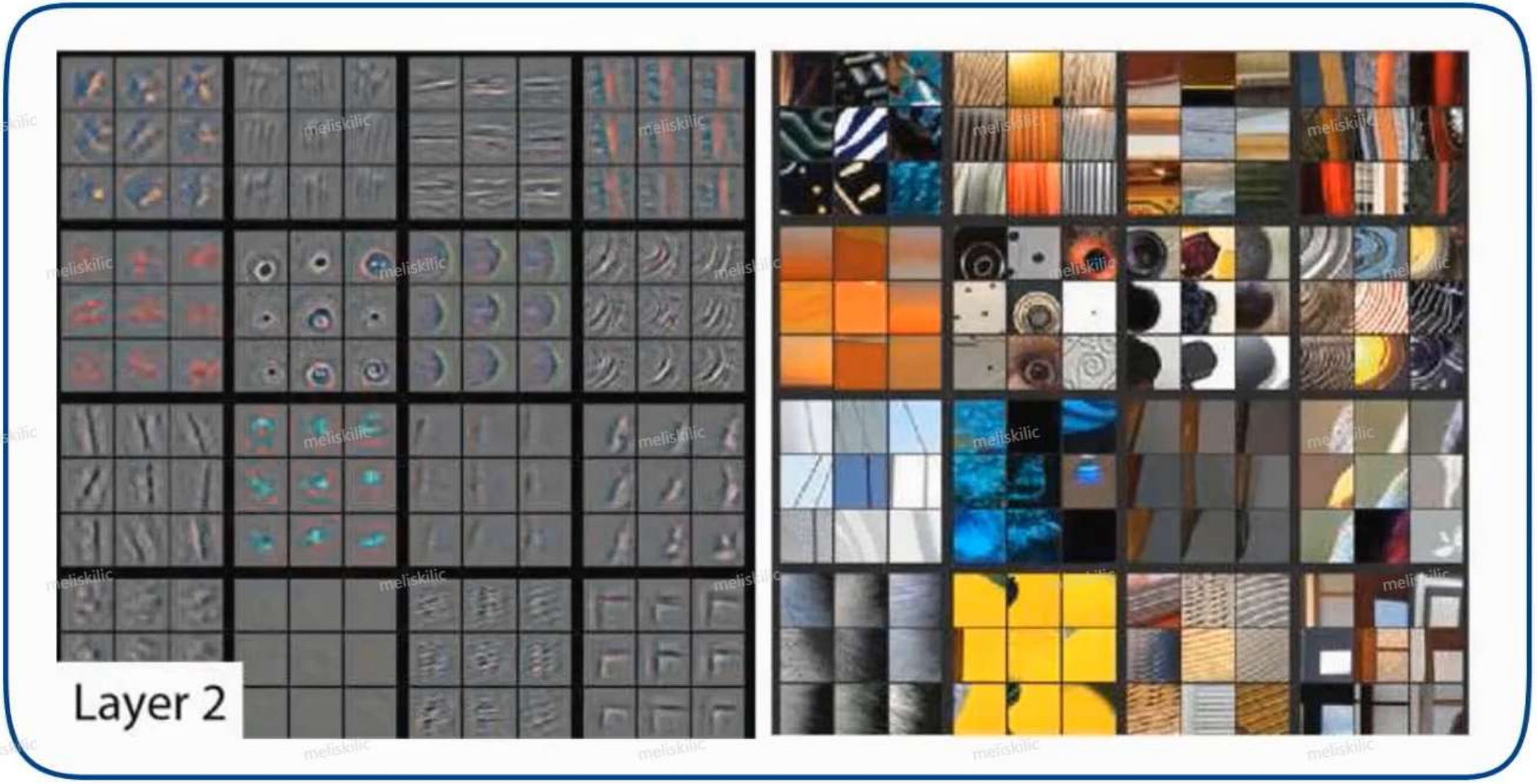
Convolution Effect (Layer 1)



- In the first figure, the weights of a convolution kernel are visualized.
- The second figure shows the local information of an input image generated by the convolution kernel.

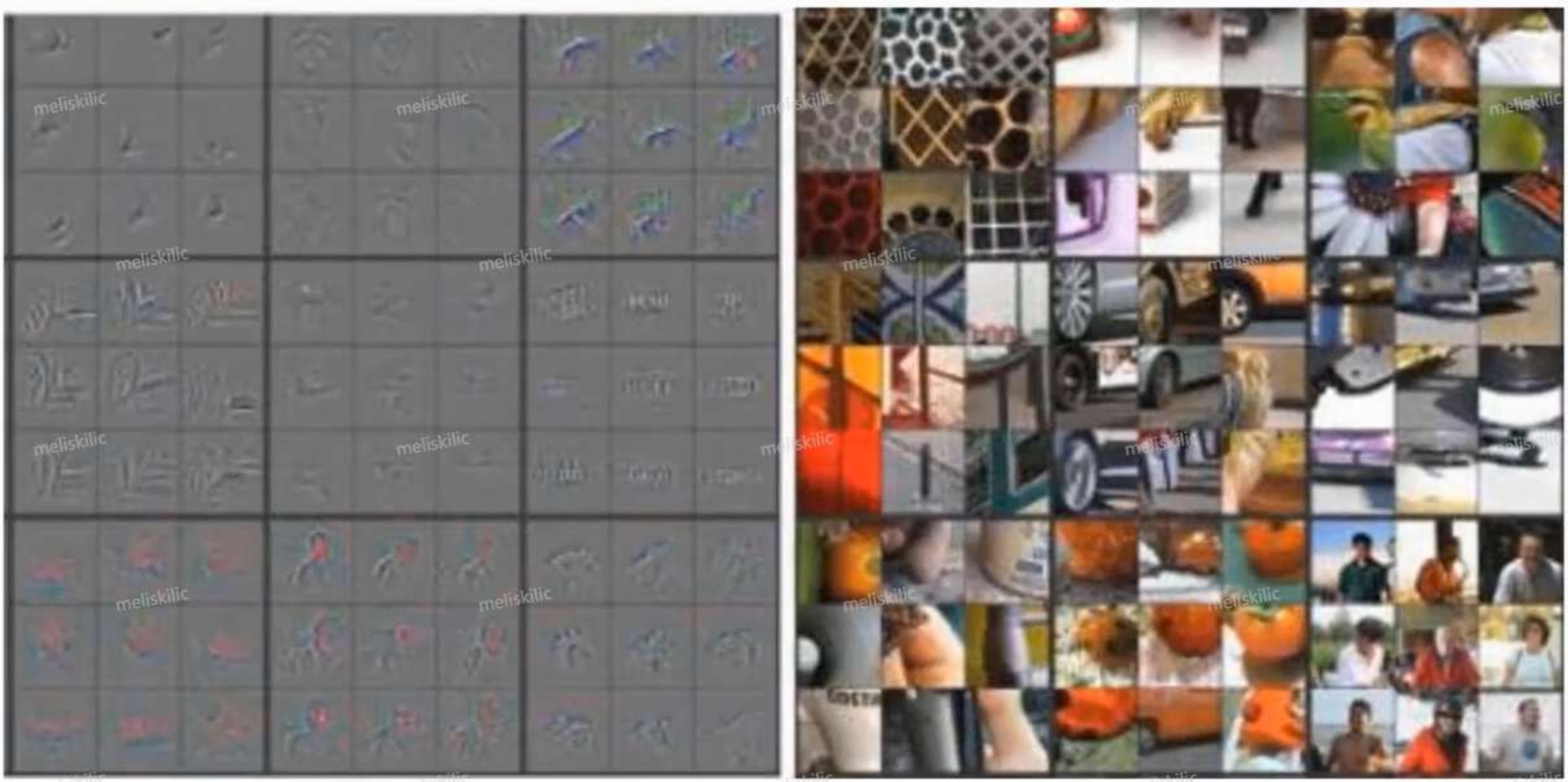
Matthew D. Zeiler. *Visualizing and Understanding Convolutional Networks*.

Convolution Effect (Layer 2)



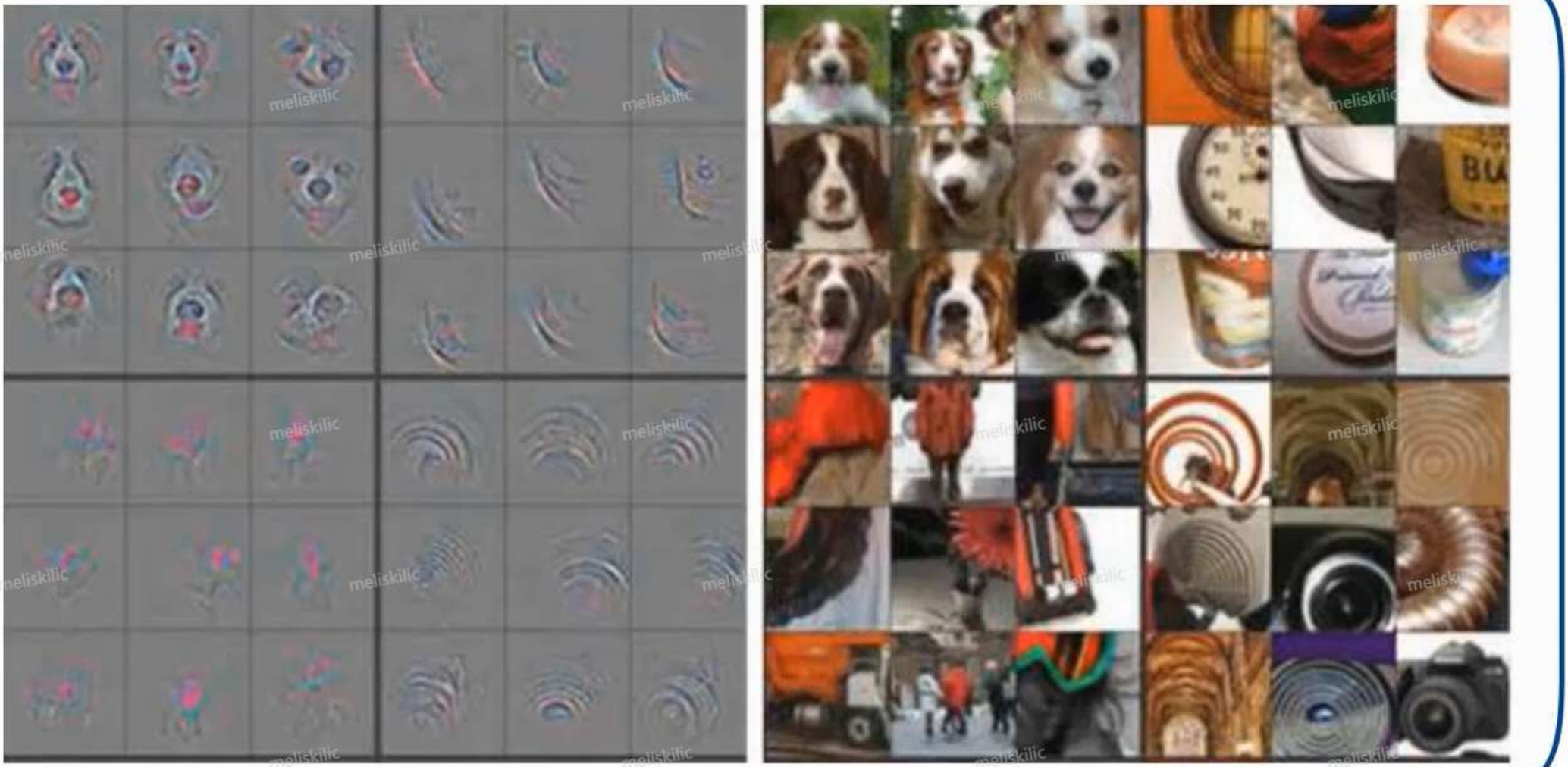
Matthew D. Zeiler. *Visualizing and Understanding Convolutional Networks.*

Convolution Effect (Layer 3)



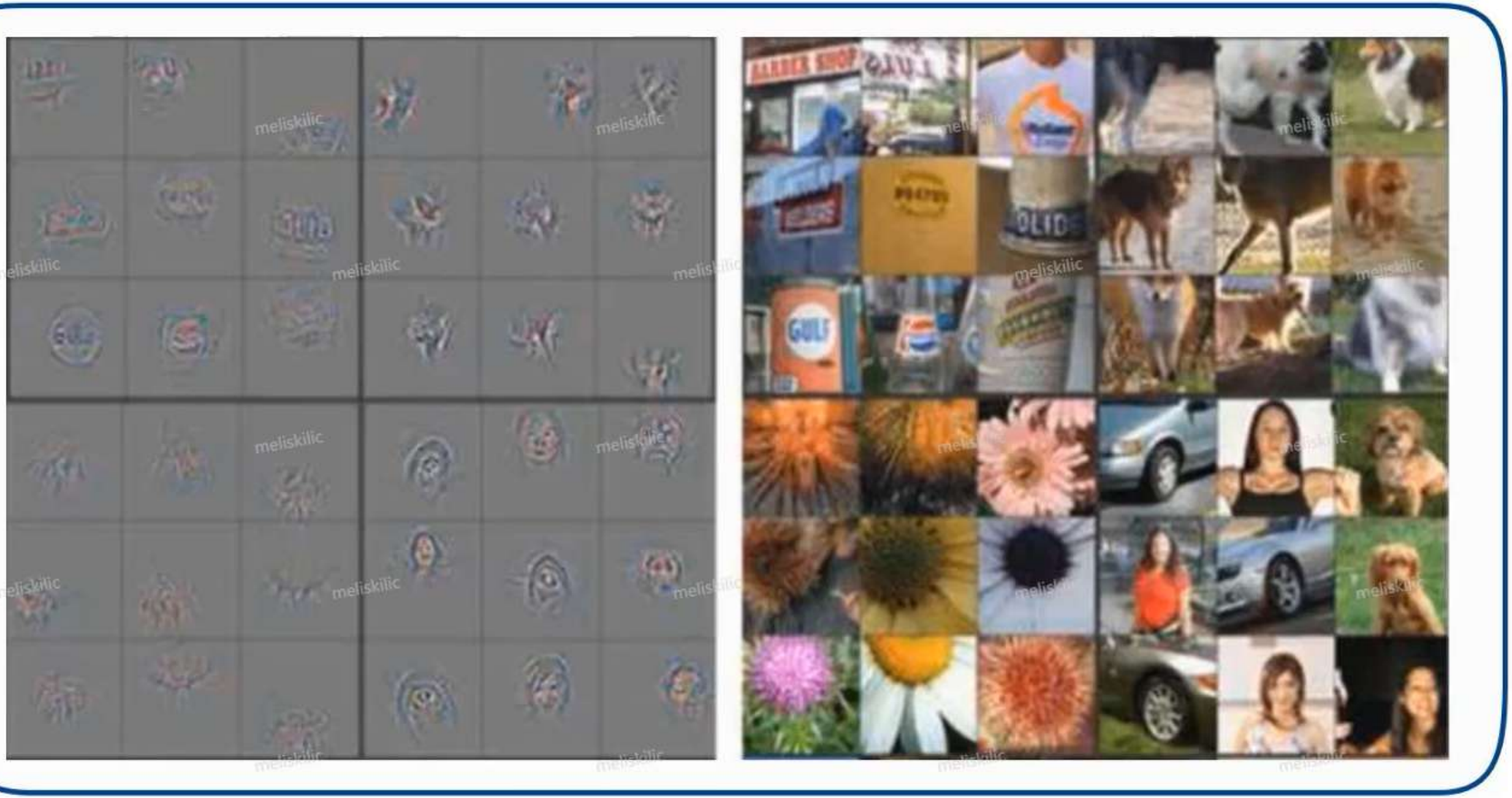
Matthew D. Zeiler. *Visualizing and Understanding Convolutional Networks.*

Convolution Effect (Layer 4)



Matthew D. Zeiler. *Visualizing and Understanding Convolutional Networks.*

Convolution Effect (Layer 5)



Matthew D. Zeiler. *Visualizing and Understanding Convolutional Networks*.

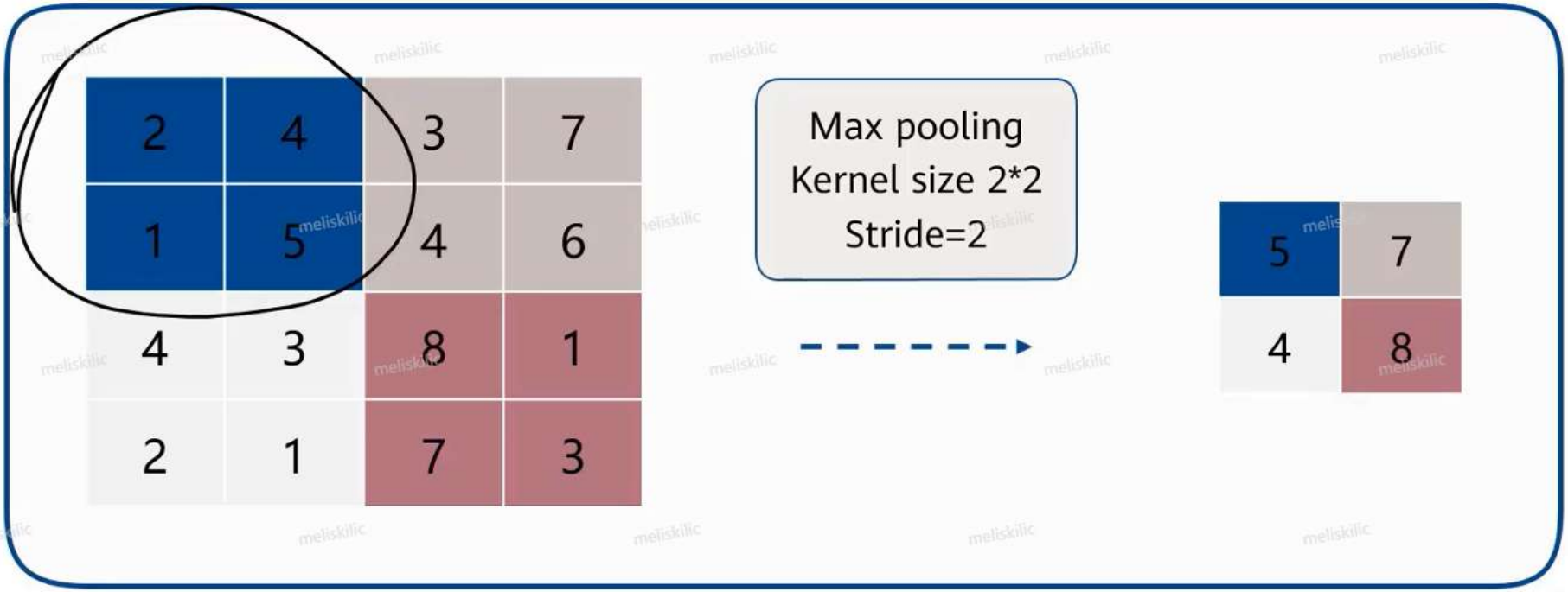
Pooling Layer

The purpose of pooling is to reduce the spatial size of feature maps.

Dishashree Gupta. *Architecture of Convolutional Neural Networks (CNNs) demystified.*

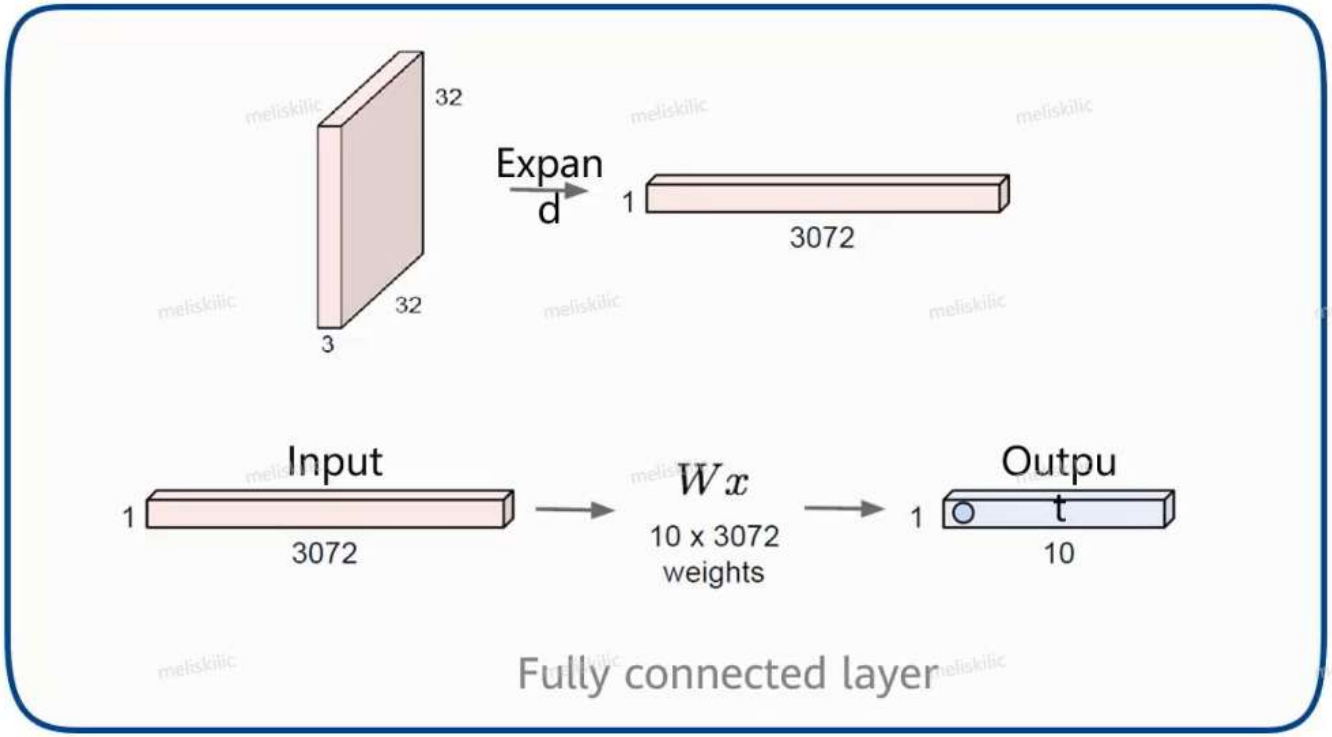
Max Pooling

For example, we set the stride to 2 and the pooling filter size to 2. The max pooling is also applied for adjusting the size of the convolution layer output.



Fully Connected Layer

Expand the feature map of the last convolutional layer before calculating it in the fully connected layer.



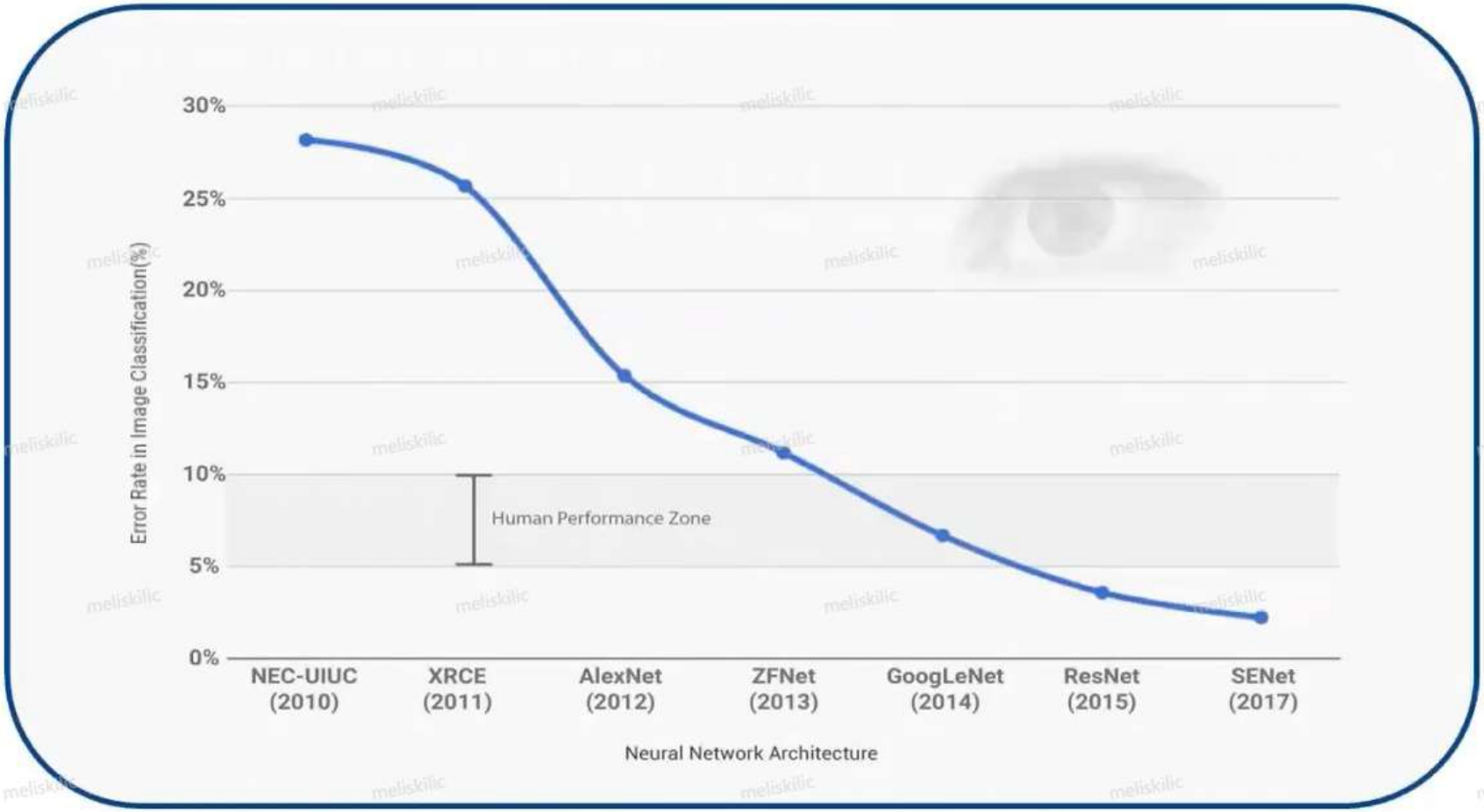
ILSVRC

Stanford University

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
- About 1.2 million images and labels
- 1000 categories
- the top-5 and top-1 error rates are used as the evaluation indicators

ILSVRC Historical Achievements

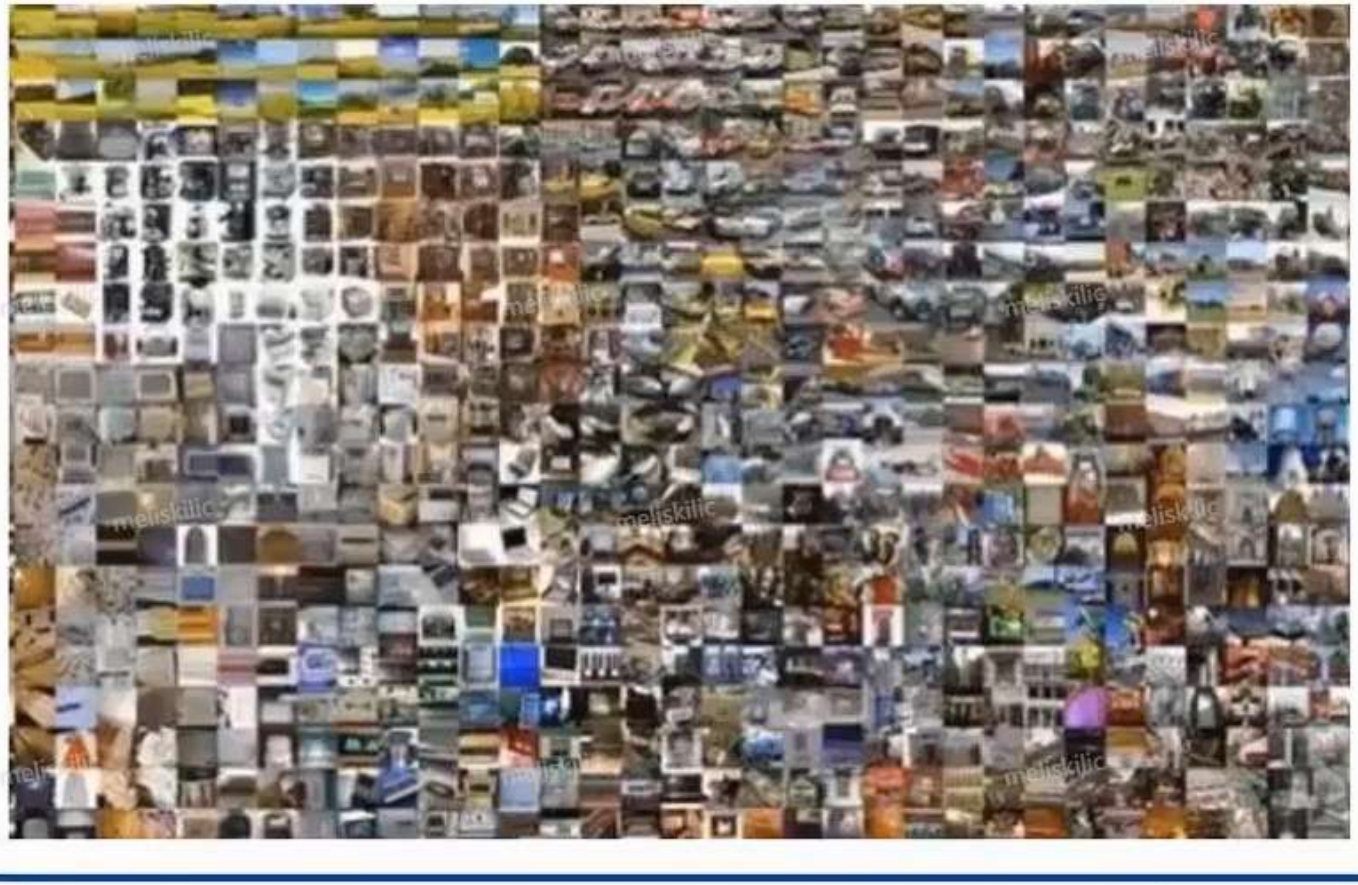
- image classification
- single-object locating
- object detection



https://twitter.com/hashtag/ilsvrc

ImageNet

- Was founded in 2007 by Li Feifei.
- aims to collect a large amount of image data.
- contains 15 million labeled high-resolution images of objects.

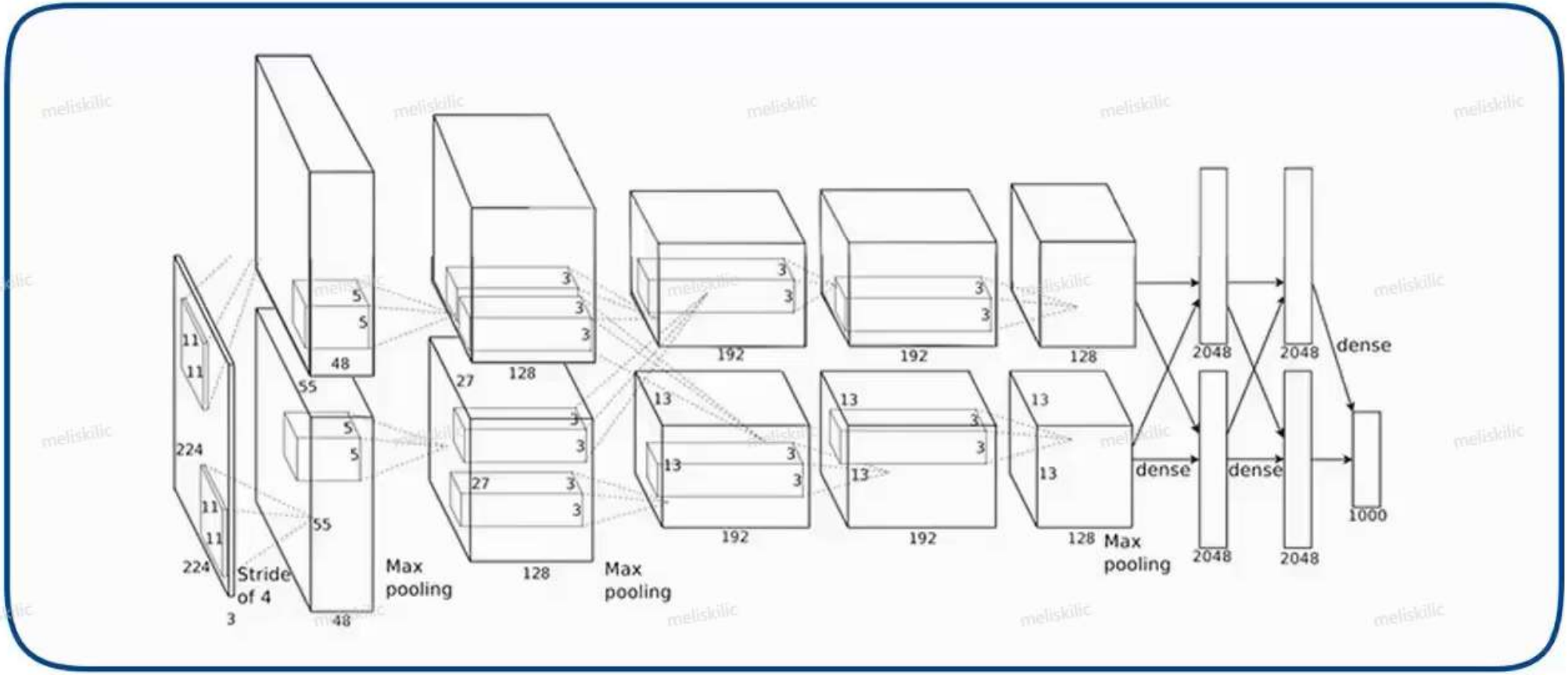


www.image-net.org

AlexNet

AlexNet, 2012

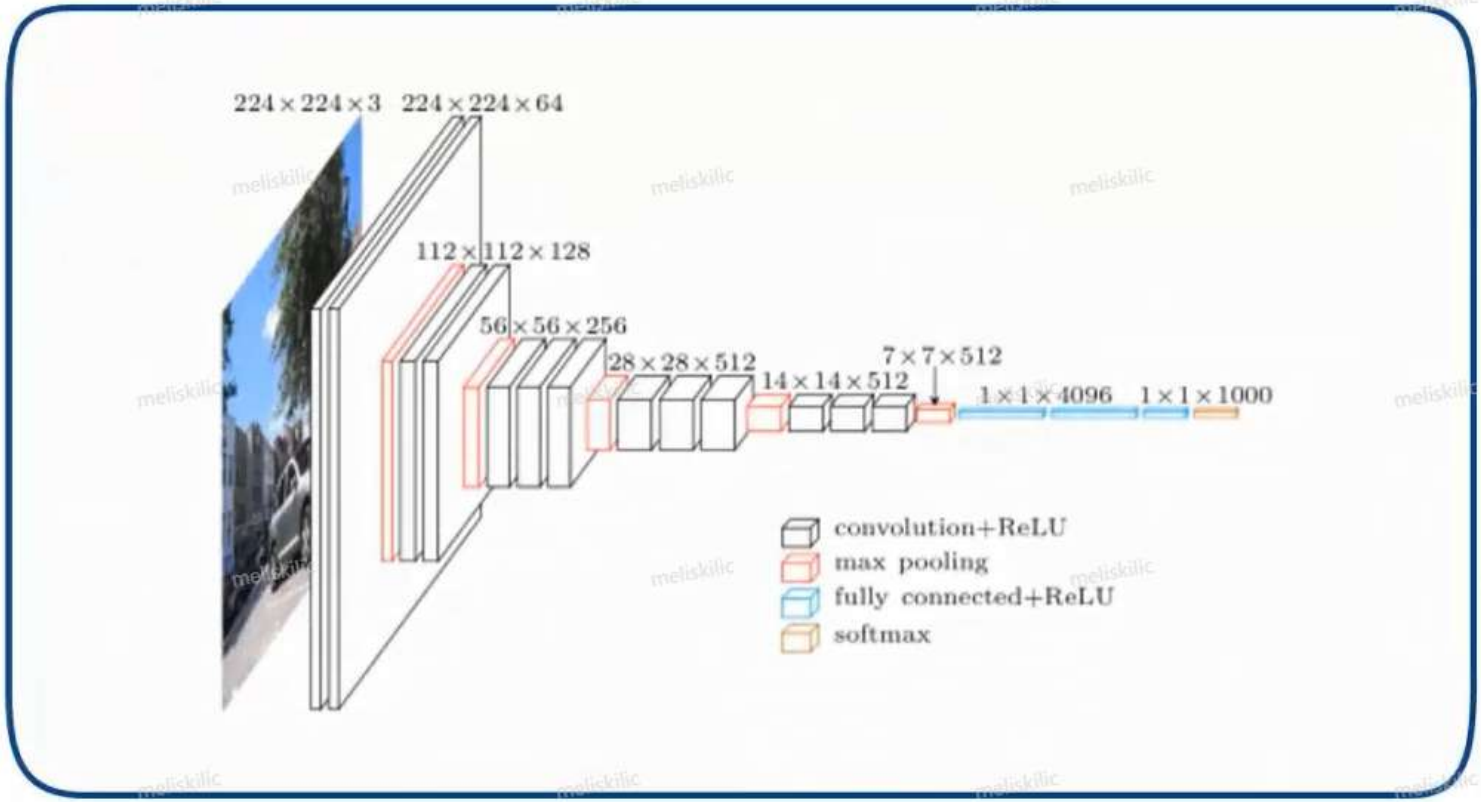
ReLU, overlapping pooling, data augmentation, dropout



Alex. *ImageNet Classification with Deep Convolutional Neural Networks.*

VGGNet

VGGNet (Visual Geometry Group), 2014



Visual Geometry Group. *VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION.*

Six Configurations of the VGG

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Visual Geometry Group. *VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION.*

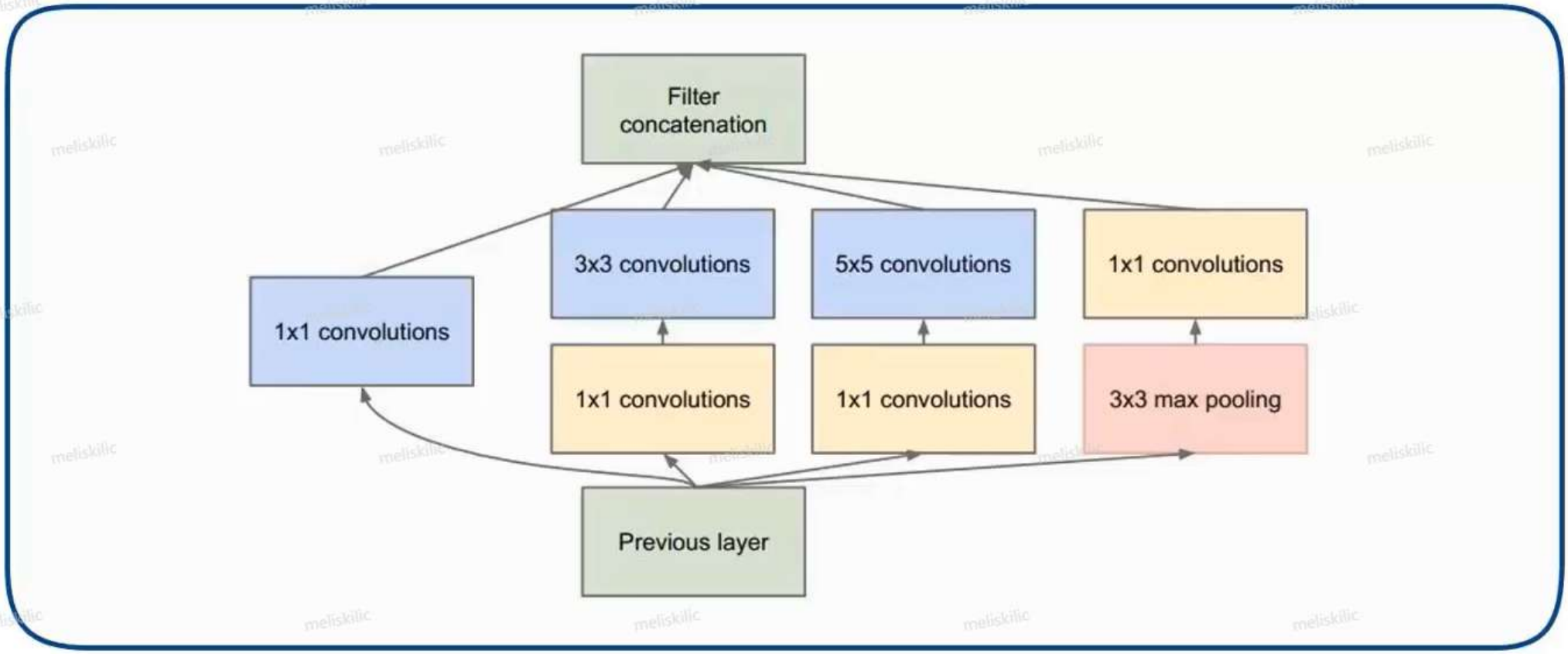
GoogLeNet

GooLeNet, 2014

Christian Szegedy. *Going Deeper with Convolutions.*

Inception Architecture

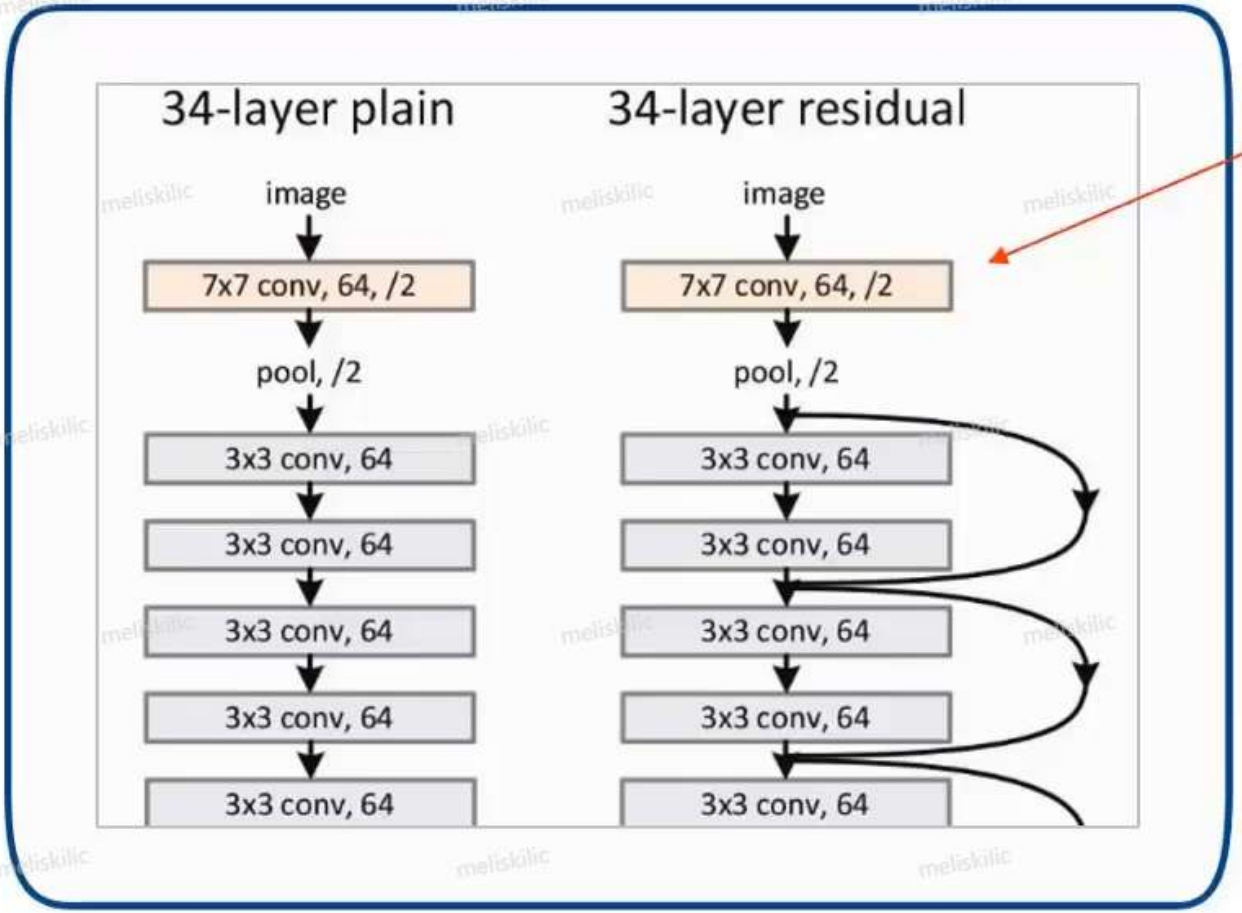
GoogLeNet uses the Inception architecture with substructures connected in parallel.

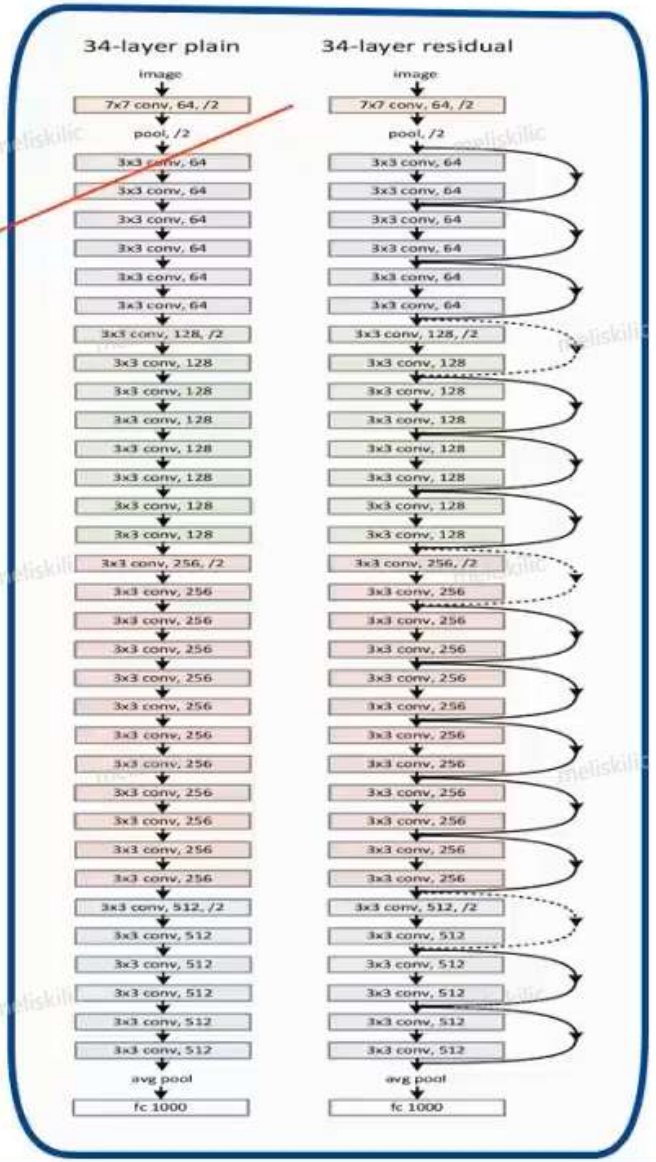


Christian Szegedy. *Going Deeper with Convolutions*.

ResNet

ResNet, 2015

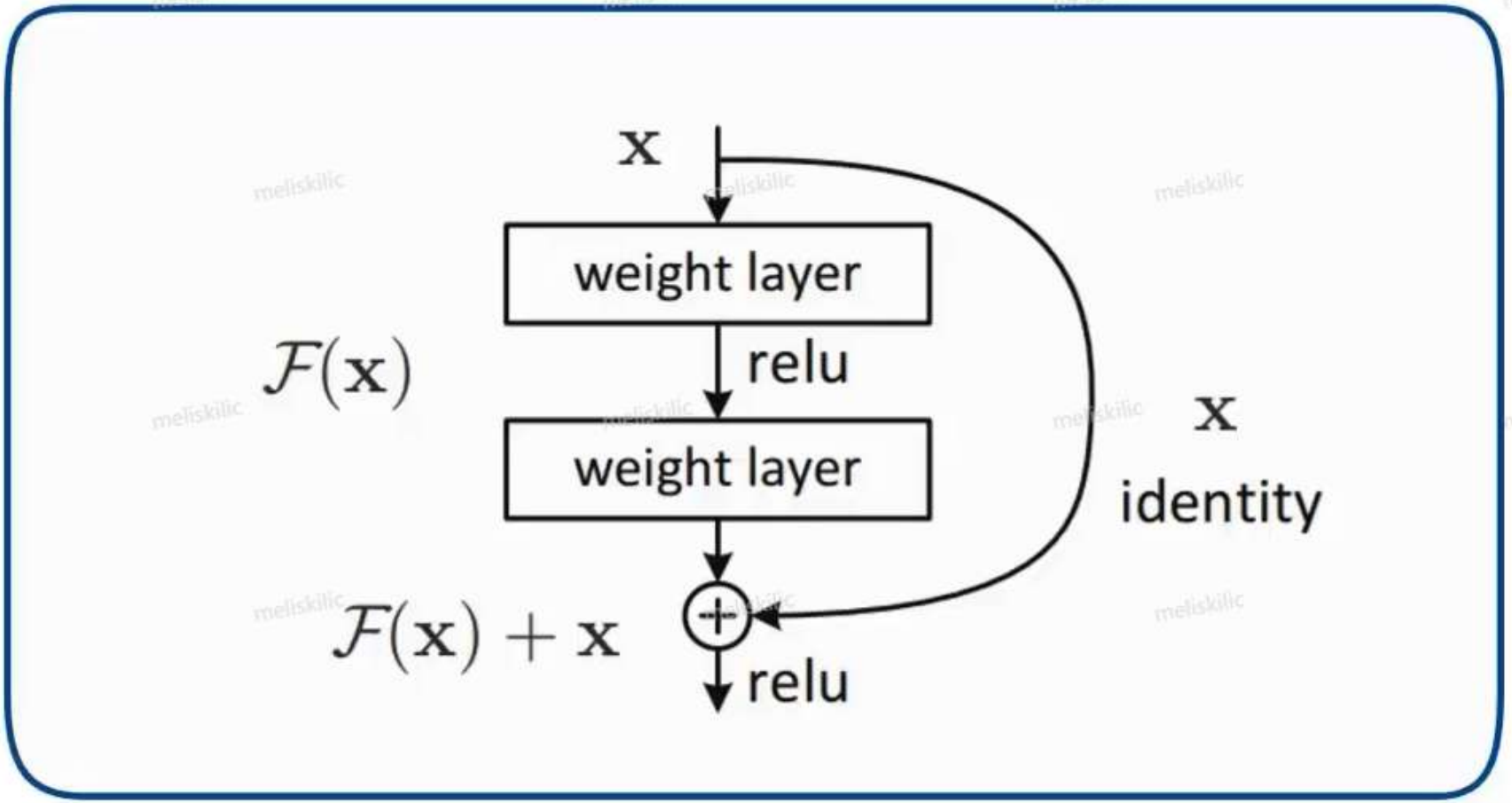




Kaiming He. *Deep Residual Learning for Image Recognition.*

Residual Architecture

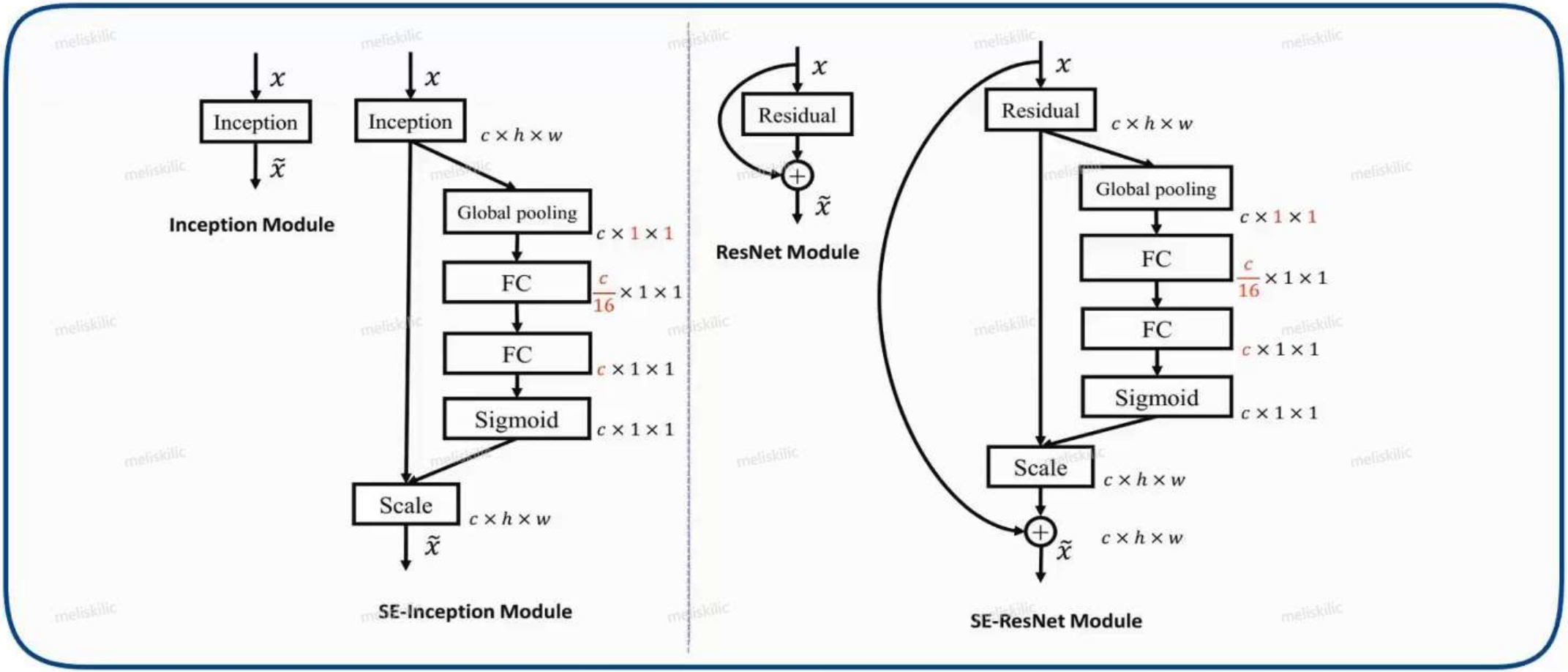
Residual architecture proposed by ResNet



Kaiming He. *Deep Residual Learning for Image Recognition.*

SENet

Squeeze-and-Excitation Networks (SENet), 2017



Jie Hu. *Squeeze-and-Excitation Networks*.