Perceptron Networks and Applications

M. Ali Akcayol
Gazi University
Department of Computer Engineering

Content

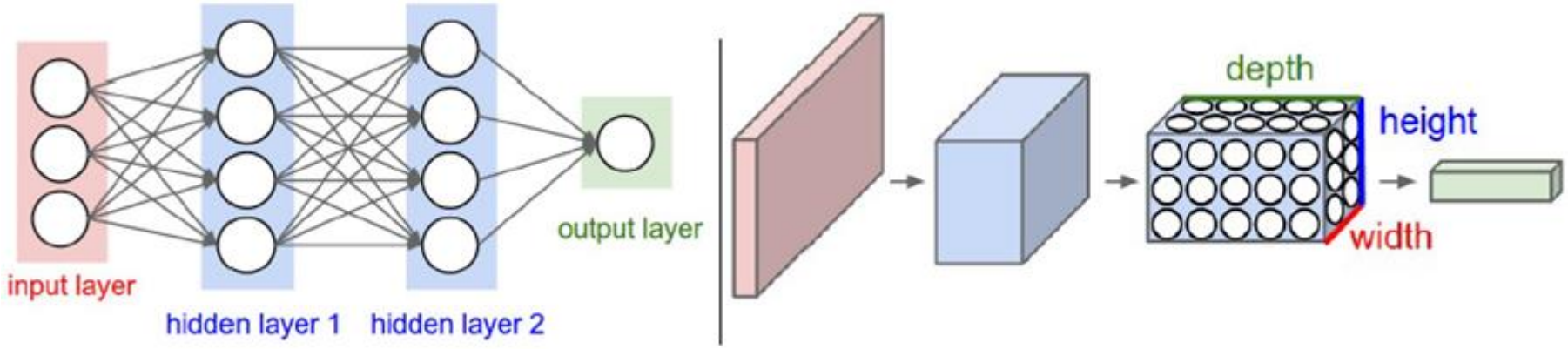
- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

Convolutional neural networks

- Convolutional neural network (CNN) is a special type of artificial neural networks.
- CNNs are deep learning architecture that is widely used especially in image problems.
- A CNN consists of neurons similar to classical neural networks and has a bias and weight values to learn.
- Each neuron takes inputs, combines them, and produces outputs, usually with a non-linear function.
- CNN applications assume the inputs as images and allow us to encode the properties into the architecture.

Convolutional neural networks

- Neurons in CNNs are arranged in three dimensions.



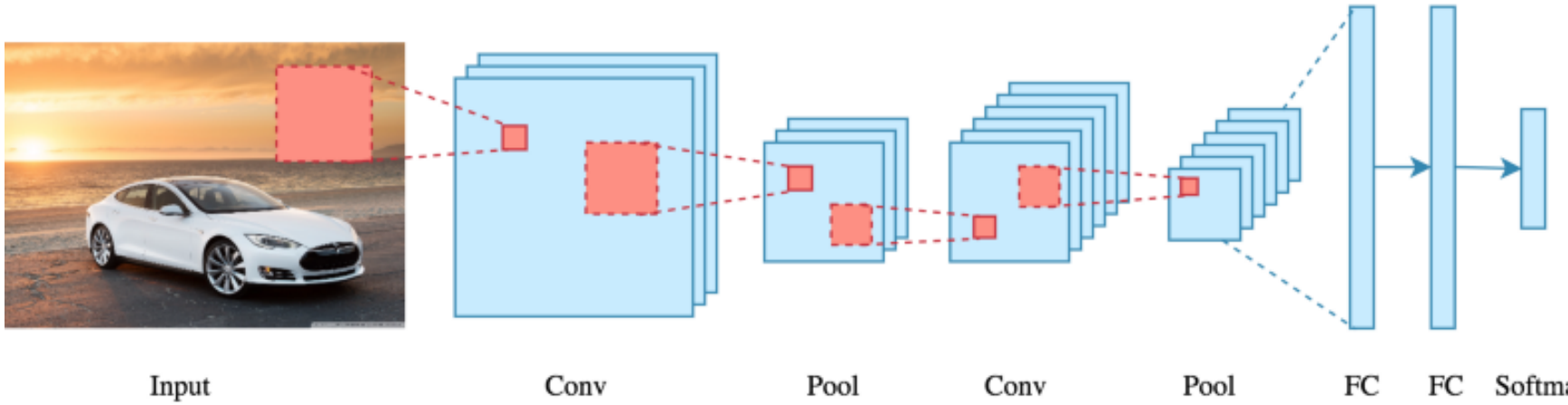
- In CNNs, each layer can receive 3D input and produce 3D output.
- The input layer gets the image.
- The width and height of the input layer is equal to the width and height of the image.
- The depth of the input layer can be 3 (red, green, blue).

Content

- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

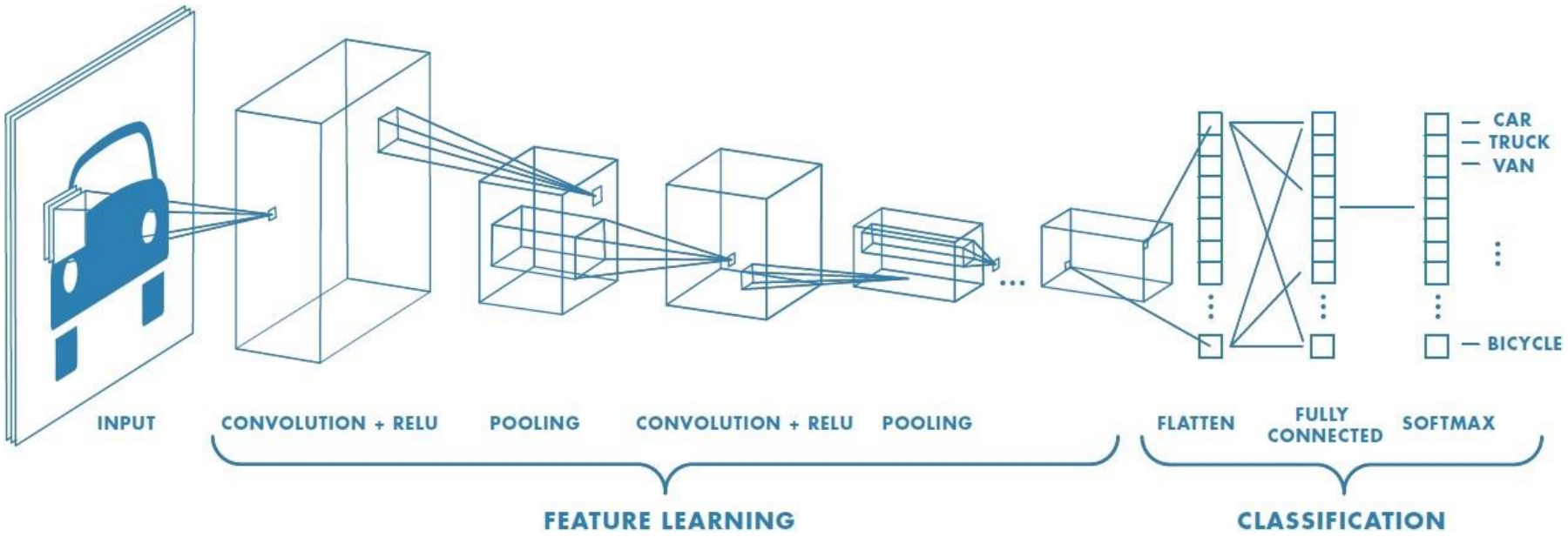
Structure of the CNNs

- CNN uses convolution and pooling operators.
- A CNN has three basic types of layers:
  - Convolutional layer
  - Pooling layer
  - Fully-connected layer
- Multiple convolution+pooling can be done consecutively.
- It then has several fully connected layers.
- In multi-label classification problems, there is a softmax layer at the output.



Structure of the CNNs

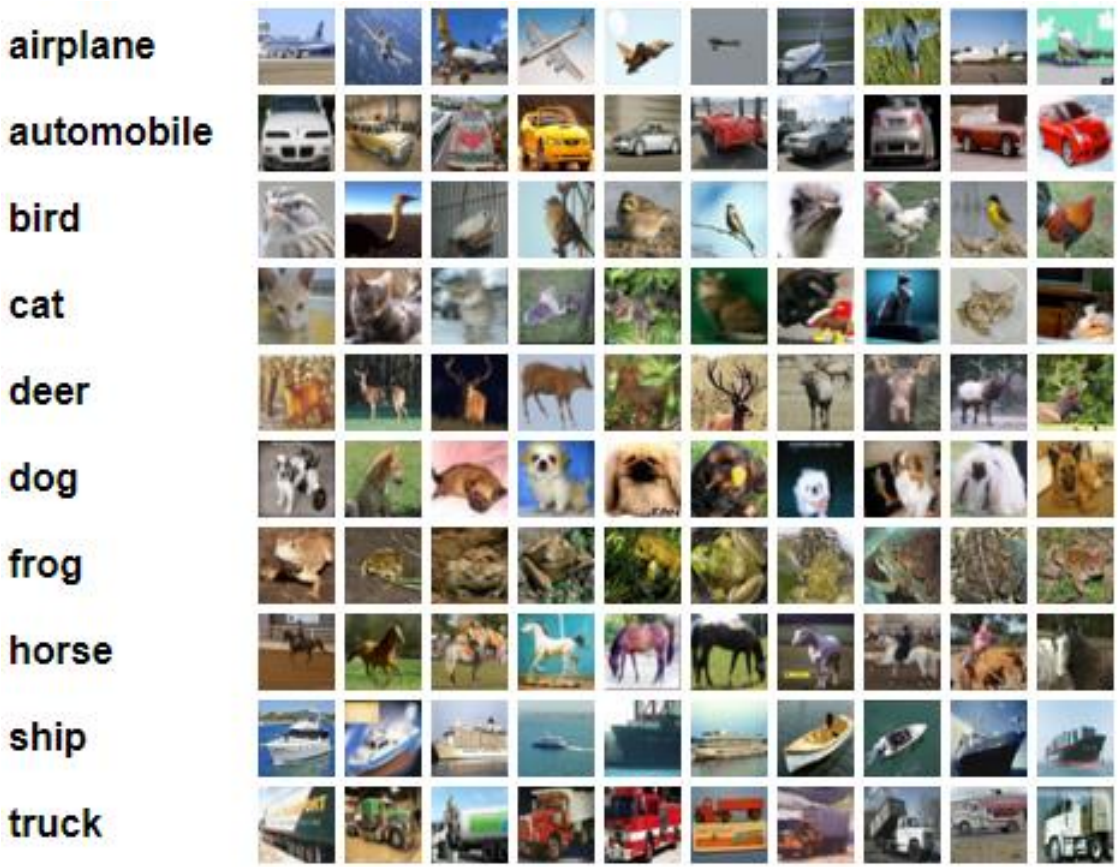
- The fully-connected layer takes the three-dimensional input by reducing it to one dimension and obtains a class label.
- Softmax layer calculates the probability distribution of the output classes.



Structure of the CNNs

Example

- **CIFAR-10*** dataset, has 60.000 32x32 color images of 10 classes (6.000 images for each class).
- It can be splitted into 50.000 for train and 10.000 for test.



*CIFAR-100 (Canadian Institute For Advanced Research) has 100 classes and 600.000 32x32 images.

Structure of the CNNs

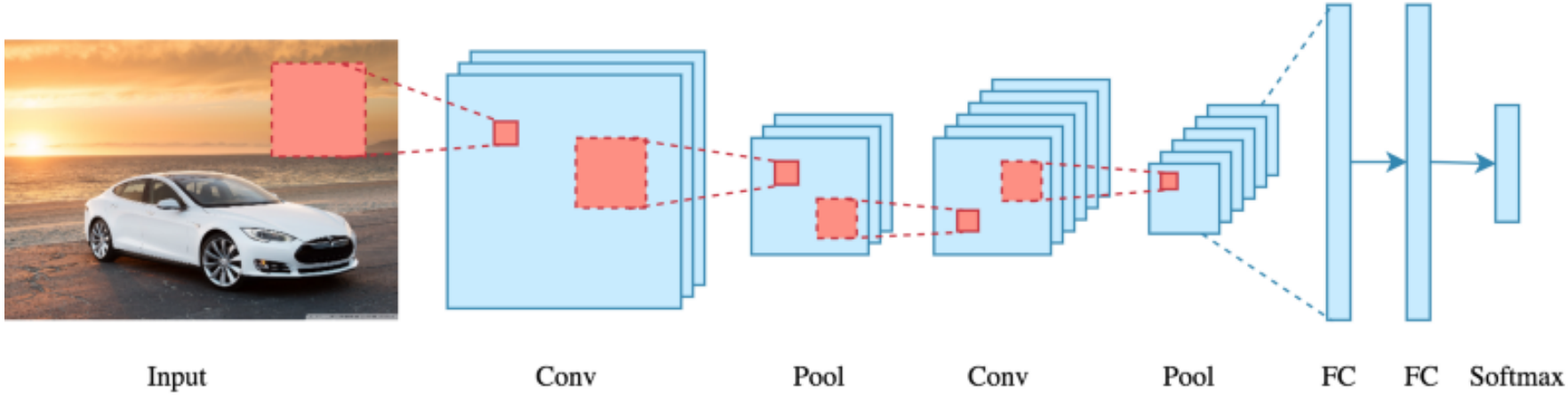
Example

- [Input-Conv-ReLU-Pool-FC] layers can be used for the CIFAR-10 dataset.
- The input layer takes 32x32x3 (red, green, blue) image pixels.
- The convolution layer calculates on the values it gets from the local regions of the input using the selected filter.
- If 12 different filters are used, the output of the convolution layer is 32x32x12 (RGB combined).
- The ReLU (Rectifier Linear Units) layer calculates the max (0, x) activation function result and produces a 32x32x12 output.

Structure of the CNNs

Example

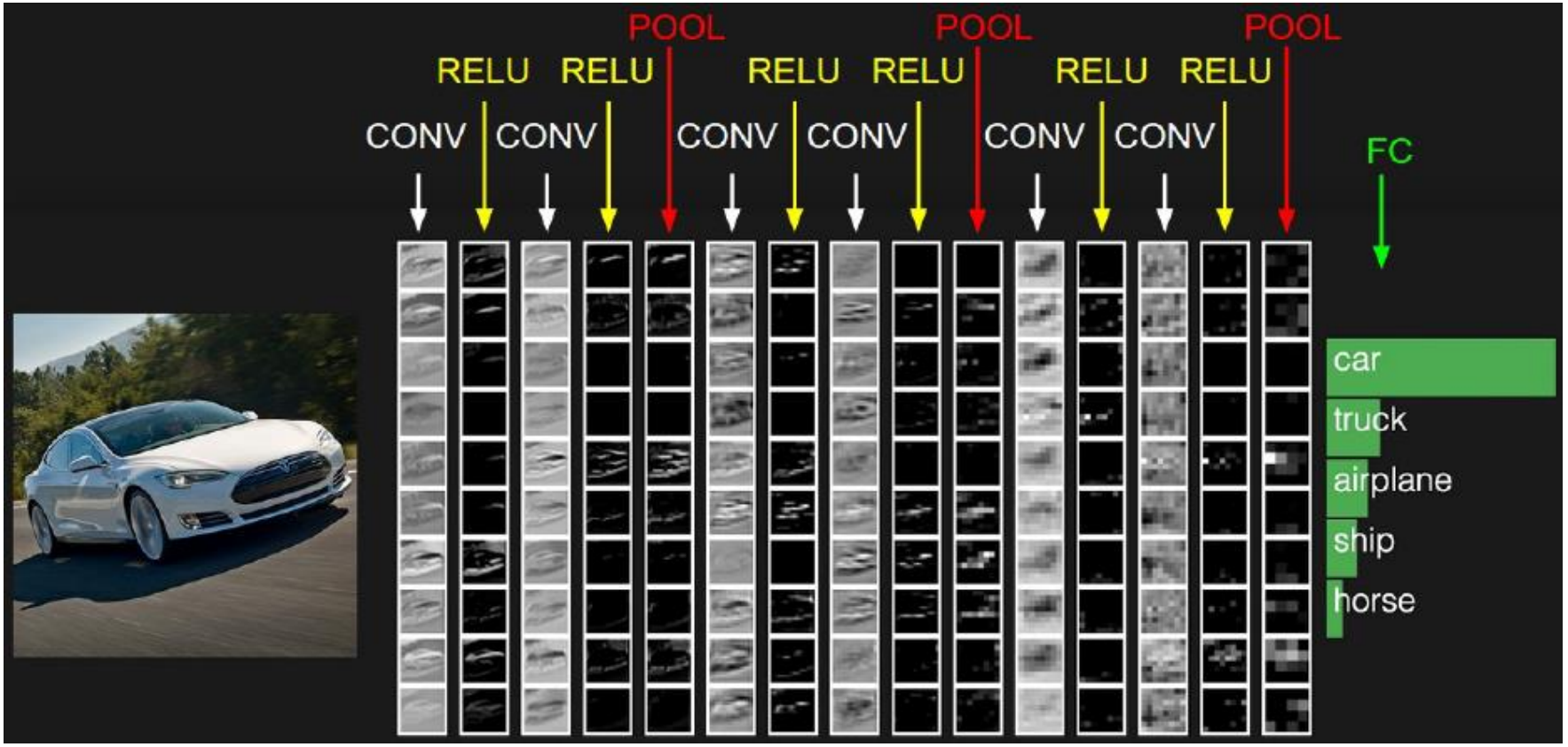
- The pool layer performs a downsampling operation and the output size can be, for example, 16x16x12.
- The fully connected layer calculates the value of the output class with 1x1x10.
- More successful results can be obtained by using different numbers of CONV + RELU + POOL layers consecutively depending on the problem type.



Structure of the CNNs

Example

- An example application for CIFAR-10 dataset can be found at http://cs231n.stanford.edu/



Content

- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

Convolution

- The main block in CNN is the convolution layer.
- Convolution is the mathematical operation that allows two sets to be combined.
- Convolution filter (kernel) is applied to the input to create a feature map.

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Input

1	0	1
0	1	0
1	0	1

Filter / Kernel

Convolution

- In the example, the input is 5x5 and the filter is 3x3.
- The convolution process is done by sliding the filter over the input matrix.
- The result of matrix multiplication with mutual elements creates one element of the feature map matrix.
- In the figure, convolution is done on 2D with a 3x3 filter.

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

Input x Filter

4		

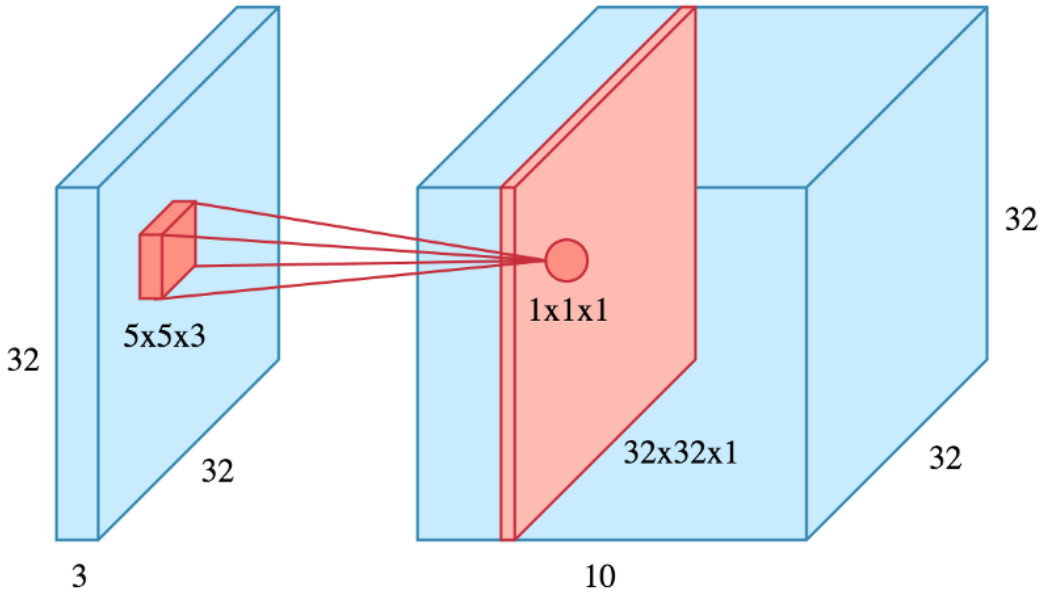
Feature Map

Convolution

- In real applications the image is shown in 3D (height, width and depth).
- Depth shows the color channels in the image.
- For RGB, the depth is taken as 3.
- Different convolution operations with different filters can be performed on one input.
- The output feature map of each filter is different.
- By combining all feature maps, a feature map is obtained as a result.

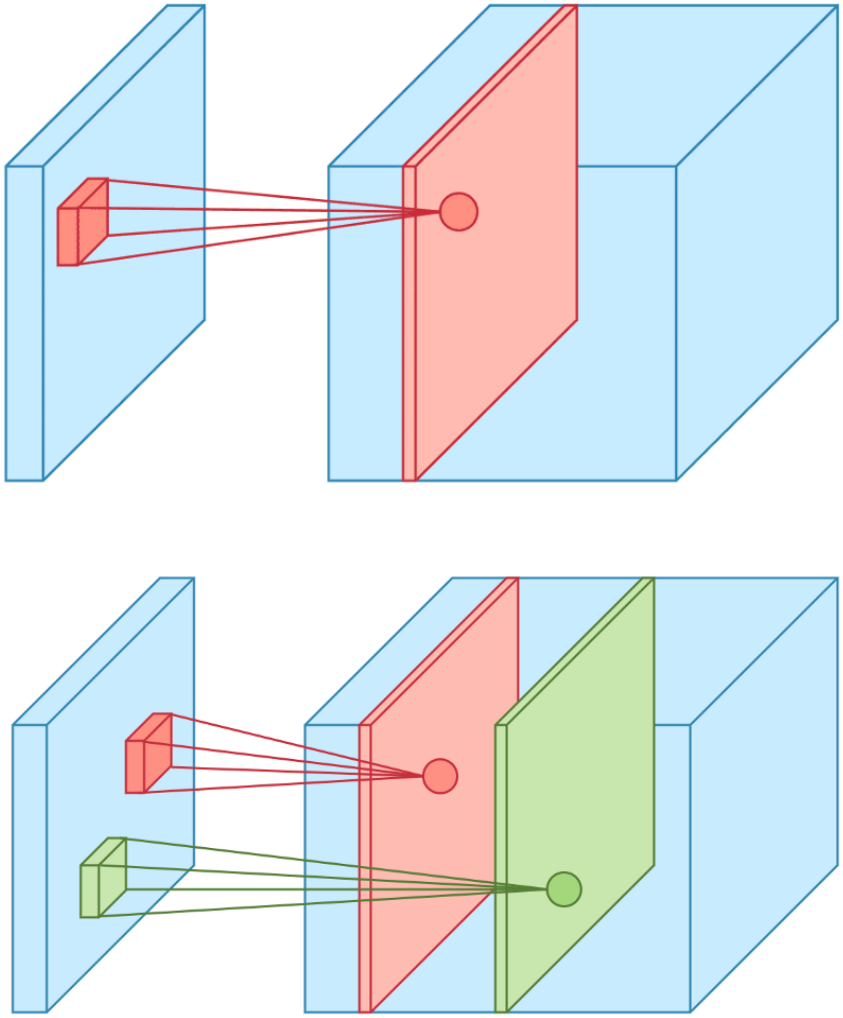
Convolution

- In the figure, a 32x32x3 image and a 5x5x3 filter are used.
- A 1x1x1 value is obtained by adding three 5x5x1 matrices.
- The feature map obtained is 32x32x1.
- If 10 different filters are used, the convolution layer consists of 32x32x10.



Convolution

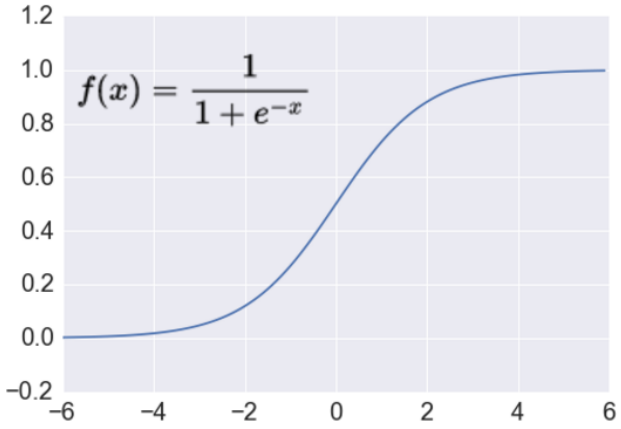
- The feature map is obtained by shifting the filter at the entire input matrix.



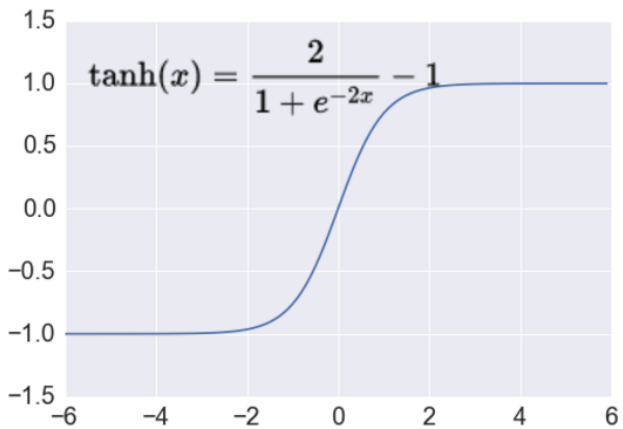
Convolution

- The result of the convolution operator is given as an input to the activation function.
- The activation function is chosen depending on the problem.

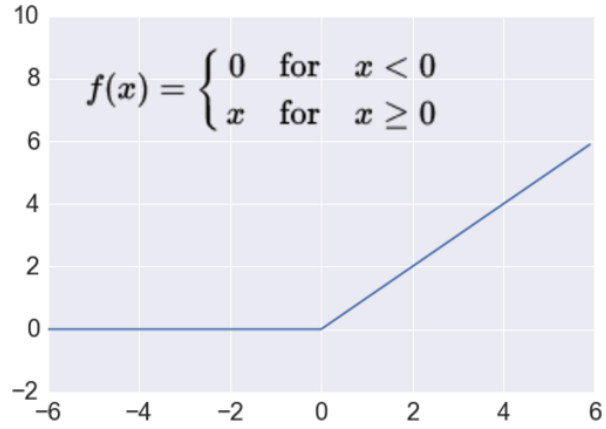
Sigmoid



TanH



ReLU

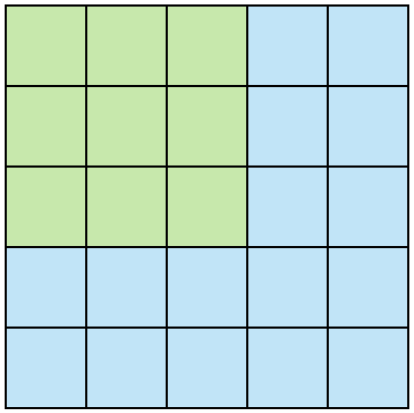


Content

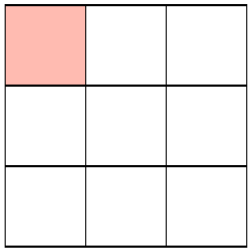
- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

Stride and padding

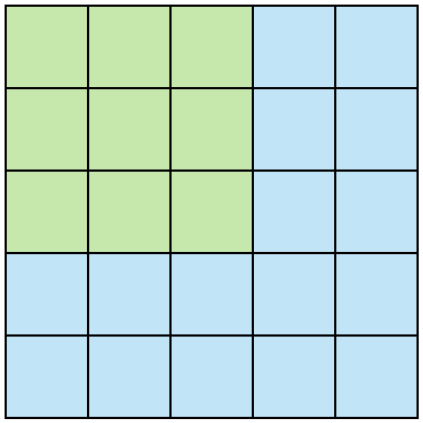
- Stride determines the movement size of the convolution filter at each step (default = 1).
- As the movement step size increases, the size of the feature map to be obtained becomes smaller.



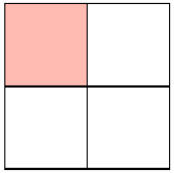
Stride 1



Feature Map



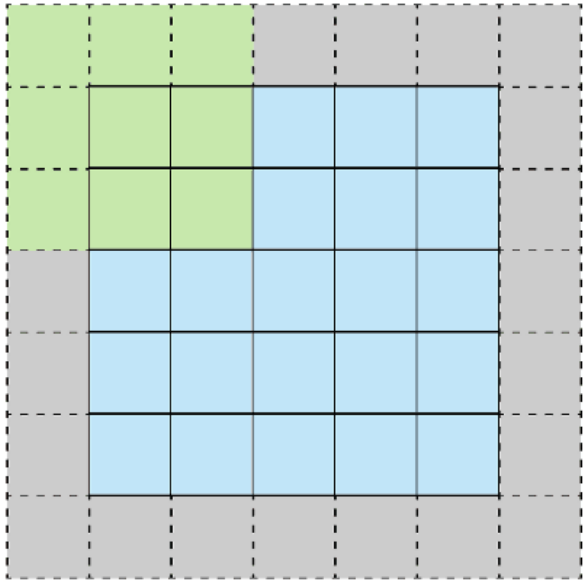
Stride 2



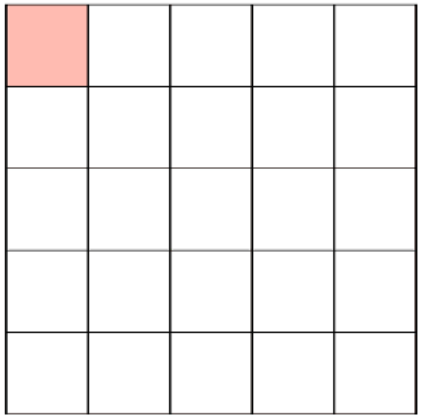
Feature Map

Stride and padding

- Padding is used to create the same size feature map as the input.
- Cells with a value of 0 around the input matrix are added as padding.



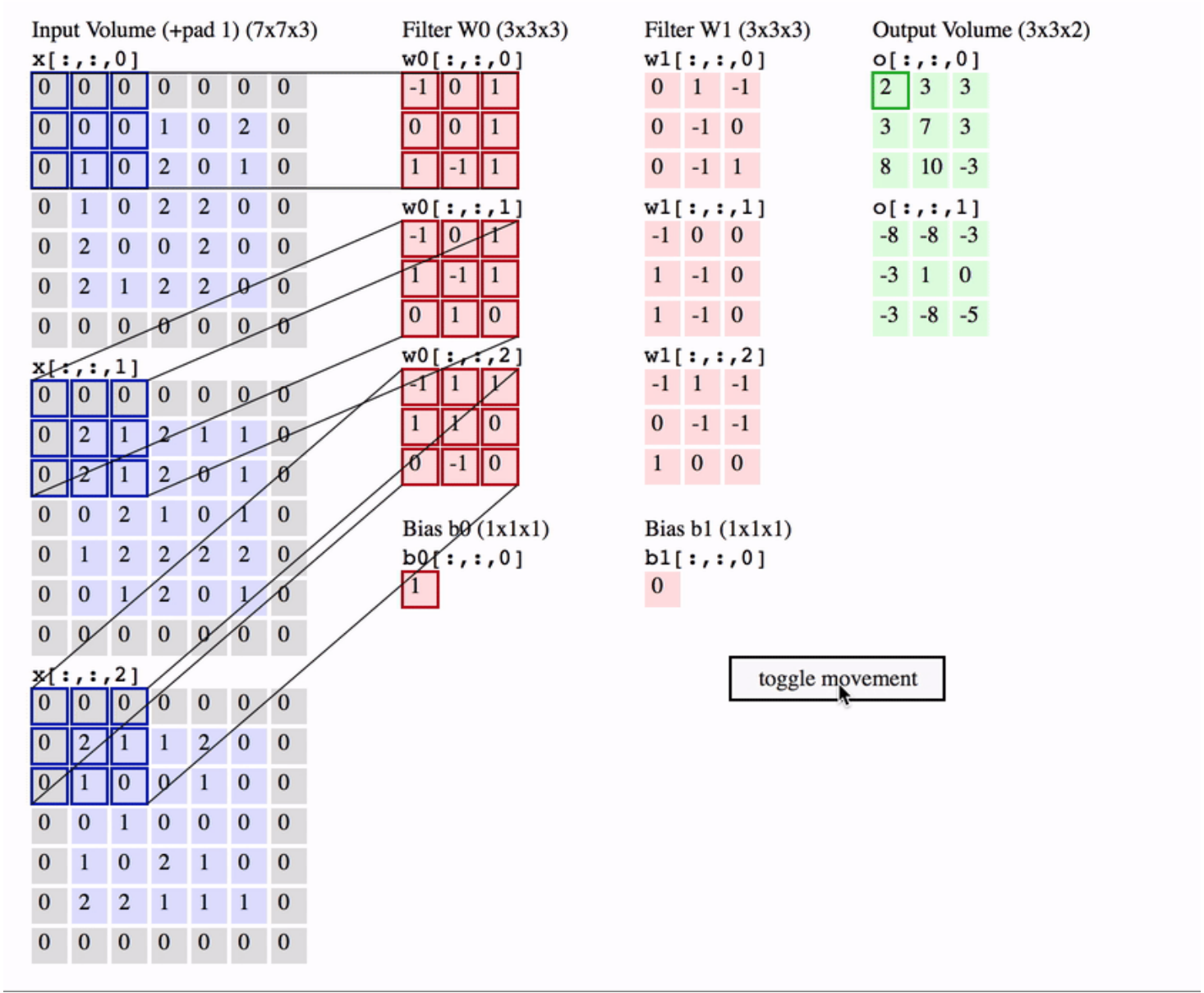
Stride 1 with Padding



Feature Map

Stride and padding

- **Example:** Inputs = 5x5x3, Padding= 1, Stride= 2

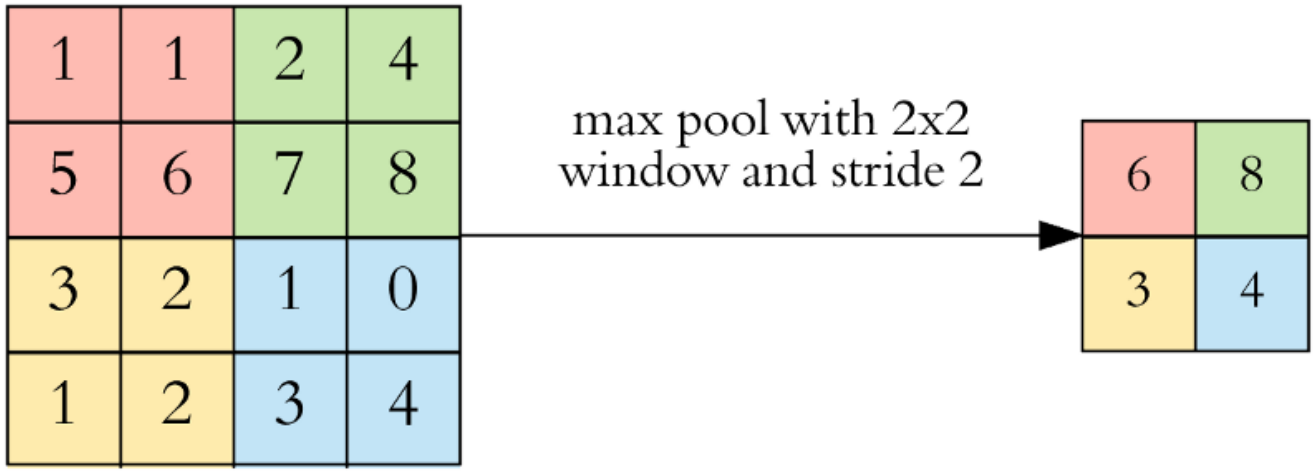


Content

- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

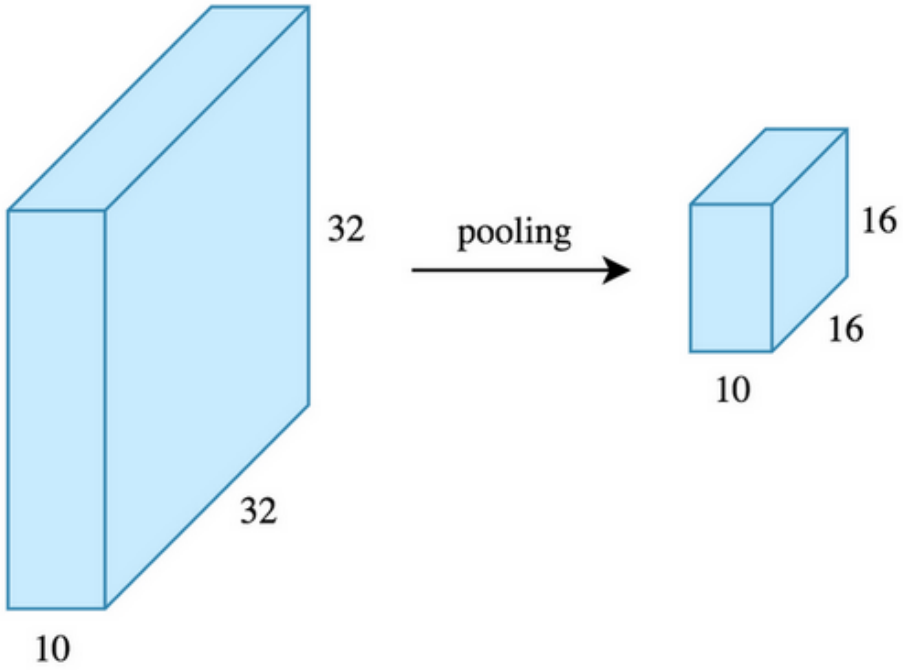
Pooling

- Pooling is applied after the convolution process and performs dimension reduction.
- The pooling layer samples by reducing the height and width of the feature map (the depth remains the same).
- Max pooling is the most widely used method.
- Window size and stride values are specified depending on the problem.



Pooling

- Typically, the values for window size and stride are chosen so that half of the feature map in the input is obtained.
- After pooling, the size of the feature map is reduced in half.

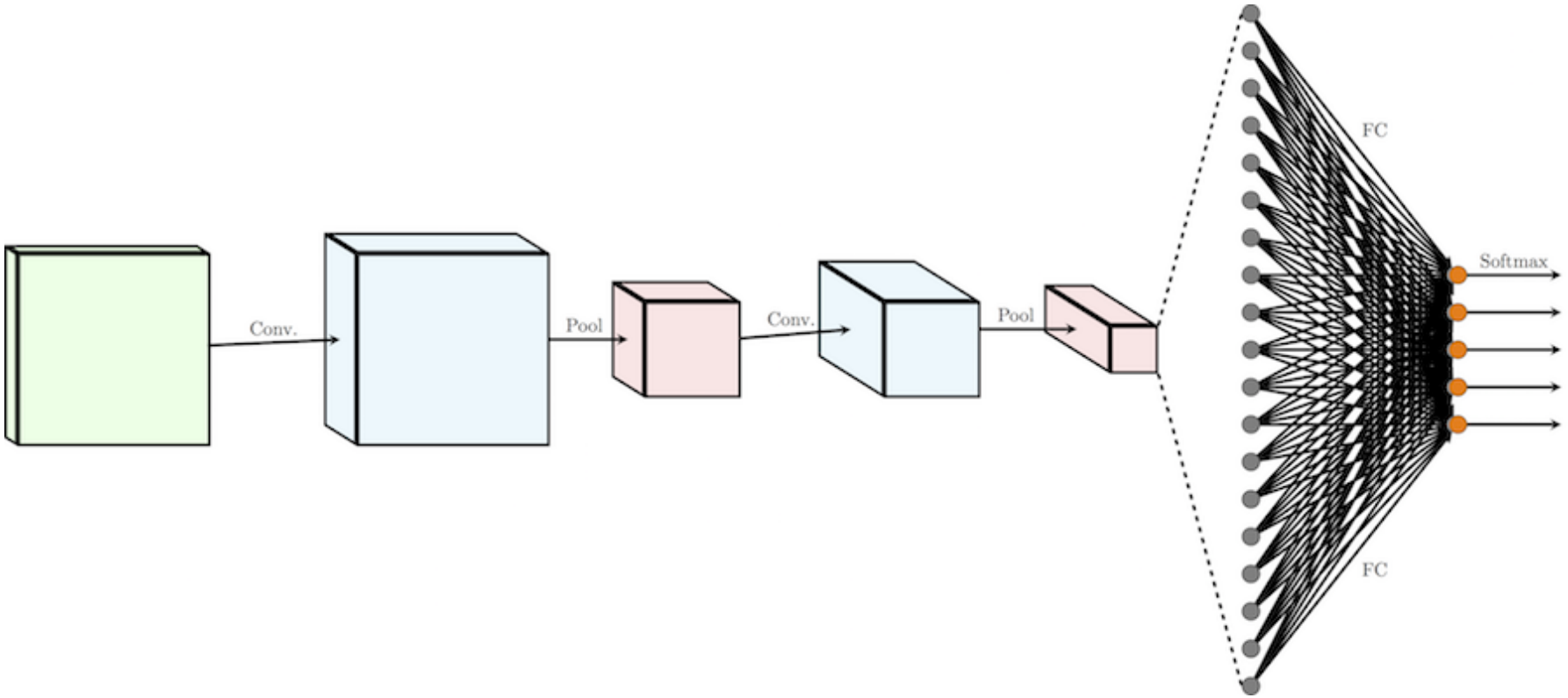


Content

- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

Fully connected layer

- After the pooling layer, a fully connected ANN is placed.
- Pooling layer output is taken in 3D and reduced to 1D at the fully connected ANN
- ANN obtaines a 1D output vector which is size equals to number of classes.

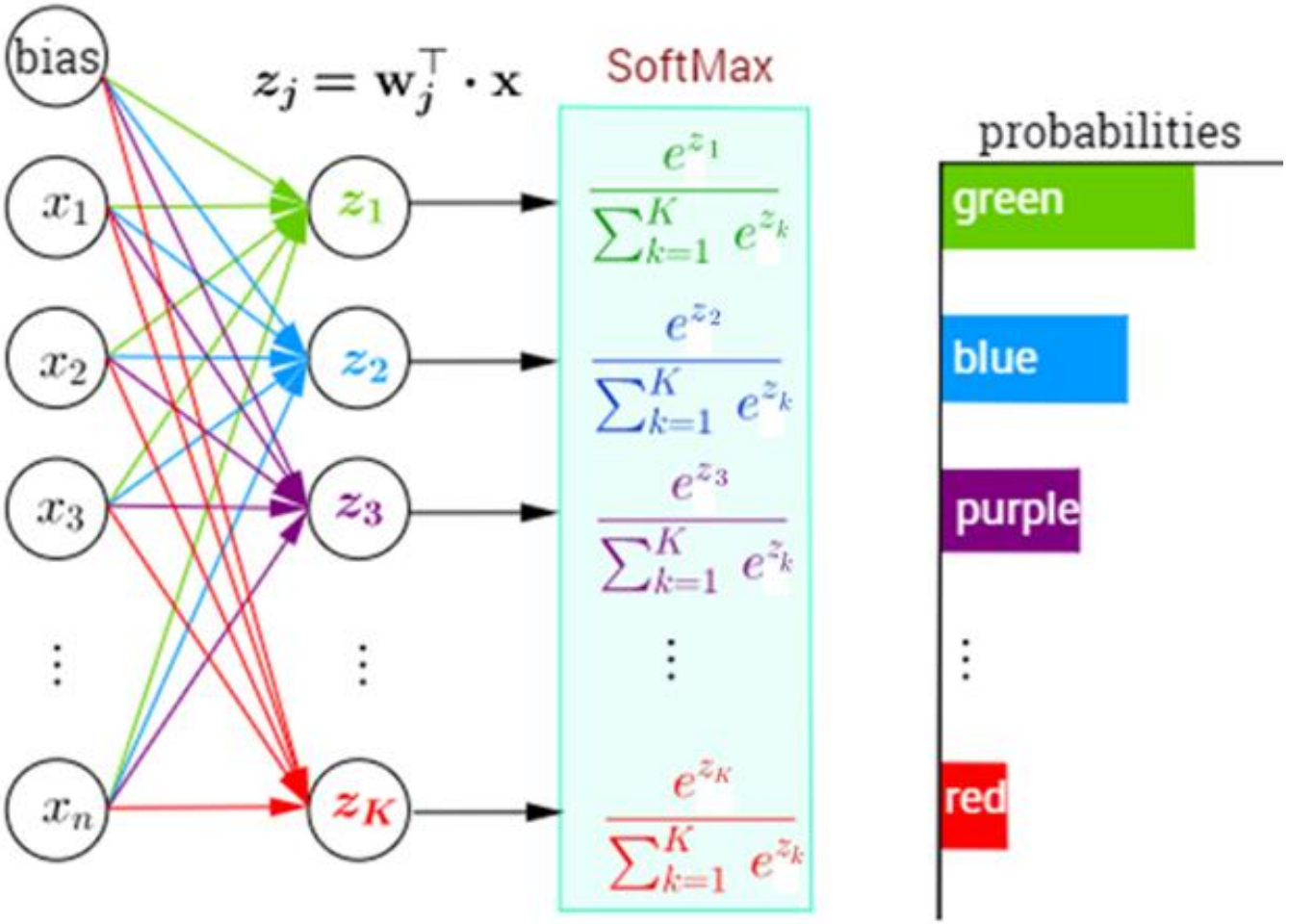


Content

- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

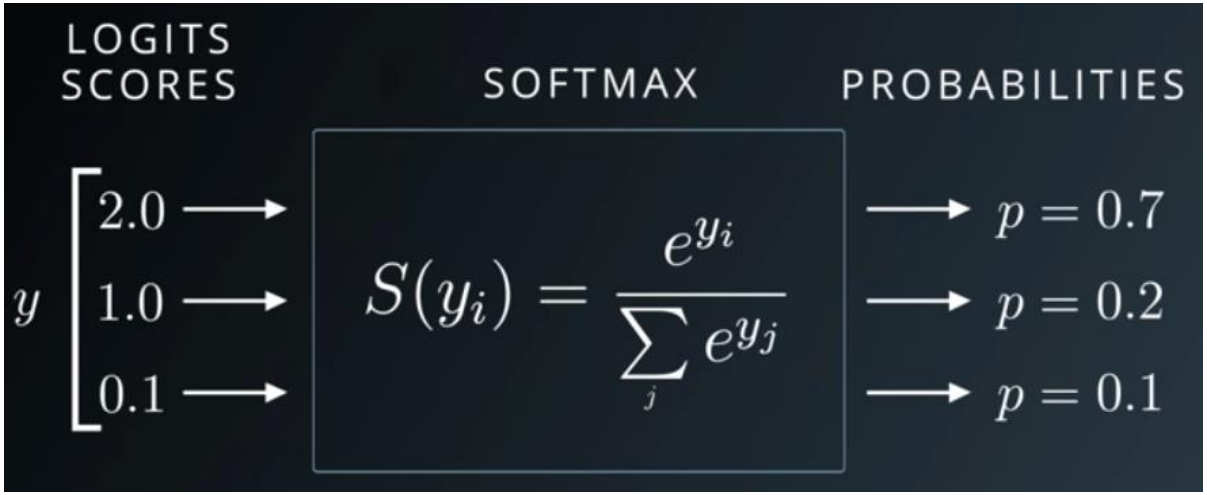
Softmax

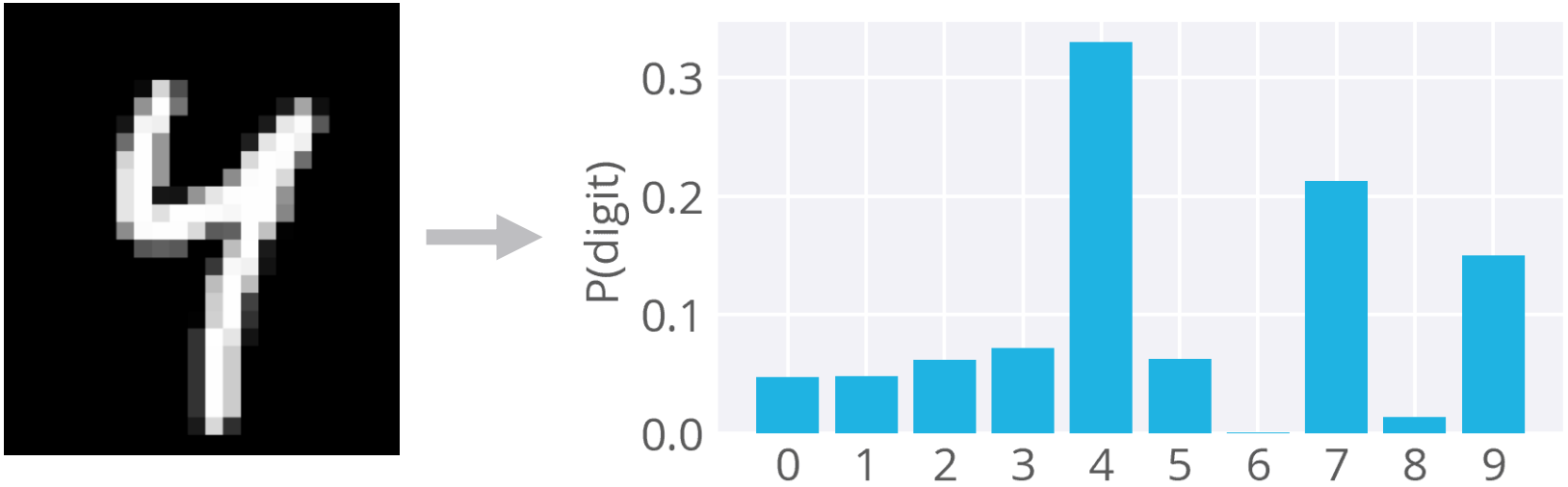
- Softmax function is used in classification problems.
- The softmax layer calculates the probability distribution of the output classes.



Softmax

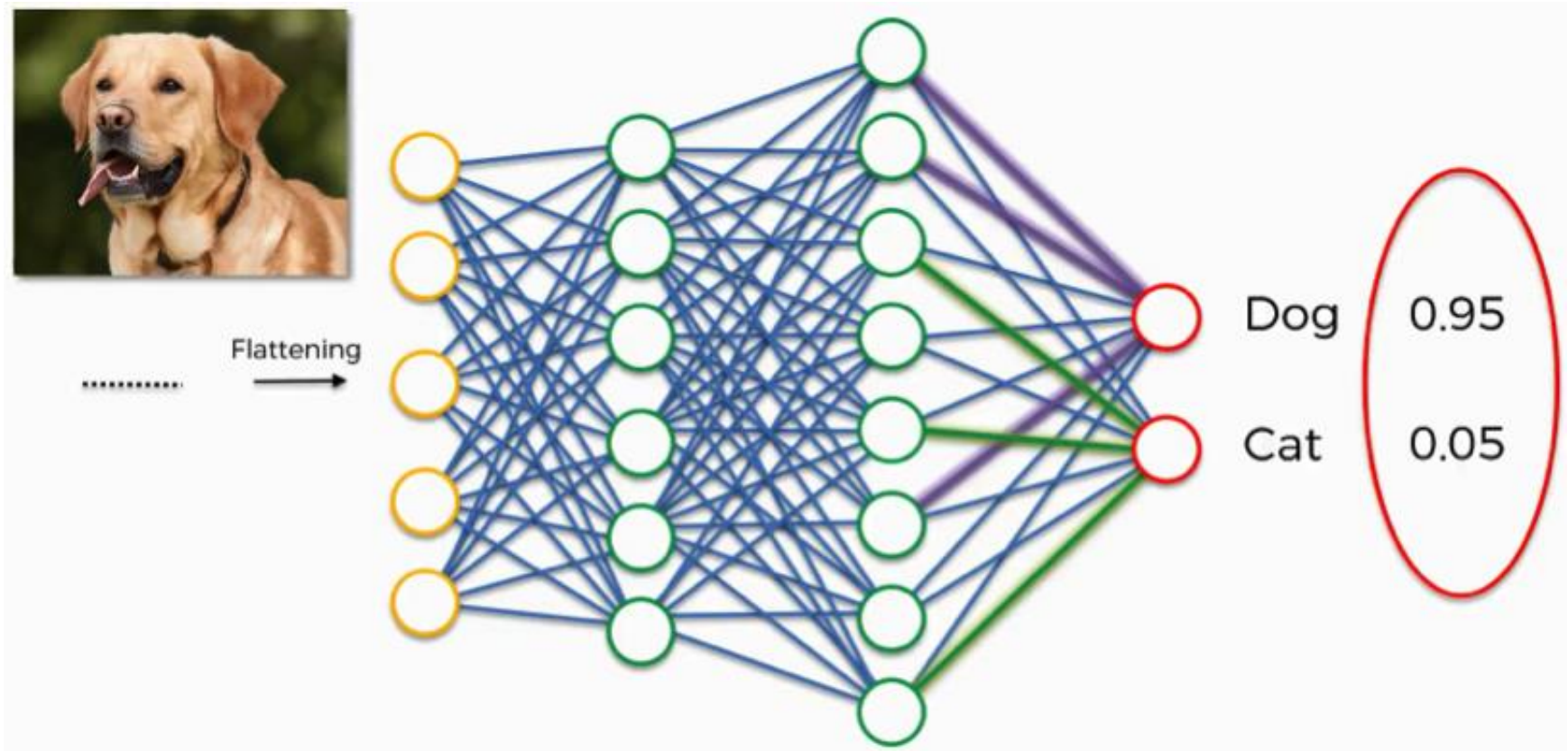
- Softmax gives the distribution of the probability that the output belongs to classes.





Softmax

- Usually, the number of the output neurons is taken as the number of class labels.
- The output label that has high probability is assigned for given input images.



Content

- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

Hyperparameters

- Hyper parameters are not learned directly, but determine the properties of the model.
- The following hyper parameters are used in CNN:
  - **Filter size:** Usually 3x3 is used, but may be larger depending on the problem.
  - **Number of filters:** The more filters are used, the more powerful the model is obtained. However, a large number of parameters increase the risk of overfitting.
  - **Stride:** Usually 1 is chosen for stride, but a different value can be chosen depending on the problem.
  - **Padding:** Usually taken as padding 1, but may not be used depending on the problem.

Content

- Convolutional neural networks
- Structure of the CNNs
- Convolution
- Stride and padding
- Pooling
- Fully connected layer
- Softmax
- Hyperparameters
- Applications

Applications

- CNN is a successfully applied model for **image related problems.**
- CNN has been successfully implemented in recommendation systems, NLP and many other areas.
- CNN automatically **detects important features** in the input data.
- CNN model **can classify images** better and faster than human.
- CNN model **can identify objects** very fast and with high accuracy.

Applications

Image Classification

- Image classification involves assigning a label to an entire image or photograph.
- This problem is also referred to as "object classification" or "image recognition".
- Some examples of image classification include:
  - Labeling an x-ray as cancer or not (binary classification).
  - Classifying a handwritten digit (multiclass classification).
  - Assigning a name to a photograph of a face (multiclass classification).

Applications

Image Classification

- A popular example of image classification used as a benchmark problem is the MNIST dataset.

Applications

Image Classification

- A popular real-world version of classifying photos of digits is The Street View House Numbers dataset.

Applications

Image Classification

- There are many image classification tasks that involve photographs of objects.
- Two popular examples include the CIFAR-10 and CIFAR-100 datasets.
- The Large Scale Visual Recognition Challenge is an annual competition in which teams compete for the best performance using ImageNet database.
- There have been significant achievements in image recognition/classification applications.

Applications

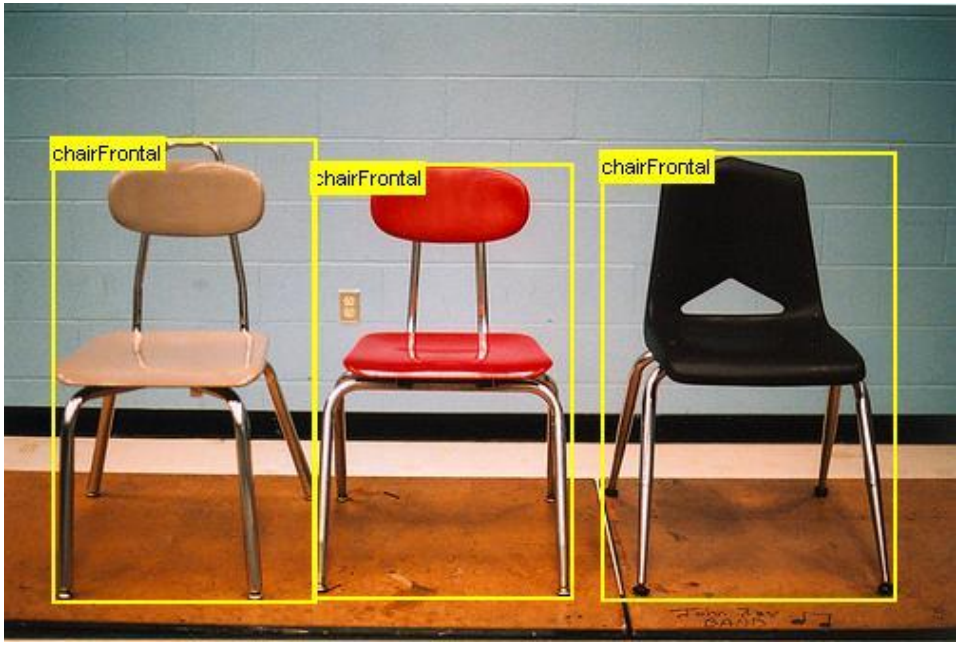
Image Classification With Localization

- Image classification with localization involves assigning a class label and showing the location of the object by a bounding box.
- This is a more challenging version of image classification.
- Some examples of image classification with localization include:
- Labeling an x-ray as cancer or not and drawing a box around the cancerous region.
- Classifying photographs of animals and drawing a box around the animal in each scene.
- A classical dataset for image classification with localization is the PASCAL Visual Object Classes dataset.

Applications

Image Classification With Localization

- This task may sometimes be referred to as "object detection."
- The ILSVRC2016 Dataset for image classification with localization is comprised of 150,000 photographs with 1,000 categories of objects.



Applications

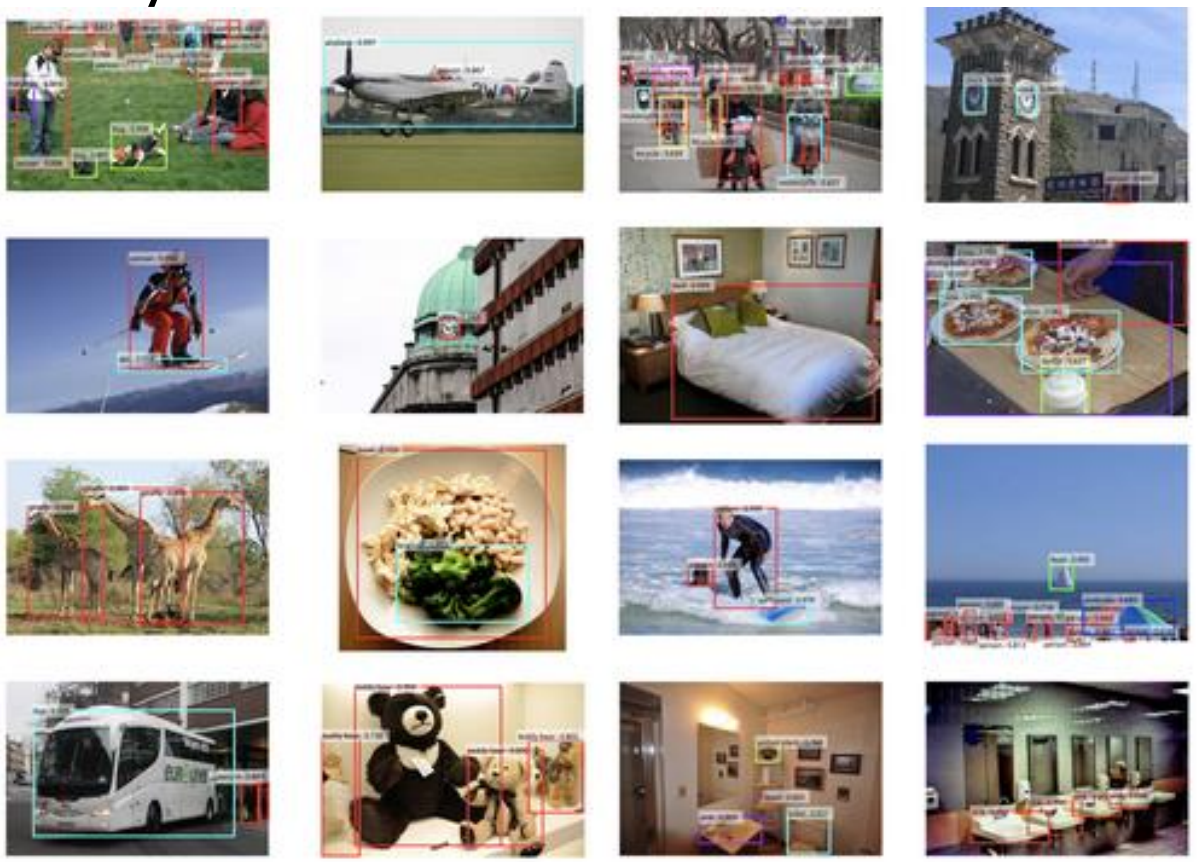
Object Detection

- Object detection is the task of image classification with localization.
- This is a more challenging task than simple image classification or image classification with localization.
- Often, techniques developed for image classification with localization are used and demonstrated for object detection.
- Some examples of object detection include:
  - Drawing a bounding box and labeling each object in a street scene.
  - Drawing a bounding box and labeling each object in an indoor photograph.
  - Drawing a bounding box and labeling each object in a landscape.

Applications

Object Detection

- The PASCAL Visual Object Classes dataset is a common dataset for object detection.
- Another dataset is Microsoft's Common Objects in Context Dataset, namely COCO.



Applications

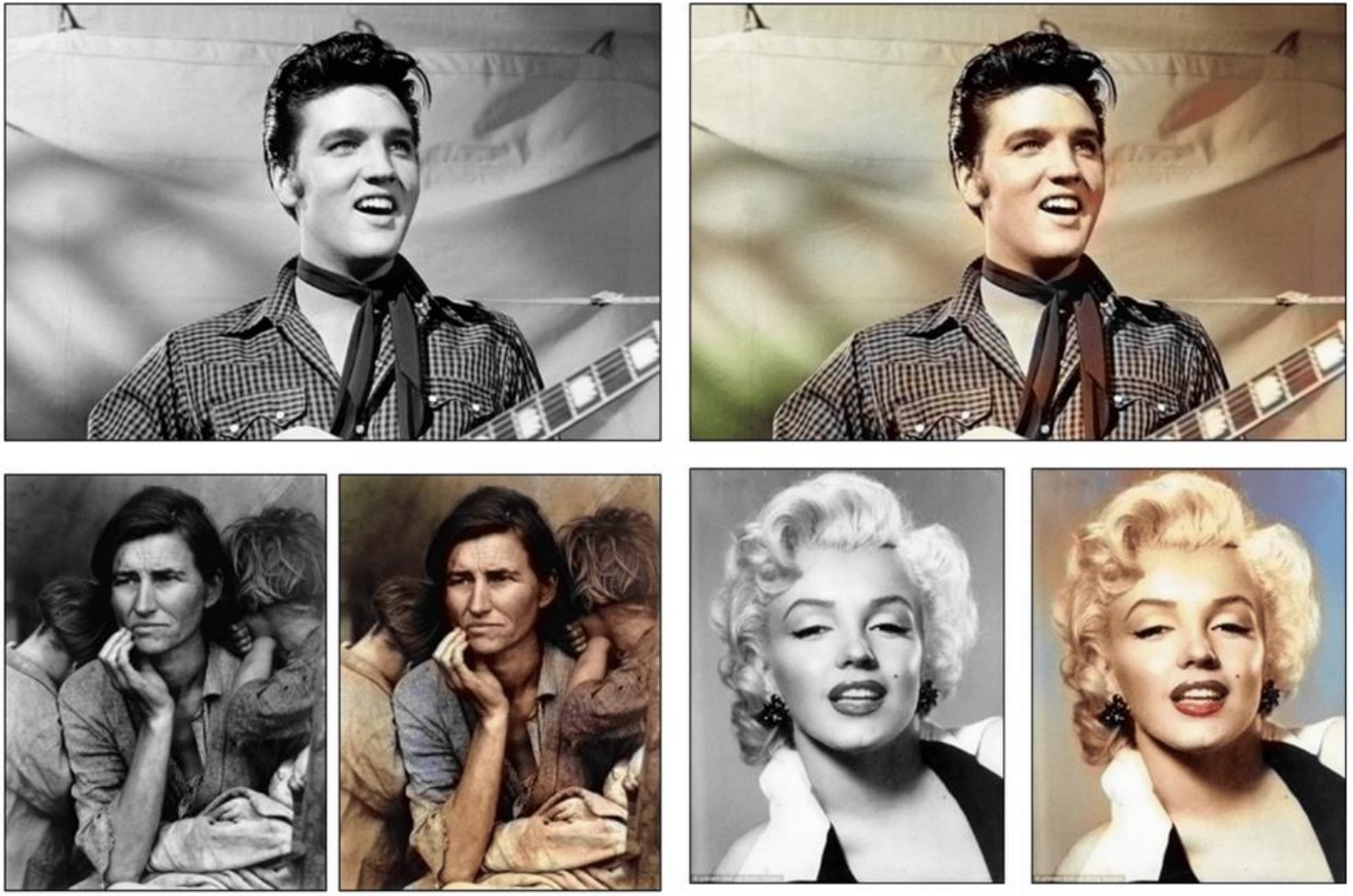
Image Colorization

- Image colorization involves converting a grayscale image to a full color image.
- This task can be thought of as a type of photo filter or transform that may not have an objective evaluation.
- Examples include colorizing old black and white photographs and movies.
- Datasets often involve using existing photo datasets and creating grayscale versions of photos.

Applications

Image Colorization

- Image colorization especially is used for historical or grayscale old version of the photos.



Applications

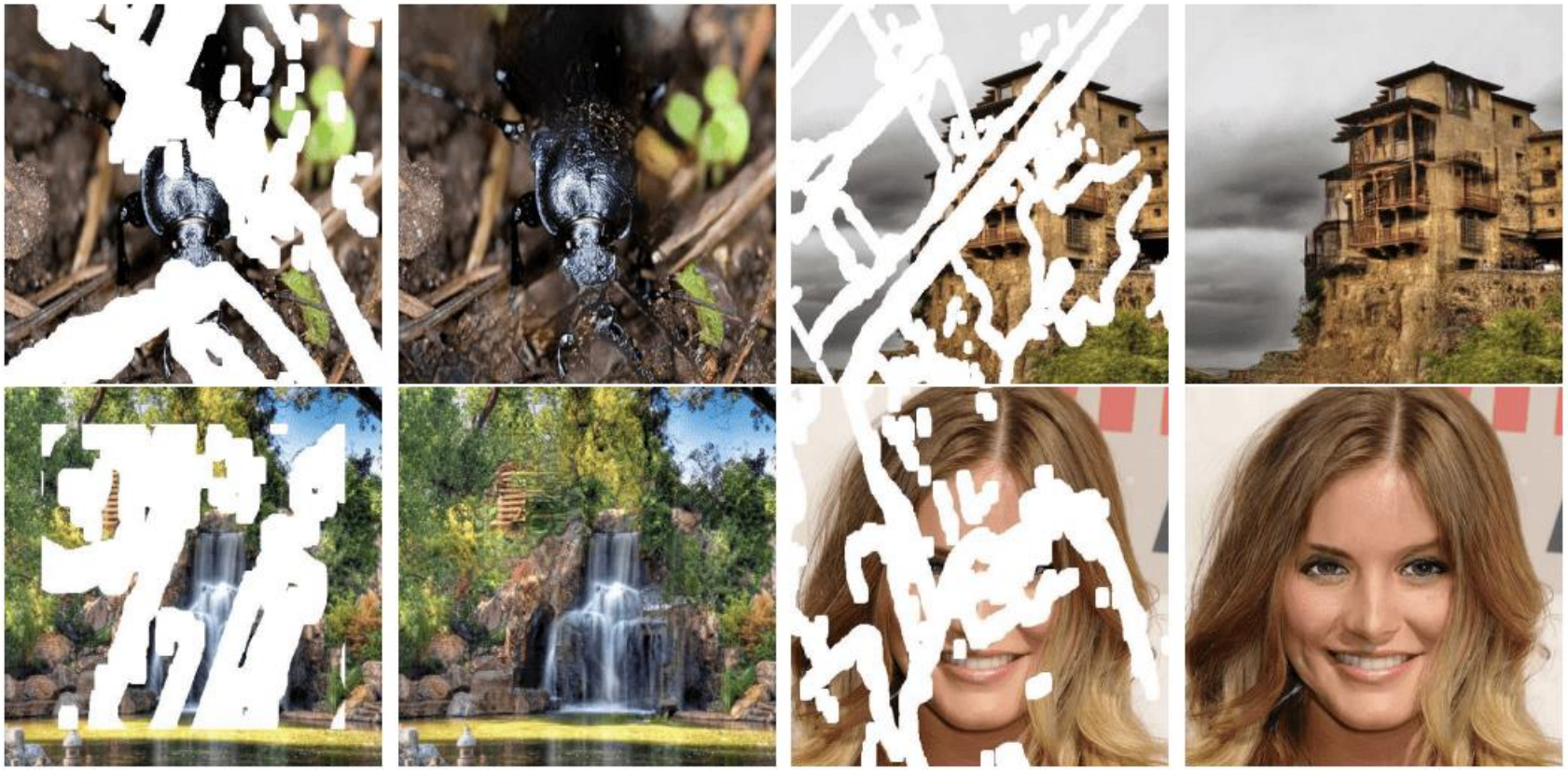
Image Reconstruction

- Image reconstruction is the task of filling in missing or corrupt parts of an image.
- This task can be thought of as a type of photo filter or transform that may not have an objective evaluation.
- Examples include reconstructing old, damaged black and white photographs and movies.
- Datasets often involve using existing photo datasets and creating corrupted versions of photos.
- The models must learn to repair using original photos and corrupted versions of the photos.

Applications

Image Reconstruction

- Image reconstruction and image inpainting is the task of filling in missing or corrupt parts of an image.



Applications

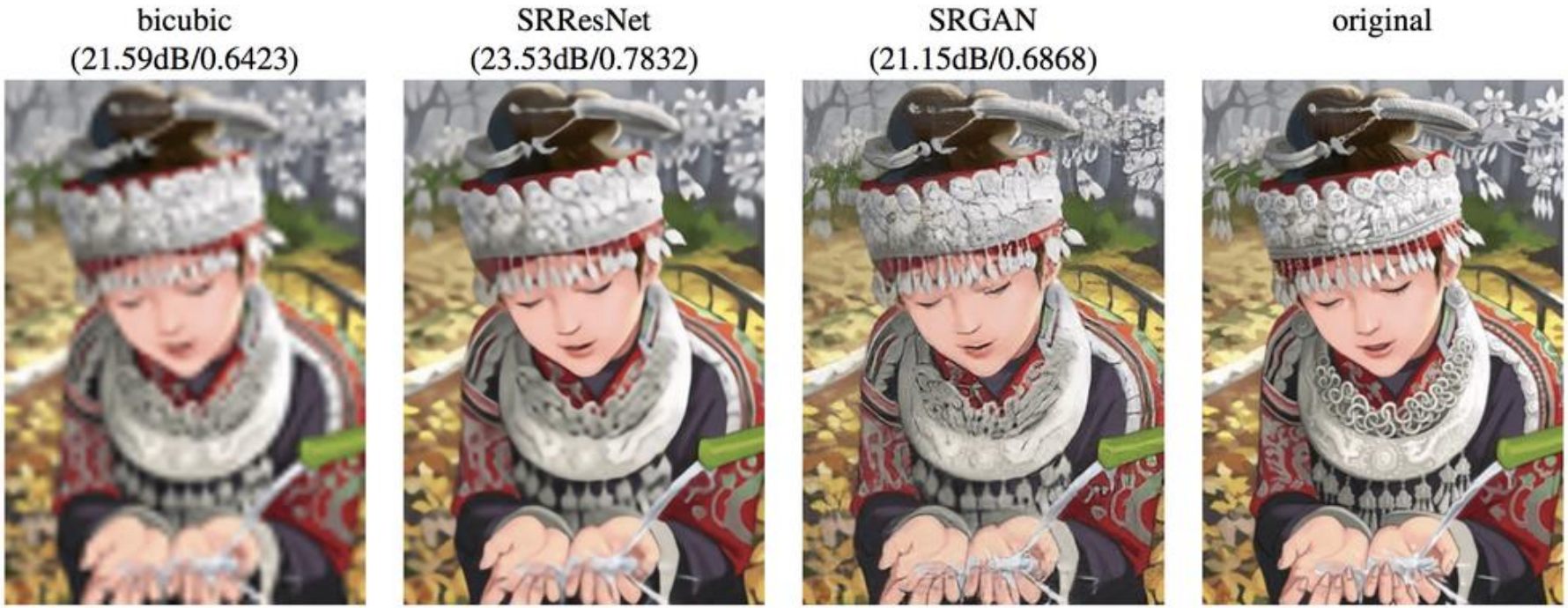
Image Super-Resolution

- Image super-resolution is the task of generating a new version of an image with a higher resolution and detail than the original image.
- Often models developed for image restoration and inpainting can be used for image super-resolution.
- Datasets often involve using existing photo and creating down-scaled version.
- The CNN models must learn to create super-resolution versions using training data set.

Applications

Image Super-Resolution

- Image super-resolution can generate a new higher resolution version using the input than the original image.



Applications

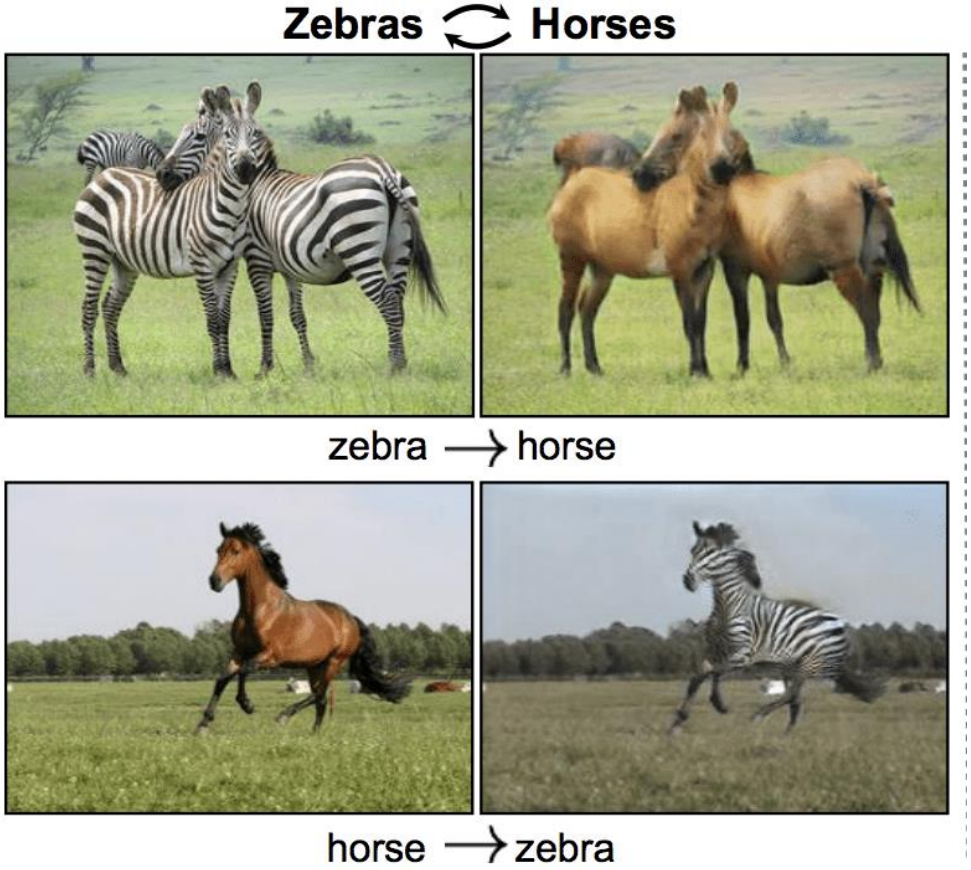
Image Synthesis

- Image synthesis is the task of generating targeted modifications of existing images or entirely new images.
- This is a very broad area that is rapidly advancing.
- It may include small modifications of image and video (e.g. image-to-image translations), such as:
  - Changing the style of an object in a scene.
  - Adding an object to a scene.
  - Adding a face to a scene.

Applications

Image Synthesis

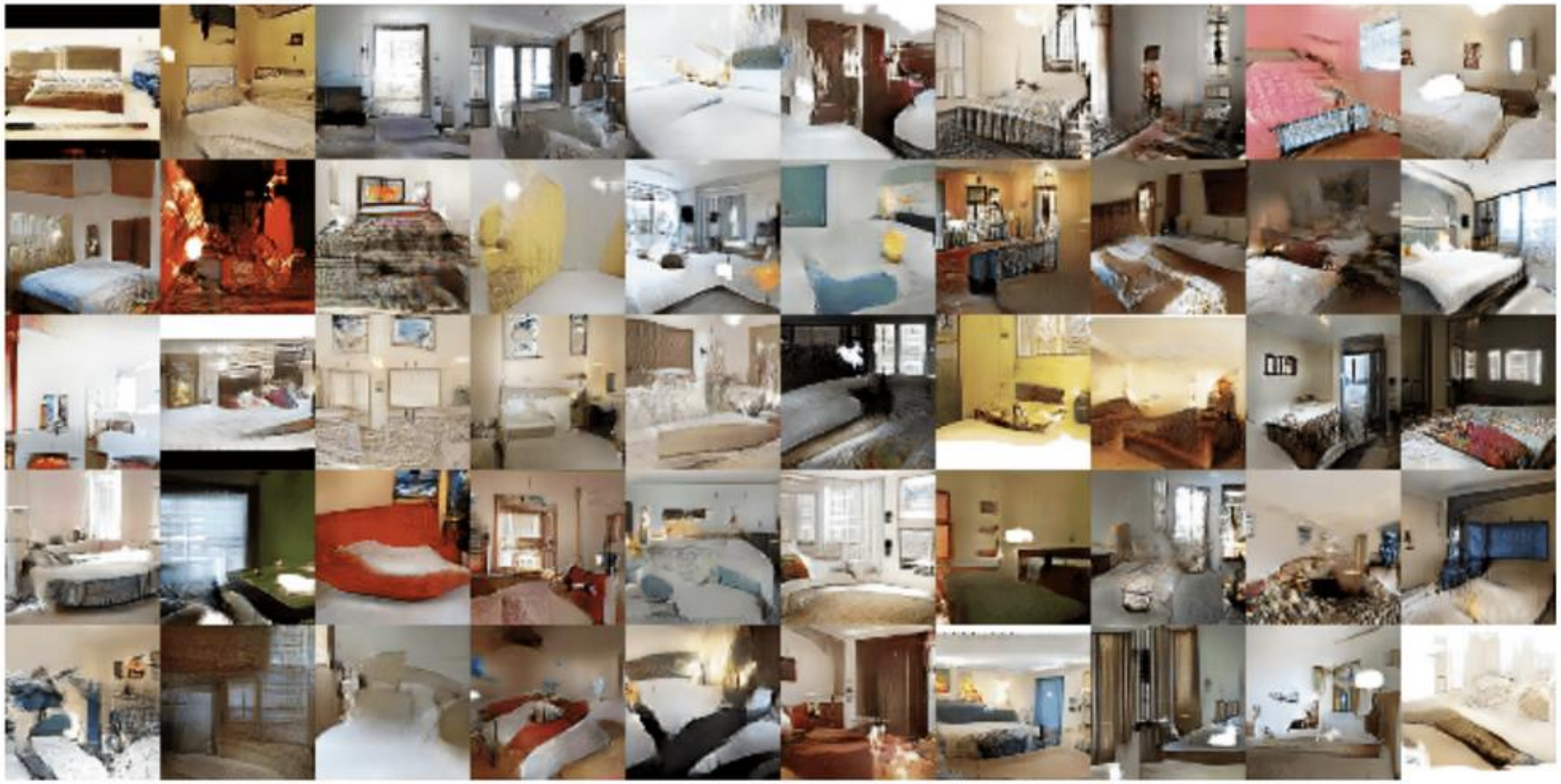
- An image with a zebra image in the figure has been modified to include a horse image.
- The patterns and colors in the image of the horse are transferred to the zebras.



Applications

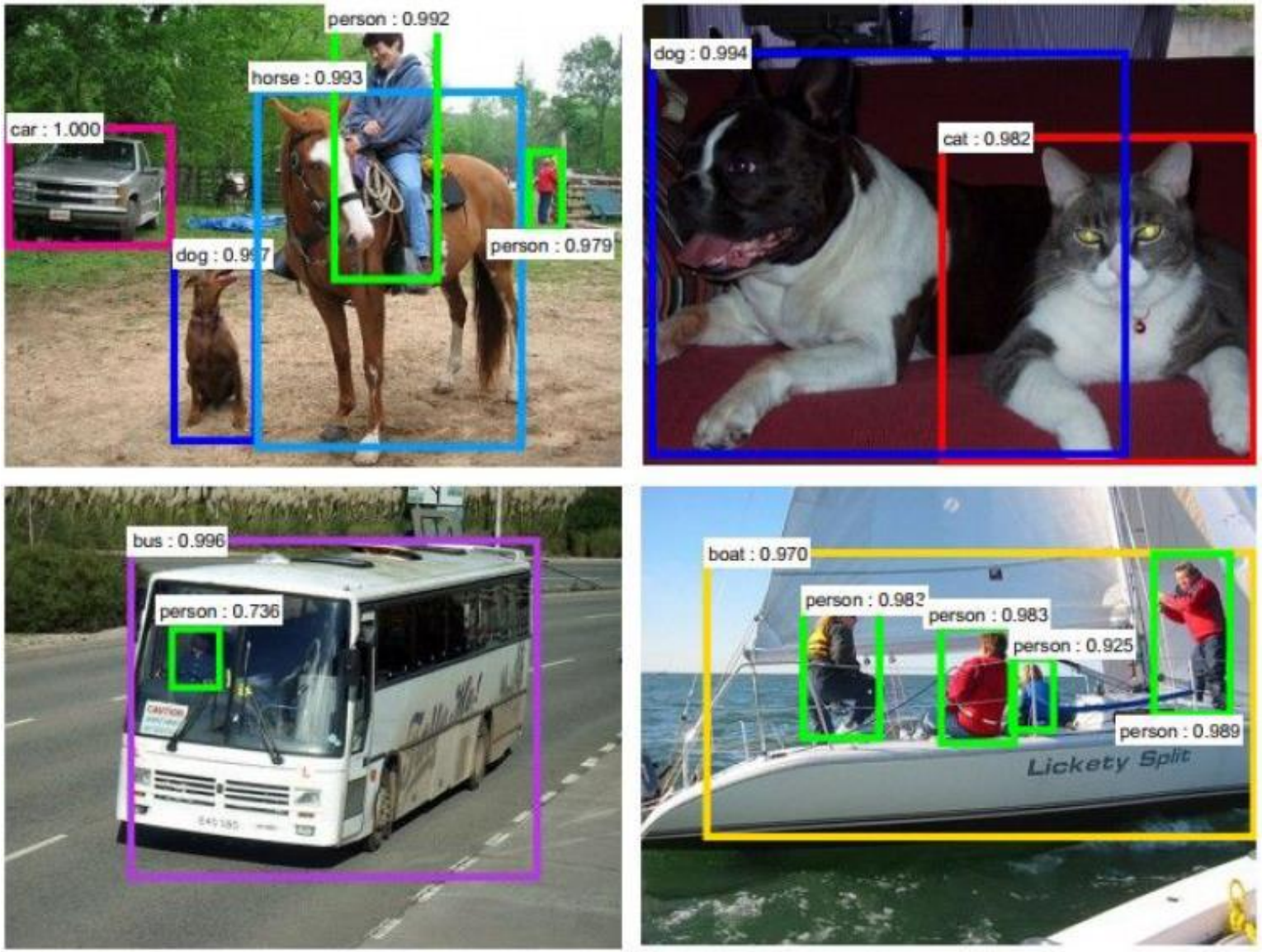
Image Synthesis

- It may also include generating entirely new images, such as:
  - Generating faces.
  - Generating bathrooms.
  - Generating clothes.



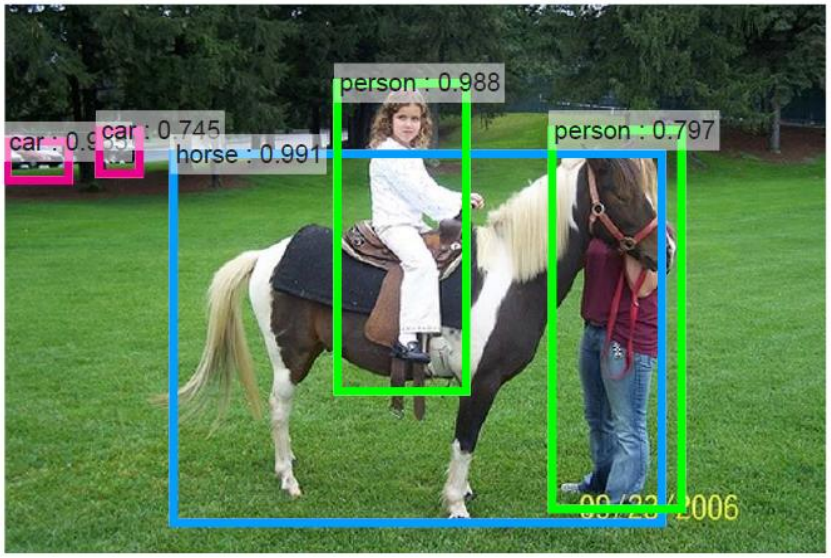
Applications

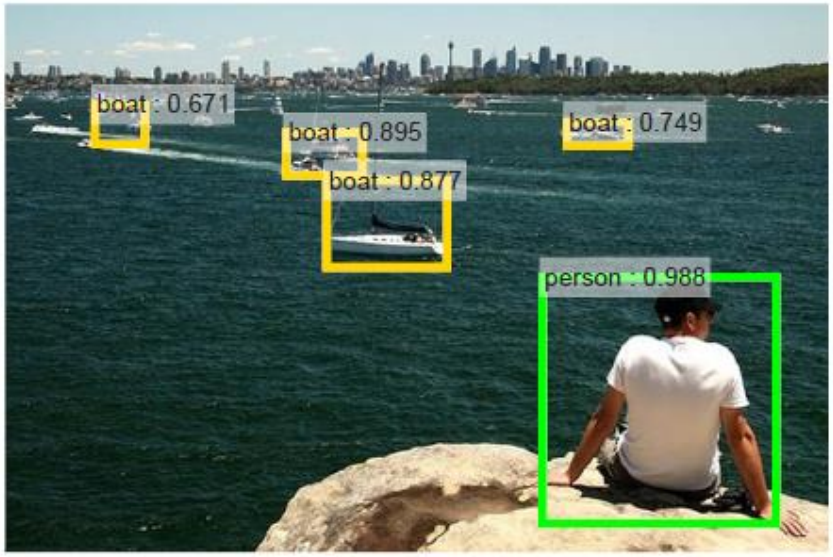
- Multiple objects recognition

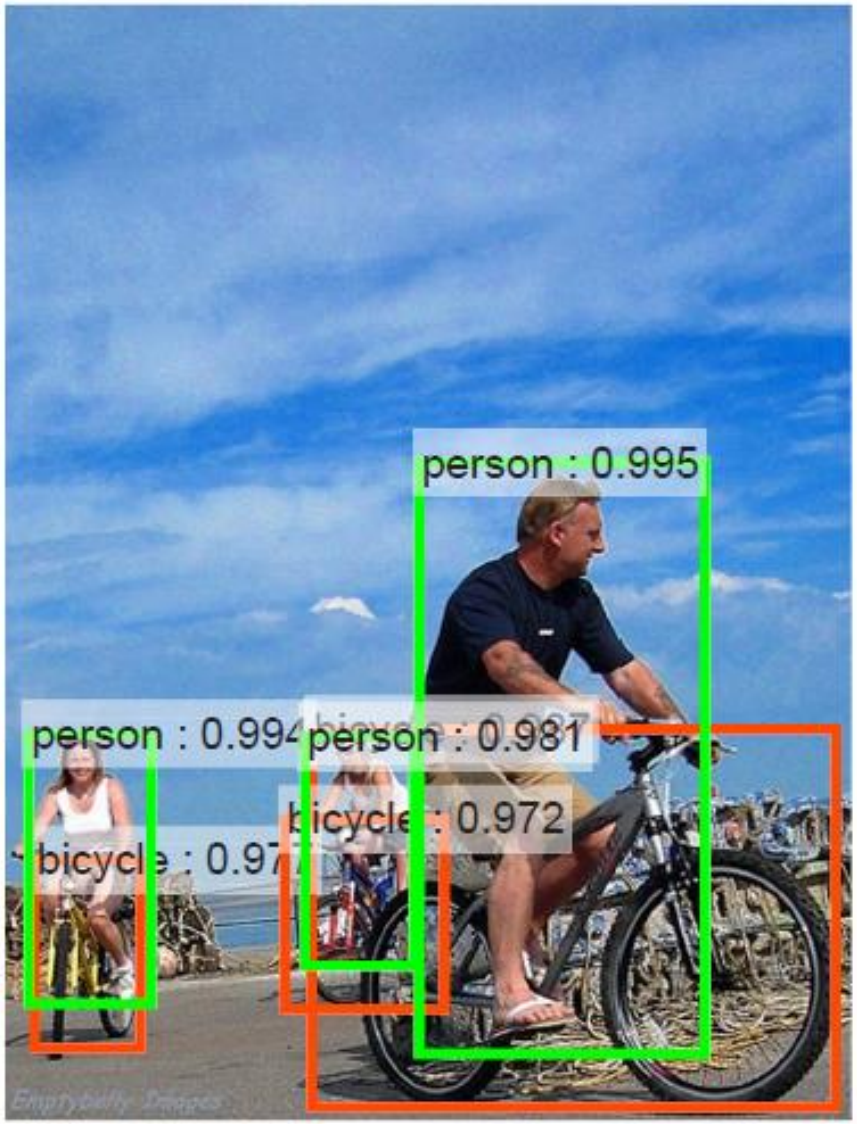


Applications

- Overlapped multiple objects recognition

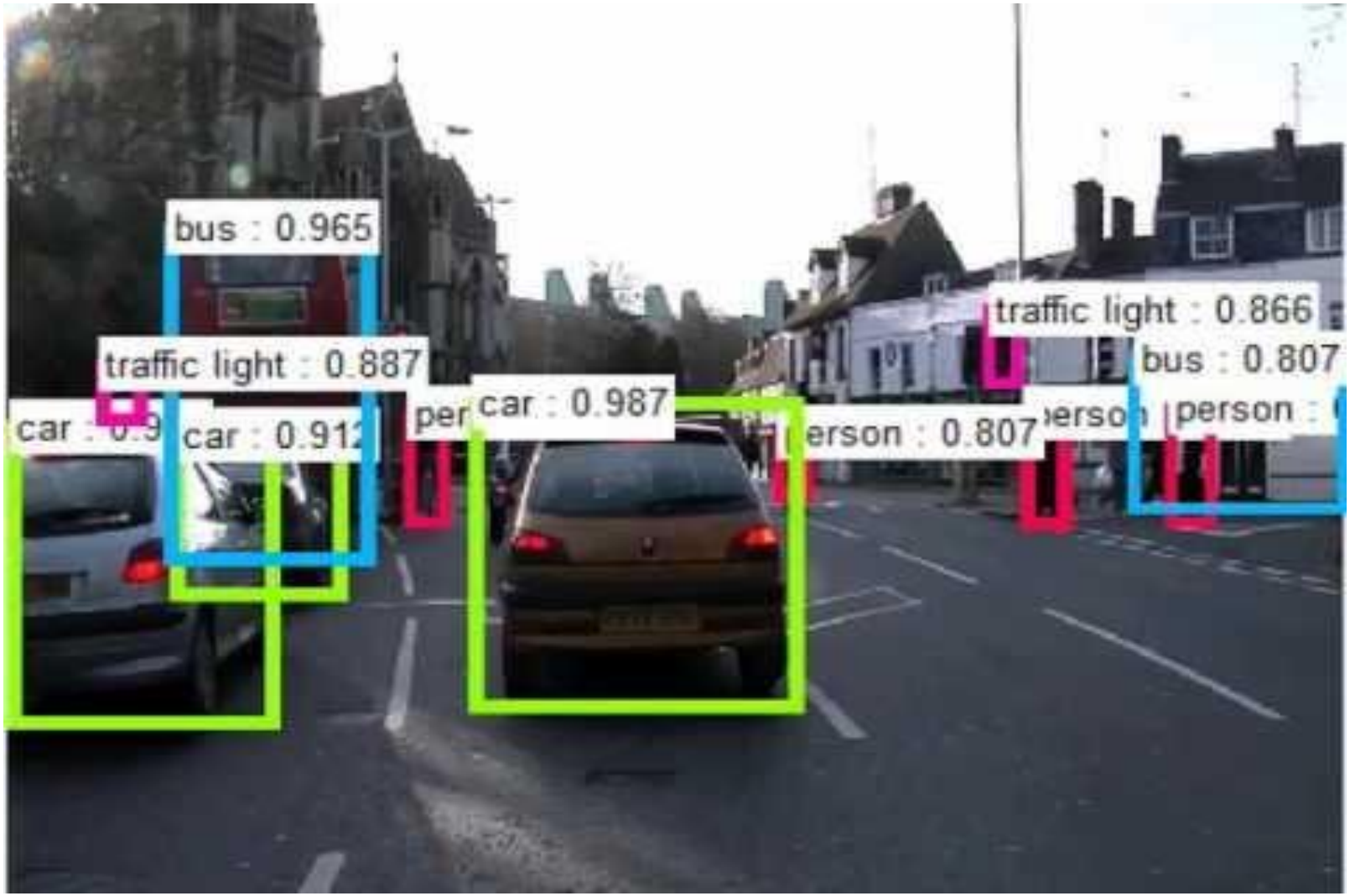






Applications

- Real time object recognition (CNN)
  https://www.youtube.com/watch?v=WZmSMkK9VuA



Applications

- Real time object recognition (CNN)
  https://youtu.be/70Kv8Rr72ag

Applications

- Image colorization (CNN)
  https://youtu.be/ys5nMO4Q0iY

Applications

- Self-driving car
  https://youtu.be/hLaEV72elj0

Applications

- Robotic

https://youtu.be/tf7IEVTDjng

Applications

- Robotic
  https://www.youtube.com/watch?v=kgaO45SyaO4

Homework

- Prepare a report on the use of convolutional neural networks in the image applications.