Two Weeks Short-term Training Program

On

Multi-modal Generative AI

16th -25th December, 2024

Organized by

Department of Artificial Intelligence

SVNIT, Surat

Deep Learning

by

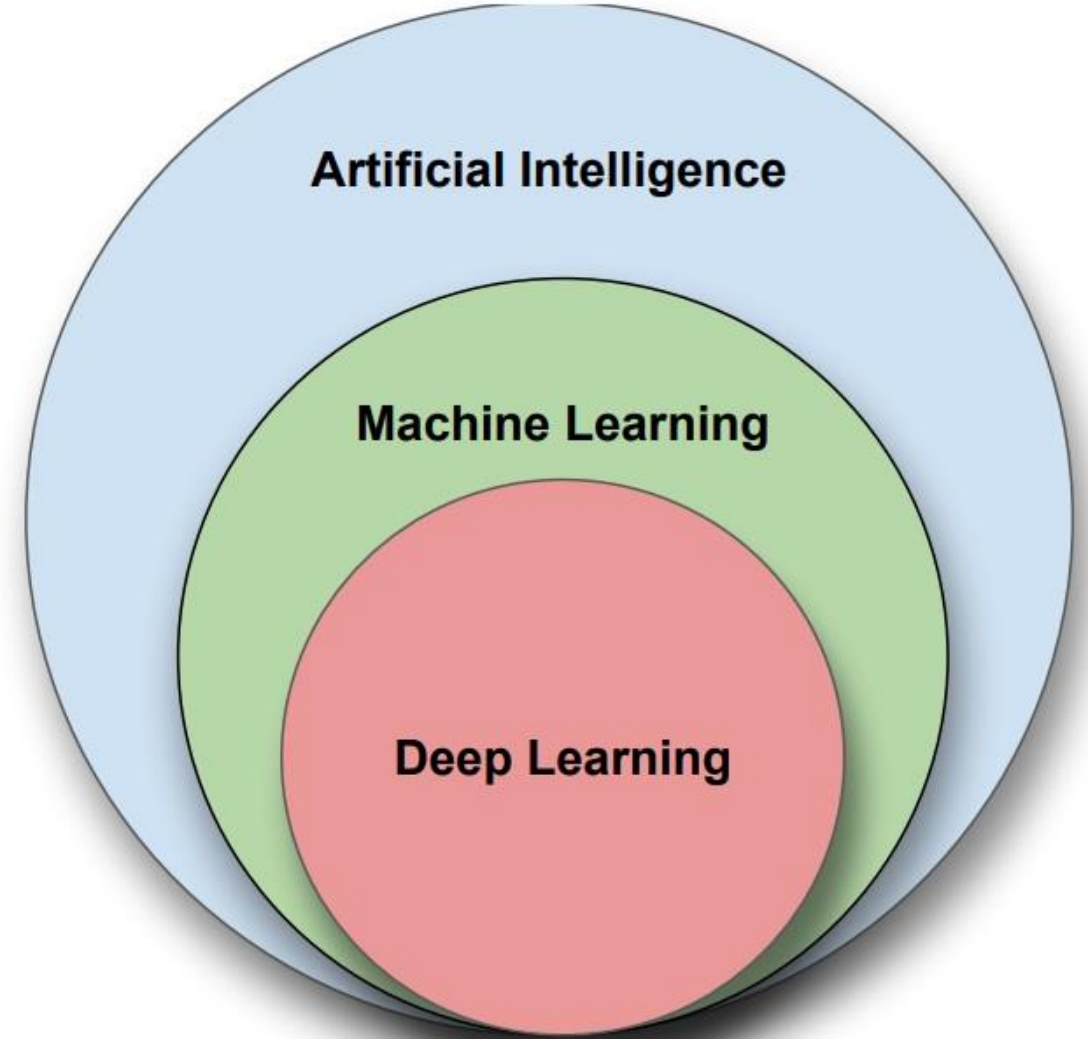
Dr. Tanmoy Hazra
Department of Artificial Intelligence
SVNIT Surat

Deep Learning

- **Artificial Intelligence (AI)**
- **Machine Learning (ML)**
- **Deep Learning (DL)**

- AI is intelligence demonstrated by machines, as opposed to natural intelligence.
- ML is the study of computer algorithms that can improve automatically through experience and using data.
- DL is part of a wider family of machine learning methods based on artificial neural networks.

Deep Learning



Types of Learning

- **Supervised**
  - (x, y) available training data set
  - For a new data x (input), predict the label of y (output)
  - This is applicable for labeled data

- **Unsupervised**
  - x is only given
  - Cluster the data based on x
  - This is applicable for unlabeled data

- **Reinforcement**
  - Learning made by rewards or penalty

Deep Learning

- Deep learning uses artificial neural networks to perform complex computations on large amount of data.

- This is also called intensed learning

- **Alternative definition:**
  - a type of machine learning based on artificial neural networks in which multiple layers of processing are used to extract progressively higher level features from data.

Key features

- Uses artificial neural networks.
- Learns from large datasets with minimal human intervention.
- Excels in complex tasks like image recognition, speech processing, and natural language understanding.

Deep Learning Algorithms

- Convolutional Neural Networks
- Recurrent Neural Networks
- Generative Adversarial Network
- Restricted Boltzman Machine
- Autoencoder
- Transformer

History

- 1943: Warren McCulloch and mathematician Walter Pitts created a model using neural networks for an electrical circuit.
- 1958: Frank Rosenblatt designed the first artificial neural network, called Perceptron.
- 1982: John Hopfield suggested creating a network which had bidirectional lines, similar to how neurons actually work.
- 1986: Neural networks use back propagation.

History

- 2006: Computer scientist Geoffrey Hinton has given a new name to neural net research as "**deep learning**".

- GoogleBrain (2012) - a deep neural network created by Jeff Dean, which focused on pattern detection in images and videos.

- AlexNet (2012): AlexNet won the ImageNet competition by a large margin in 2012.

History

- DeepFace (2014):
  - A DNN created by Facebook
  - Claimed it can recognize people with the same precision as a human can.
- DeepMind (2014):
  - it managed to beat a professional at the game Go.
- Generative Adversarial Neural Network (GAN) was introduced in 2014
- OpenAI (2015):
  - This is a non-profit organisation created by Elon Musk and others.

History

- ResNet (2015) - This was a major advancement in CNNs.
- U-net (2015)- a CNN architecture specialized in biomedical image segmentation.
- 2016: AlphaGo beat the world's number second player at Go game, beat the number one player in 2017.
- BERT (Bidirectional Encoder Representations from Transformers):
  - In 2017, Google’s BERT model transformed NLP, pre-trained models to understand context in language better;
  - applications in sentiment analysis and question answering.

History

- 2018: GPT (Generative Pretrained Transformer):
  - OpenAI released GPT, a language model capable of generating coherent and contextually relevant text.
- GPT-2 (2019):
  - OpenAI released GPT-2, a more powerful version of its predecessor,
  - capable of generating human-like text across a variety of domains.
- GPT-3 (2020):
  - OpenAI introduced GPT-3, very powerful language model, with 175 billion parameters
  - prompting new advancements in conversational AI and creative writing tools.

History

- 2021: CLIP (Contrastive Language-Image Pretraining):
  - OpenAI introduced CLIP, which linked text and images.

- 2022: Multimodal Models:
  - Increasing focus on multimodal models that combine text, images, and even video to improve performance across different types of data.

- 2023: GPT-4 (1.76 trillion parameters) and other advanced language models (Gemini, Llama, BERT, Meta, XAI, Claude, MistralAI etc.) further improved text generation, understanding, and reasoning capabilities.

History

- 2023: AI Ethics and Bias: A growing focus on responsible AI, addressing ethical concerns, bias, fairness, and transparency in deep learning systems.

- 2024: Transformer-Based Models in Multilingual NLP: Deep learning continues to improve multilingual understanding with models capable of handling multiple languages.

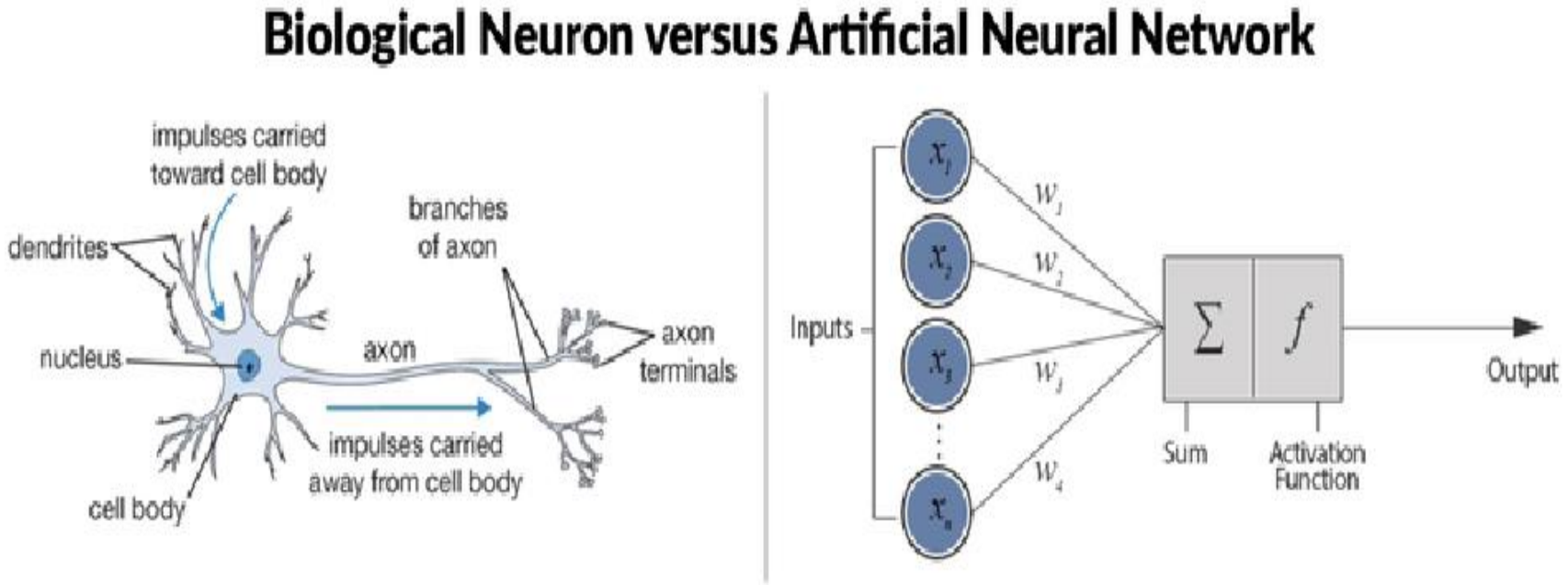
- 2024: AI in Creativity and Art: The integration of deep learning into various creative fields—such as AI-generated art, music, and even video creation

Mathematics for Deep Learning

- Probability and Statistics
  - Probability
  - Statistical inference
  - Validation
  - Estimates of error, confidence intervals
- Linear Algebra
  - Hugely useful for compact representation of linear transformations on data
  - Dimensionality reduction techniques
- Optimization Theory

Neural Networks

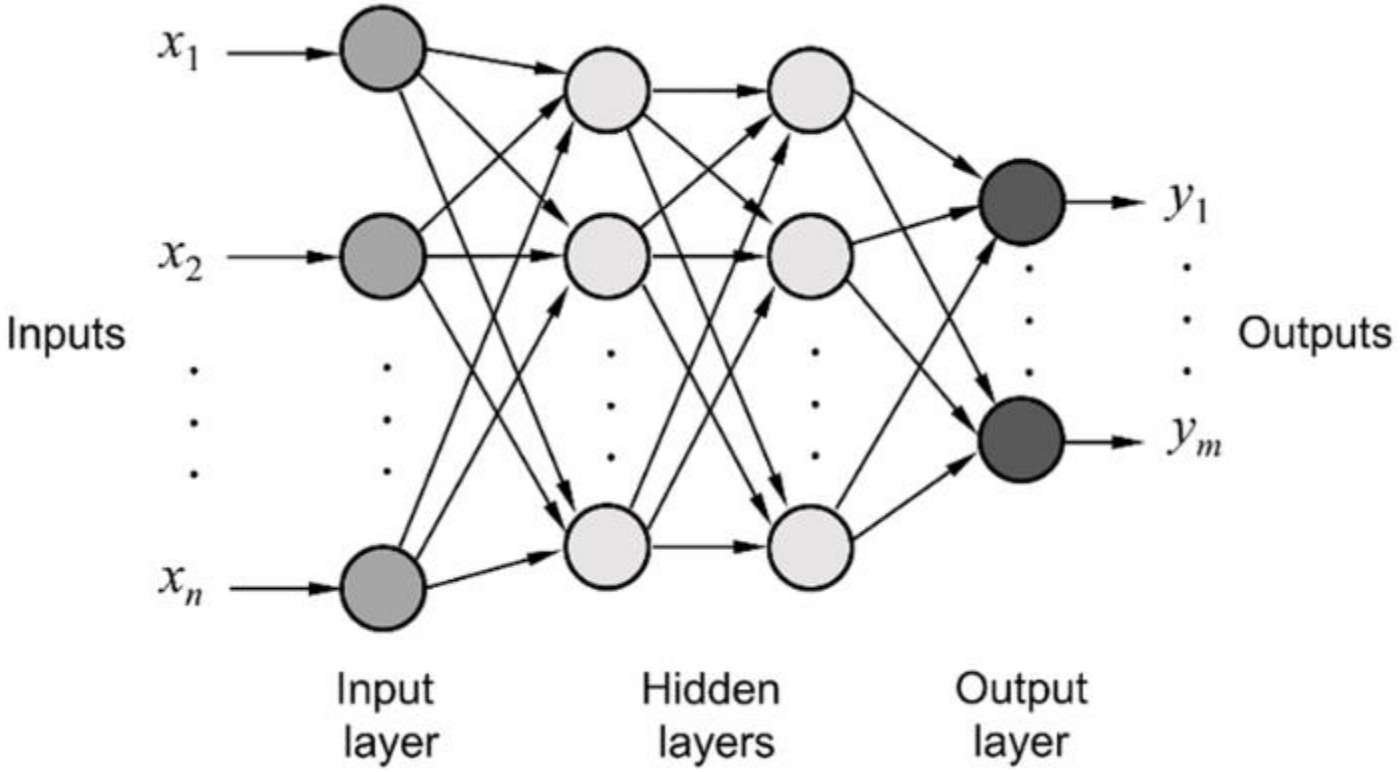
- 10 billions of neurons in human brains
- Massive parallelism
- Connectionism
- Distributed associative memory



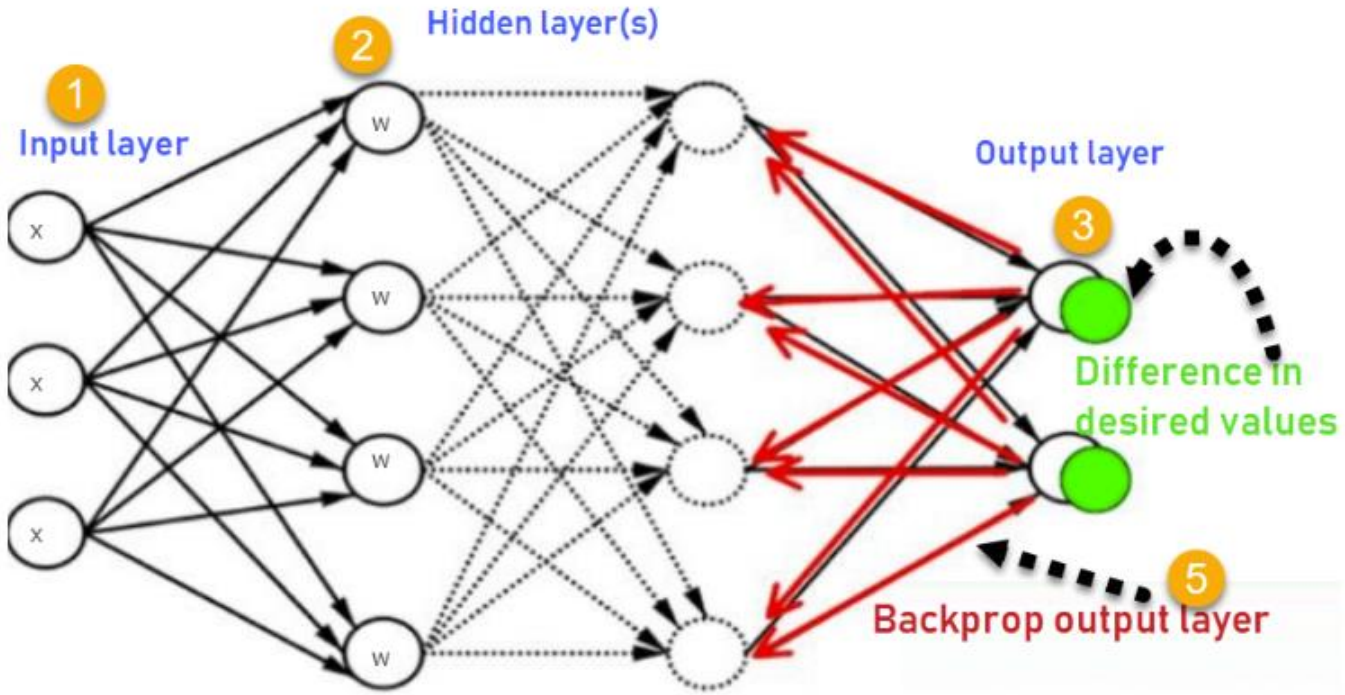
How neural network works?

- **Forward Propagation**:
  - Data moves through the network (input to output) to generate predictions.
- **Backpropagation**:
  - The error (loss) is propagated back through the network to update weights and minimize the error using optimization algorithms (e.g., gradient descent).
- **Learning Rate**:
  - Controls how much the weights are updated during training.

Multilayer Neural Networks



Multilayer Backpropagation Neural Networks



How Backpropagation Works

Implementing Logic Gates with MP Neurons

We can use McCulloch-Pitts neurons to implement the basic logic gates (e.g. AND, OR, NOT).

It is well known from logic that we can construct any logical function from these three basic logic gates.

All we need to do is find the appropriate connection weights and neuron thresholds to produce the right outputs for each set of inputs.

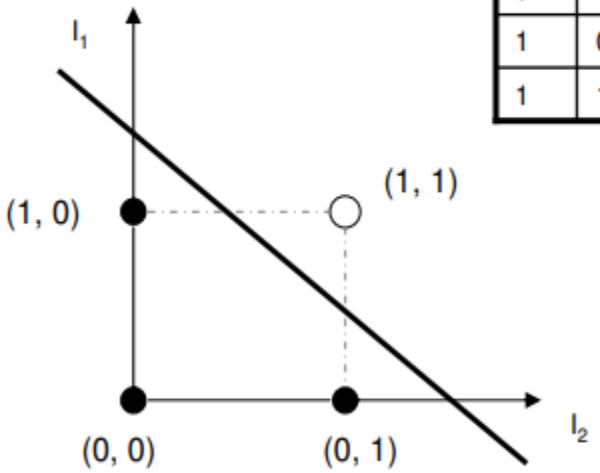
We shall see explicitly how one can construct simple networks that perform NOT, AND, and OR.

Decision Boundaries for AND and OR

We can now plot the decision boundaries of our logic gates

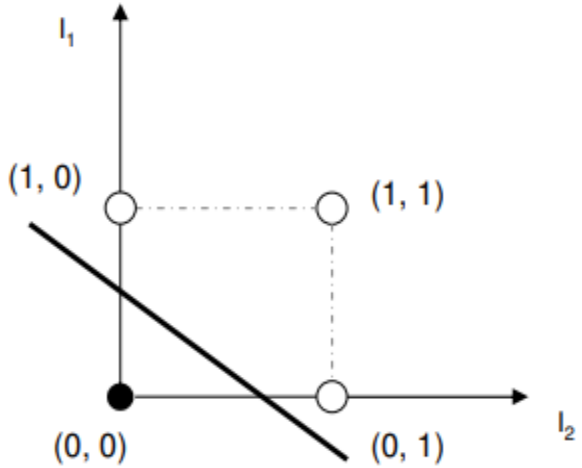
AND
w1=1, w2=1, θ=1.5

AND		
I_1	I_2	out
0	0	0
0	1	0
1	0	0
1	1	1



OR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	1

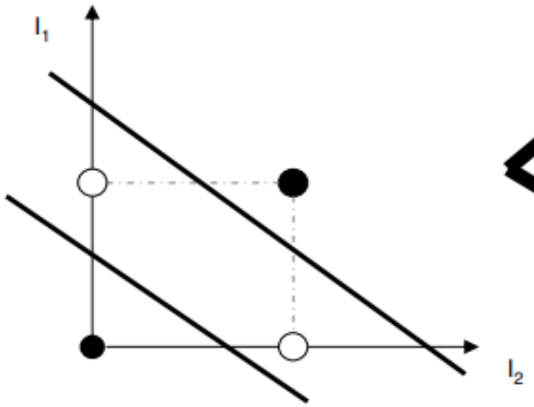
OR
w1=1, w2=1, θ=0.5

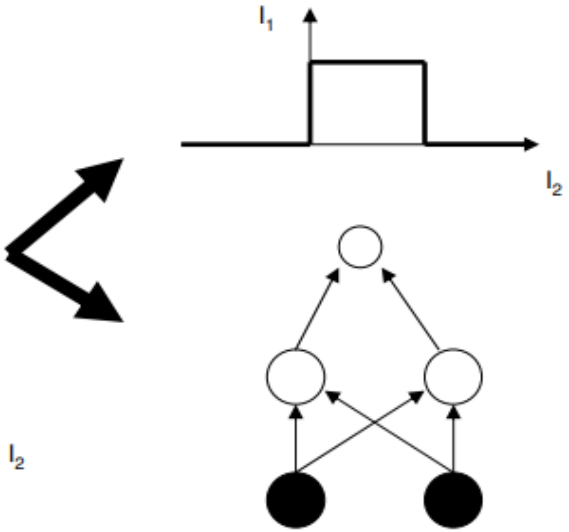


Decision Boundary for XOR

The difficulty in dealing with XOR is rather obvious. We need two straight lines to separate the different outputs/decisions:

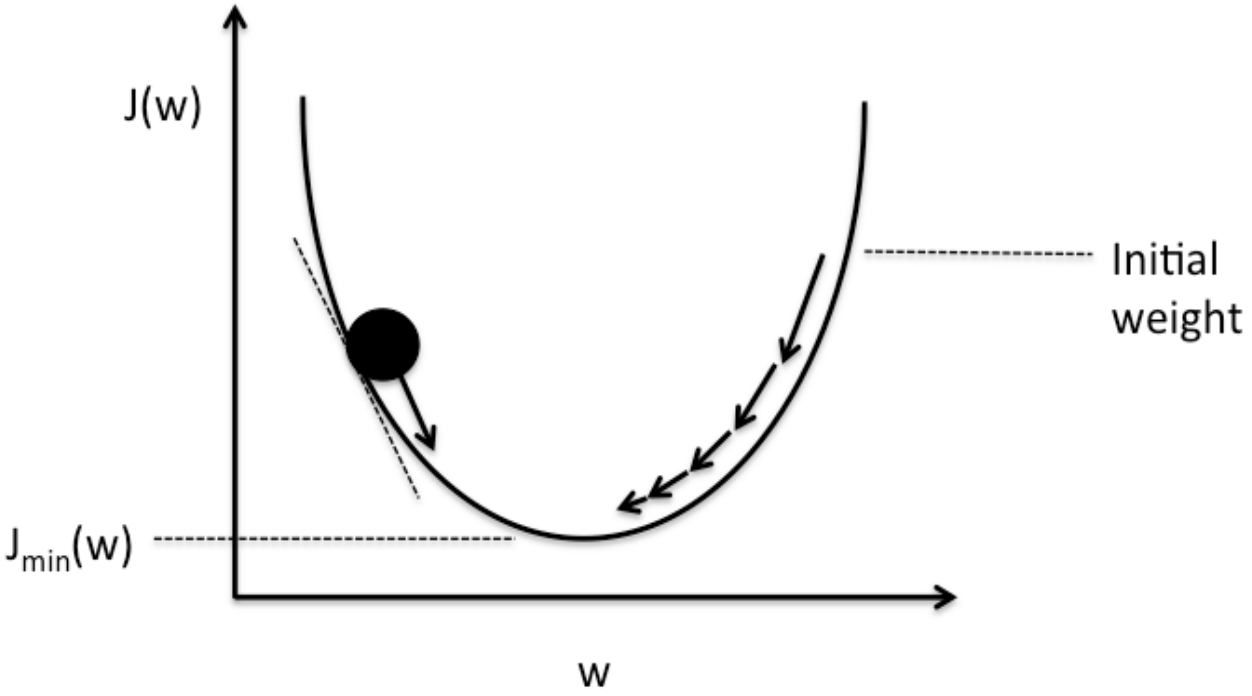
XOR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	0





Solution: either change the transfer function so that it has more than one decision boundary, or use a more complex network that is able to generate more complex decision boundaries.

Gradient descent Algorithm



sebastianraschka.com

Minimize cost function

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (wx_i + b - y_i)^2$$

$$\frac{\partial J}{\partial b} = \frac{2}{n} \sum_{i=1}^{n} (wx_i + b - y_i) \quad \Rightarrow \quad \frac{\partial J}{\partial b} = \frac{2}{n} \sum_{i=1}^{n} (pred_i - y_i)$$

$$\frac{\partial J}{\partial w} = \frac{2}{n} \sum_{i=1}^{n} (wx_i + b - y_i)\, x_i \quad \Rightarrow \quad \frac{\partial J}{\partial b} = \frac{2}{n} \sum_{i=1}^{n} (pred_i - y_i)\, x_i$$

Updated values of w and b is given as

$$w = w - \alpha \cdot \frac{2}{n} \sum_{i=1}^{n} (pred_i - y_i)\, x_i$$

$$b = b - \alpha \cdot \frac{2}{n} \sum_{i=1}^{n} (pred_i - y_i)$$

Restricted Boltzman Machine

- Restricted Boltzmann Machine technique, used for feature selection and feature extraction
- It is a type of generative model that is capable of learning a probability distribution over a set of input data.
- A RBM is a type of stochastic neural network used primarily for unsupervised learning.
- It consists of two layers: a visible layer representing observed data and a hidden layer capturing features or patterns.

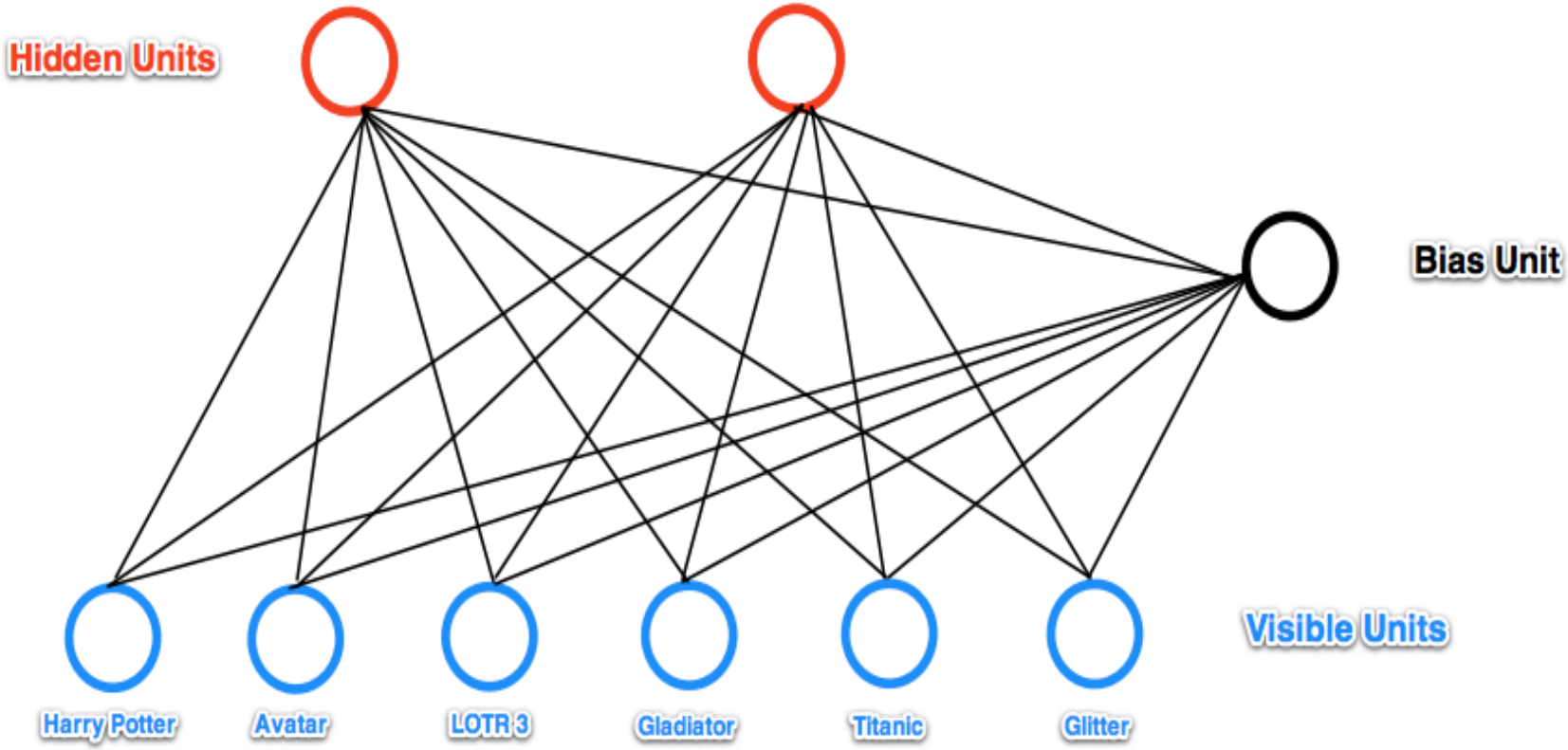
Restricted Boltzman Machine

- The RBM operates by probabilistically activating neurons in the visible and hidden layers based on the energy of the system, using the concept of a Boltzmann distribution.
- During training, the model adjusts its weights and biases to minimize the difference between the input data and its reconstructed version, often using contrastive divergence to optimize the network.

Restricted Boltzman Machine

- The "randomized" aspect comes from the probabilistic nature of neuron activation, where the model stochastically decides whether to turn neurons on or off, leading to a better exploration of the feature space.
- This process enables RBMs to learn useful representation of data without needing labeled examples.

Restricted Boltzman Machine



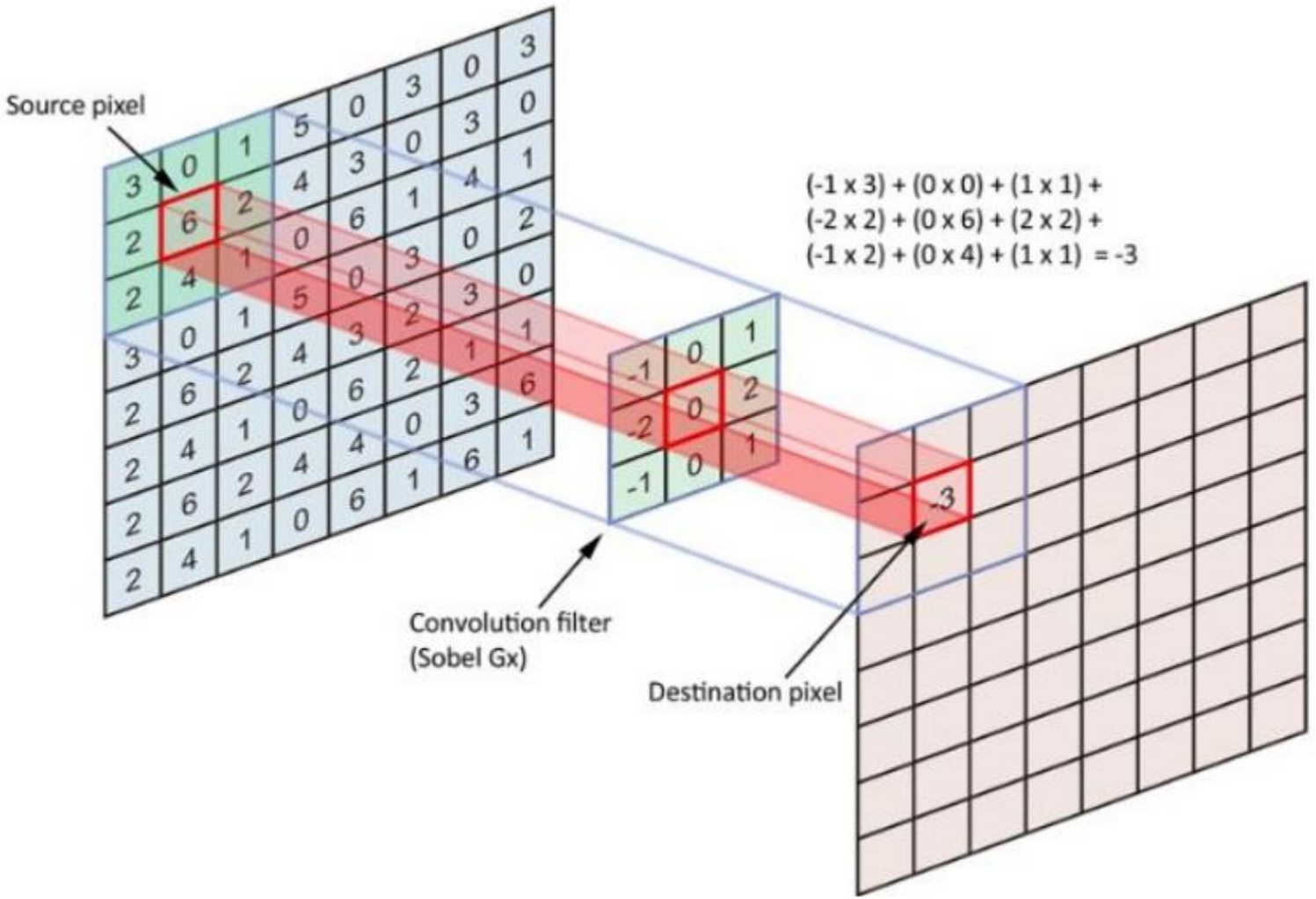
Convolutional Neural Networks

- A convolutional neural network can have tens or hundreds of layers that each learn to detect different features of an image.
- Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer.
- The filters can start as very simple features, such as brightness and edges, and increase in complexity to features that uniquely define the object.

Convolutional Neural Networks

- **Methodology:**
  - It consists of layers that perform convolution operations, where small filters or kernels slide over the input data (like an image) to detect local patterns such as edges, textures, and shapes.
  - These convolutional layers are followed by pooling layers that downsample the feature maps, reducing their dimensionality and computational complexity while retaining the most important features.
  - After several convolution and pooling layers, the network typically includes fully connected layers to make final predictions or classifications.

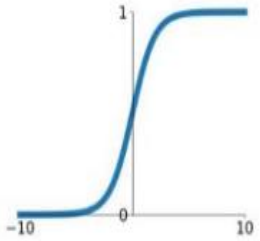
Convolution in 2D



Activation Function

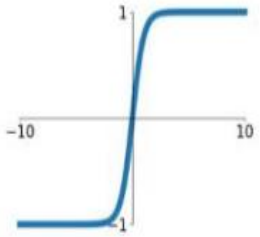
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



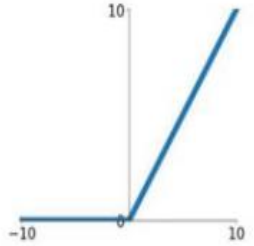
tanh

$$\tanh(x)$$



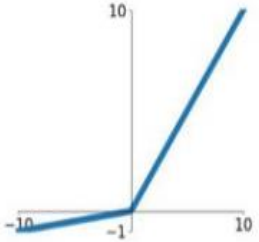
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

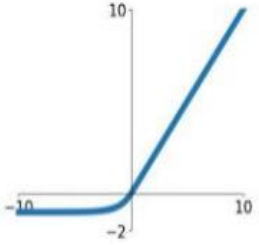


Maxout

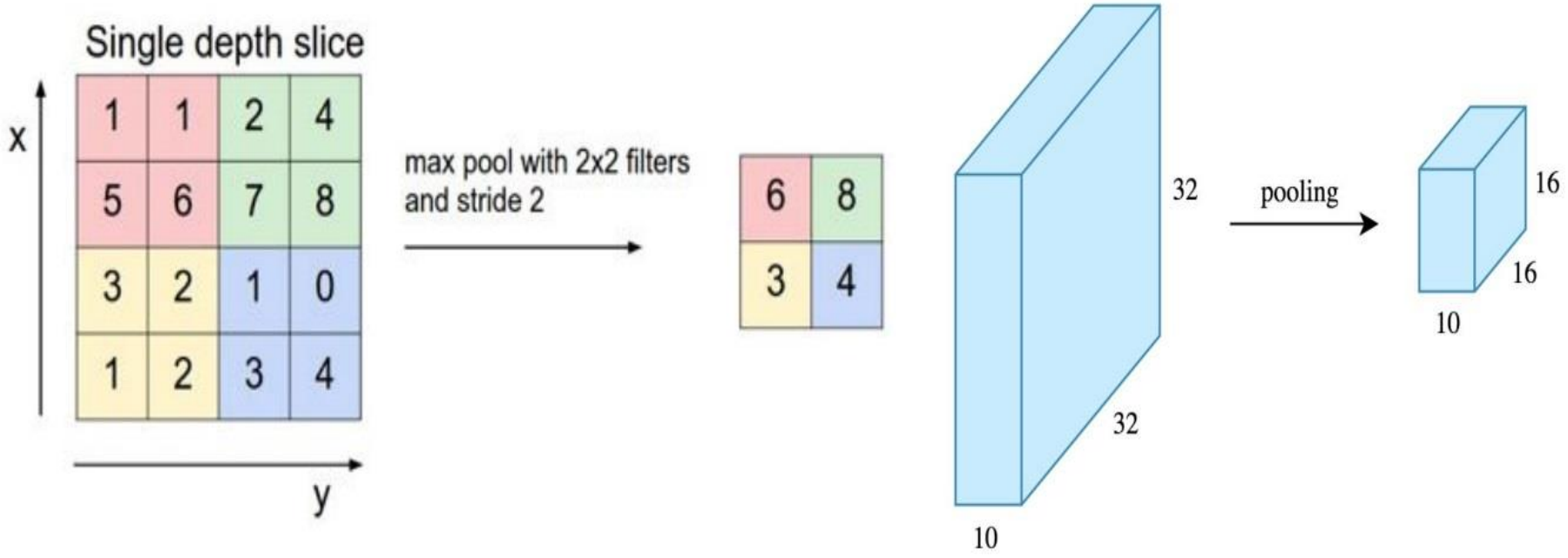
$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

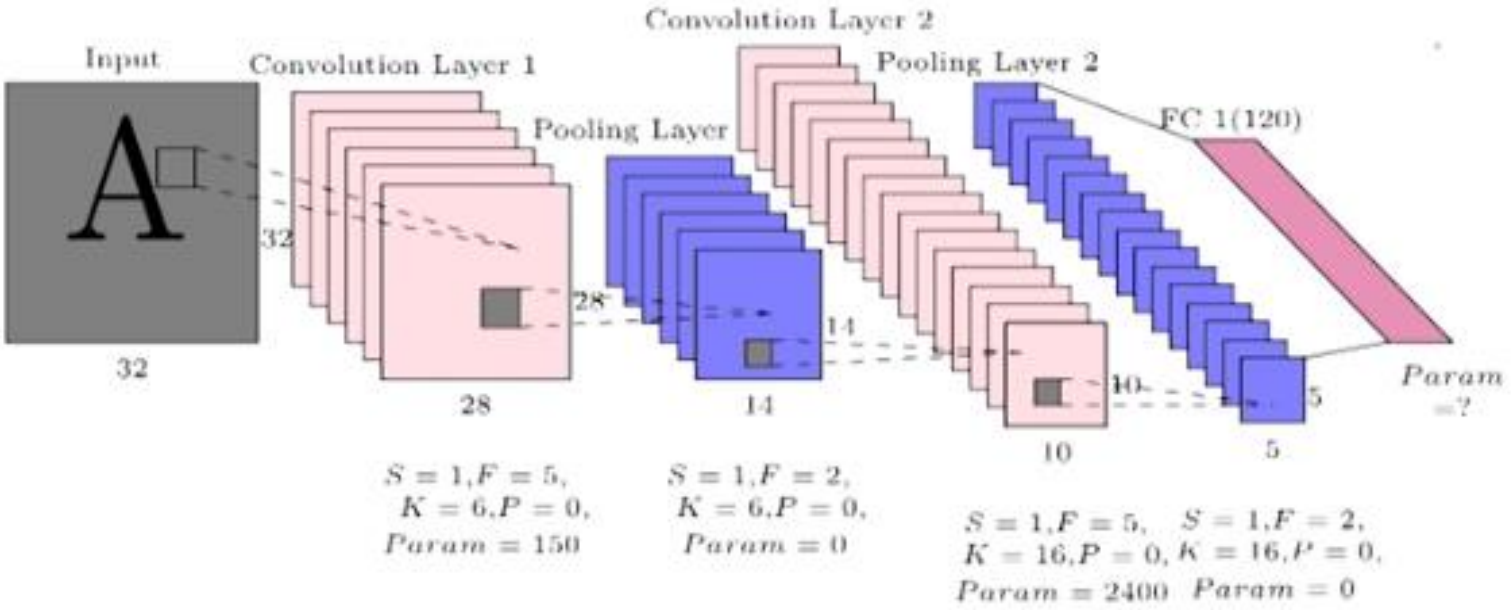
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



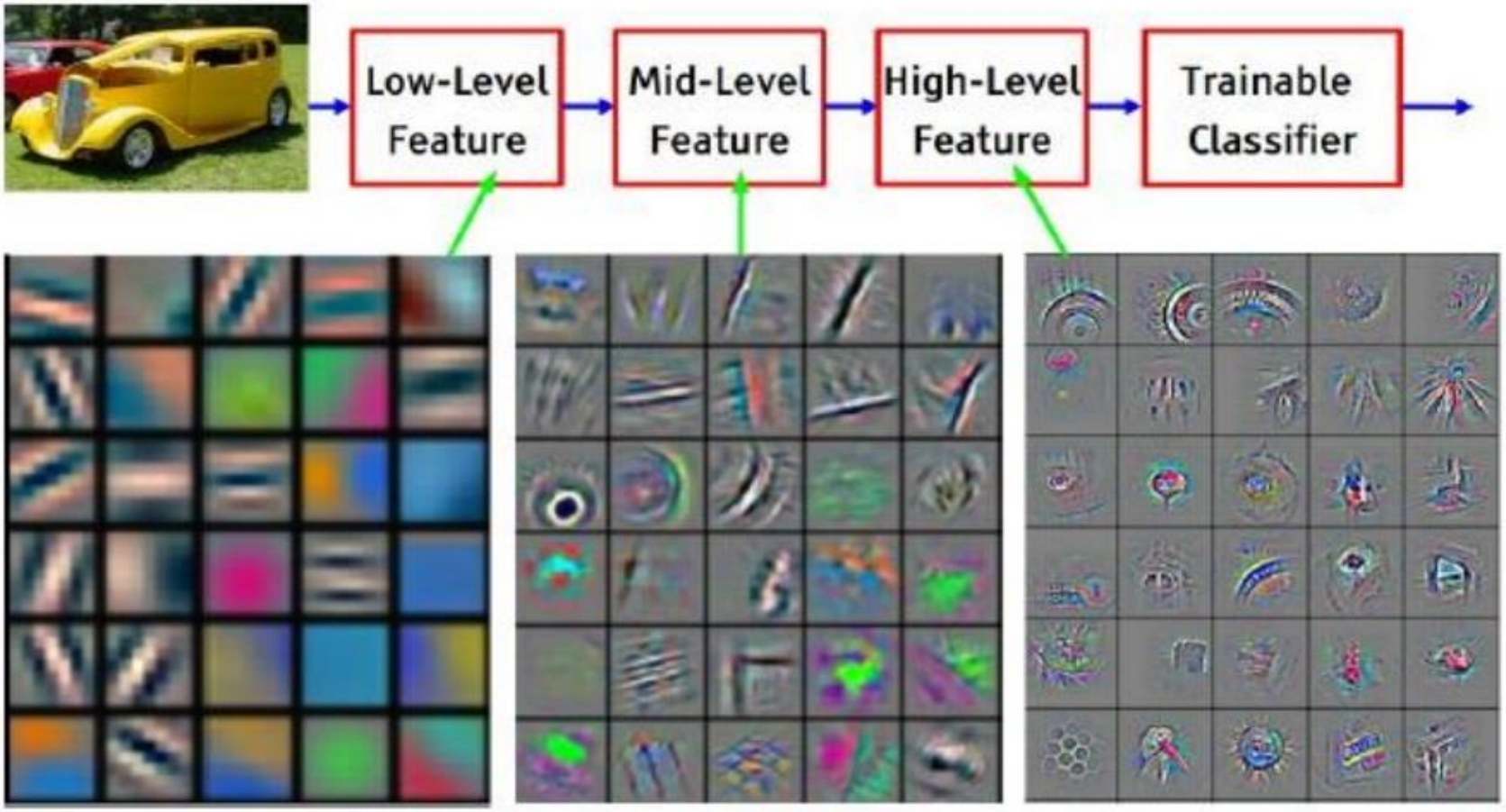
Max Pooling


Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4
32
32
10
pooling
16
16
10

Convolutional Neural Networks

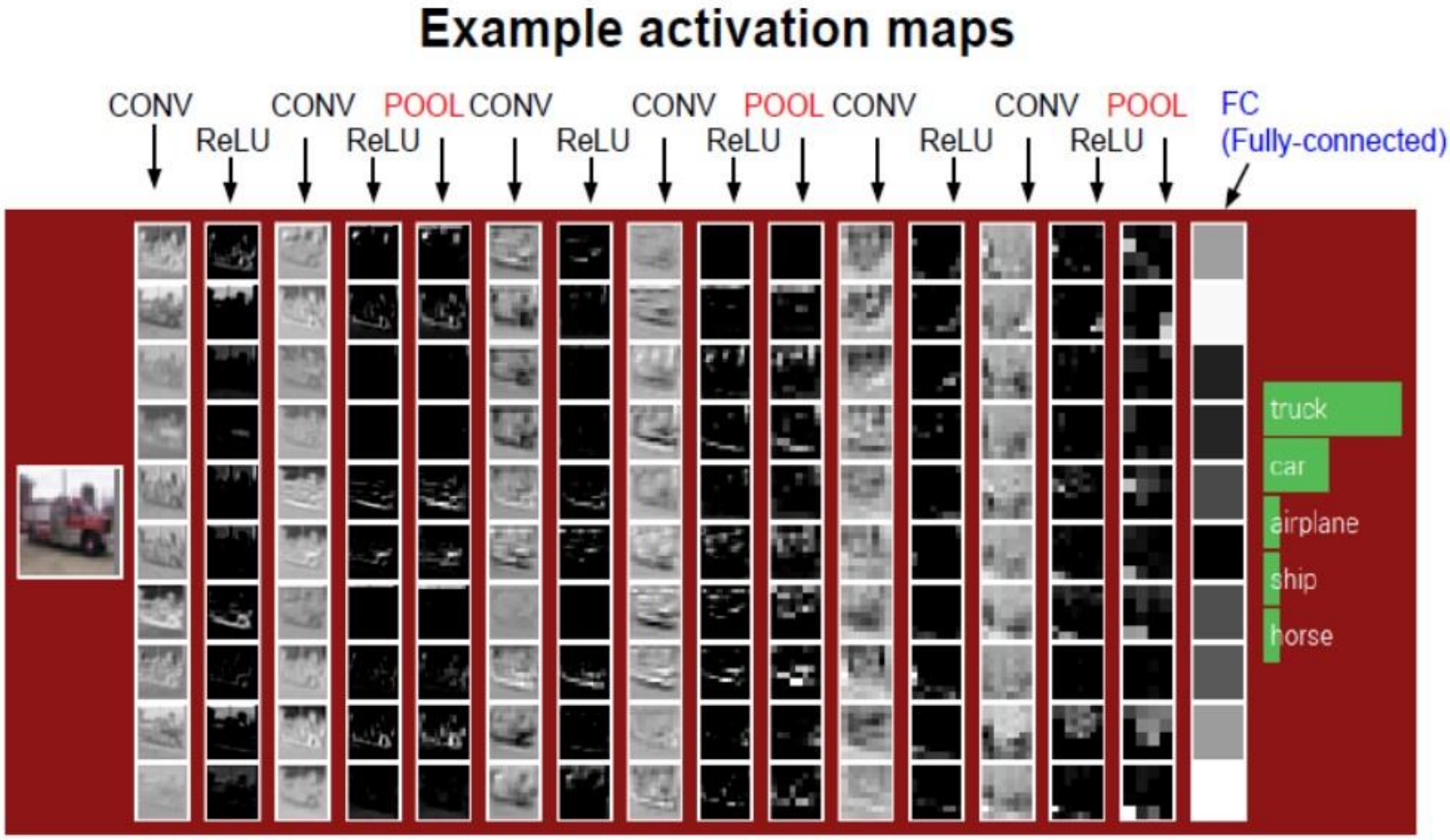


What do the neurons learn?

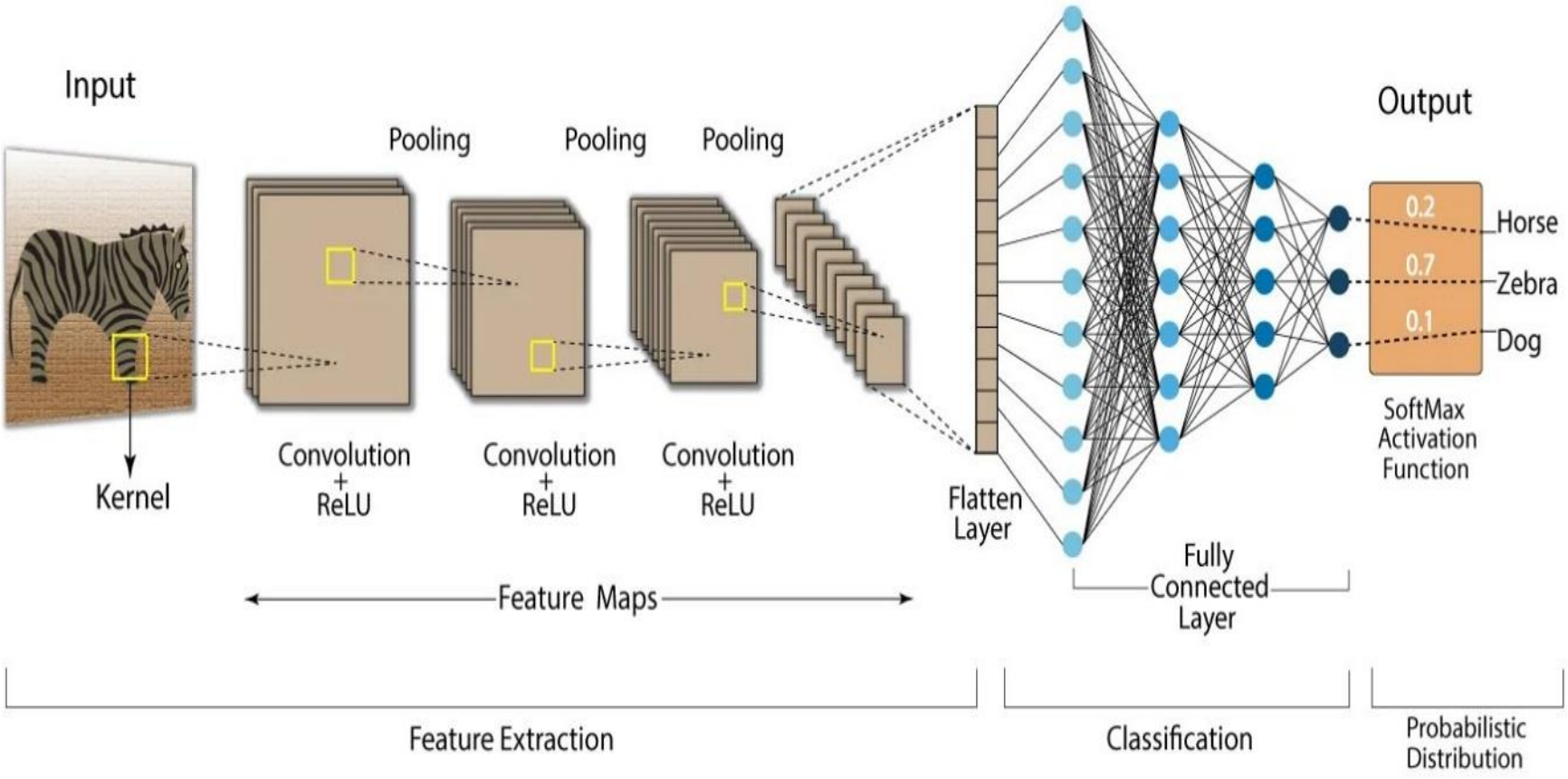


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

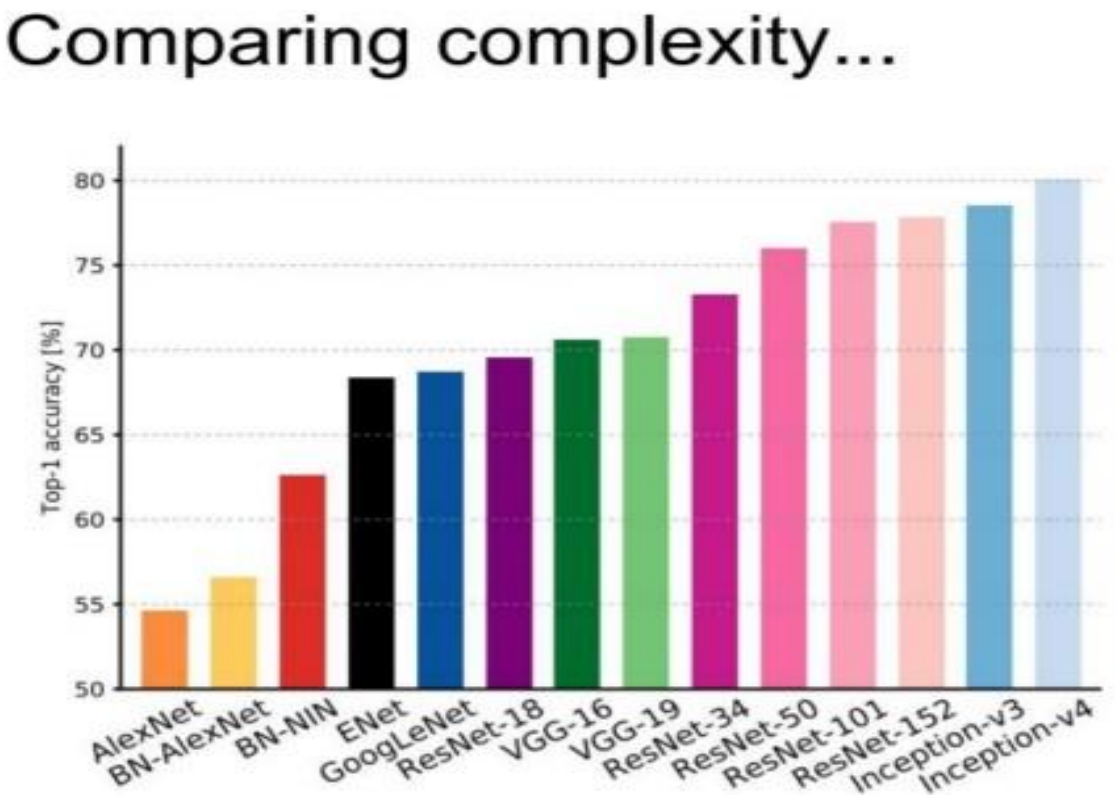
Example activation maps



Representation



Computational Complexity



Key Takeaways

- **Use Cases**: Image classification, object detection, facial recognition.
- **How It Works**:
  - Convolutional layers apply filters to detect patterns.
  - Pooling layers reduce dimensionality.
  - Fully connected layers for final prediction.
- **Example**: ImageNet classification (e.g., AlexNet, ResNet).

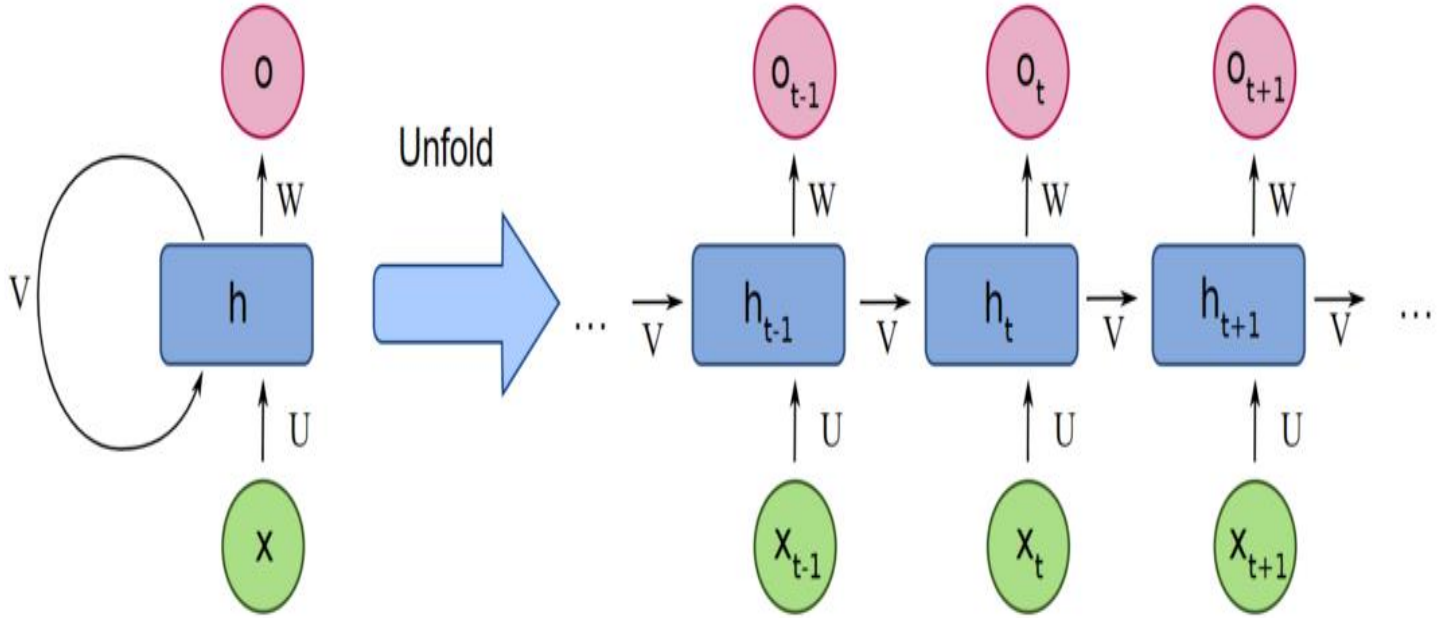
Recurrent Neural Network

- A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data.
- These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (NLP), speech recognition, and image captioning

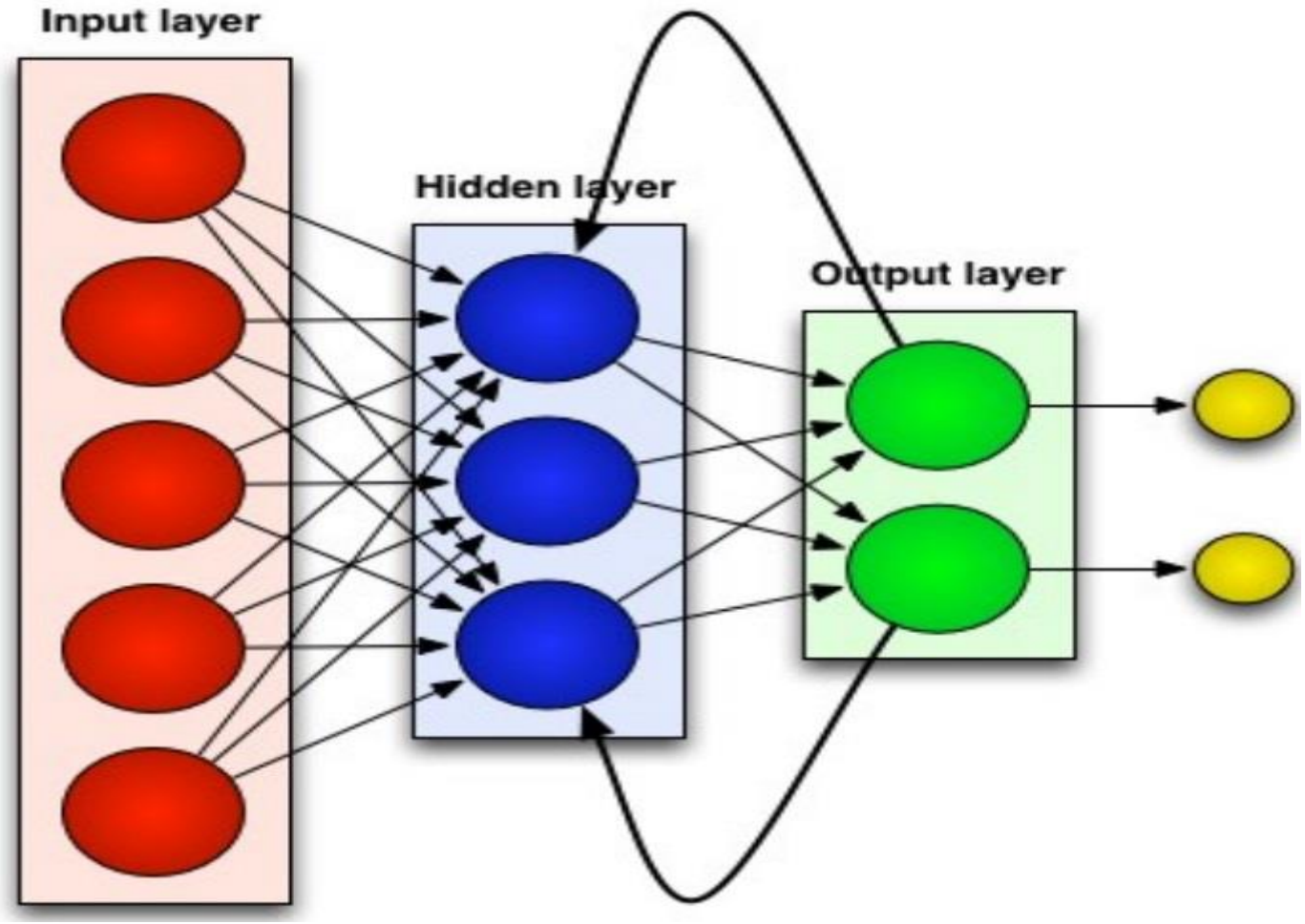
Recurrent Neural Network

- RNNs have hidden states that are updated at each time step based on the current input and the previous hidden state, effectively creating a form of memory.
- However, traditional RNNs can struggle with long-term dependencies due to issues like vanishing gradients, which is why advanced variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are often used to improve performance.

Recurrent Neural Network



Recurrent Neural Network



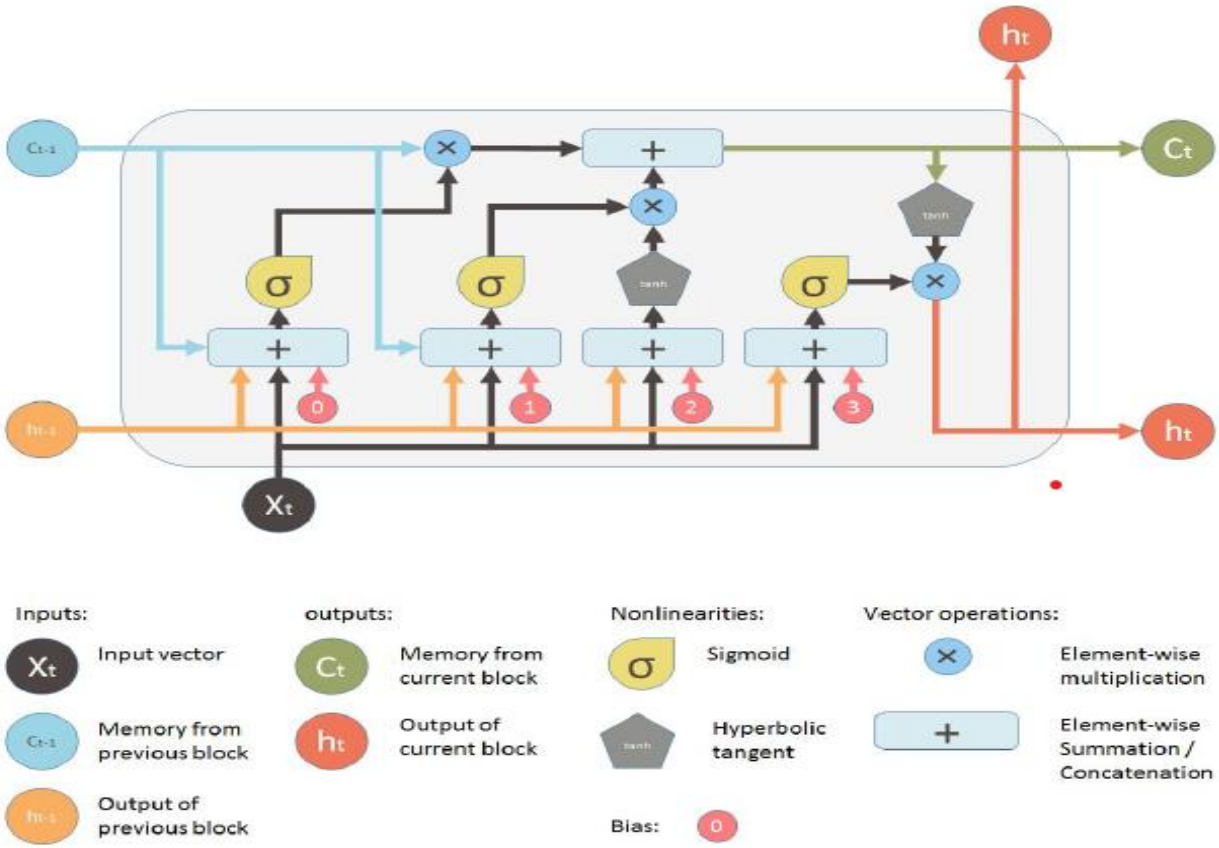
Key Takeaways

- **Use Cases**: Sequence data, speech recognition, text generation.
- **How It Works**:
  - RNNs have loops allowing information to persist.
  - They are good at handling temporal dependencies in data.
- **Variants**: Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU).

Long Short Term Memory

- It is special kind of recurrent neural network that is capable of learning long term dependencies in data.
- LSTMs are predominantly used to learn, process, and classify sequential data because these networks can learn long-term dependencies between time steps of data.
- Common LSTM applications include sentiment analysis, language modeling, speech recognition, and video analysis.

Long Short Term Memory



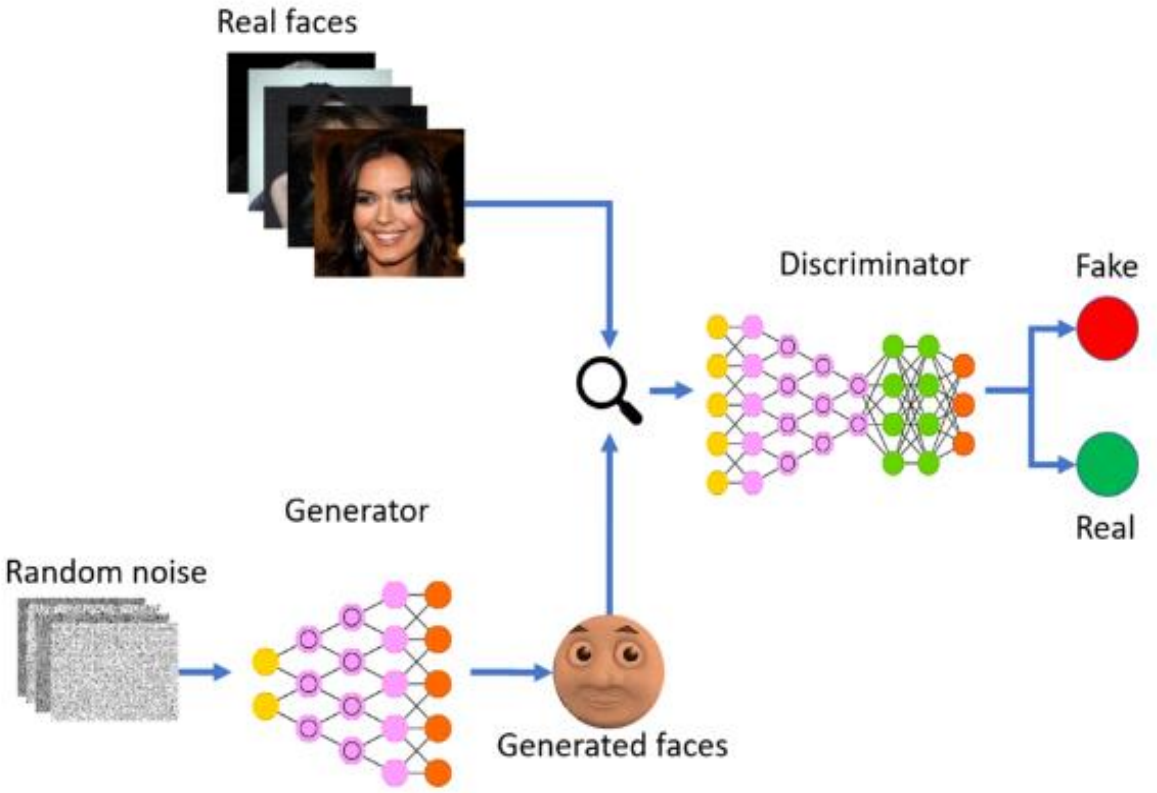
Generative Adversarial Network

- A generative adversarial network (GAN) is a deep learning architecture.
- It trains two neural networks (discriminator and generator) to compete against each other to generate more authentic new data from a given training dataset.
- During training, the generator improves its ability to create more convincing data, while the discriminator becomes better at identifying fake data.

Generative Adversarial Network

- This adversarial process continues until the generator produces data that is nearly indistinguishable from real data, and the discriminator is unable to tell the difference.
- For instance, we can generate new images from an existing image database or original music from a database of songs.

Generative Adversarial Network



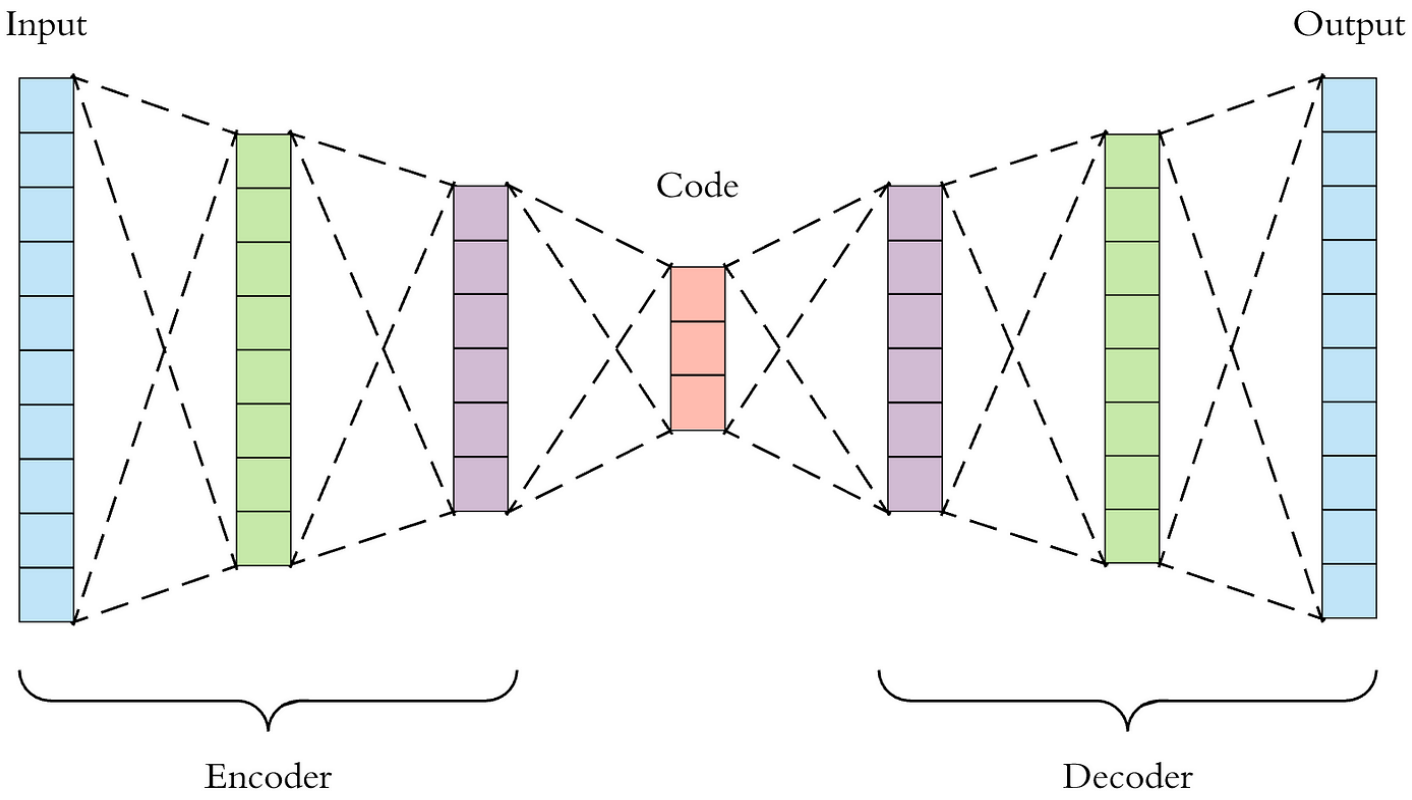
Key Takeaways

- **What is a GAN?**: A type of neural network architecture where two networks (Generator and Discriminator) compete against each other.
- **Use Cases**: Image generation, style transfer, data augmentation.
- **How it Works**:
  - Generator creates fake data.
  - Discriminator evaluates if the data is real or fake.
  - Both networks improve through adversarial training.

Autoencoder

- Autoencoders are neural network models primarily used for unsupervised learning tasks such as dimensionality reduction, data compression, and feature extraction.
- They learn to reconstruct the input data and capture its essential patterns, making them useful for anomaly detection and image-denoising tasks
- An autoencoder learns two functions: an encoding function that transforms the input data, and a decoding function that recreates the input data from the encoded representation.

Autoencoder



Transformer

- Self-Attention:
  - Allows the model to weigh the importance of each word in a sequence relative to others; capture long-range dependencies.
- Multi-Head Attention:
  - Uses multiple attention mechanisms in parallel to capture different aspects of relationships within the input.
- Position Encoding:
  - Adds information about the order of tokens since the Transformer doesn’t inherently process sequences step by step like RNNs.

Transformer

- Feed-Forward Networks:
  - Each token's representation is passed through a simple fully connected network.

- Encoder-Decoder Structure:
  - Encoder processes the input sequence.
  - Decoder generates the output sequence.

- Parallelization:
  - Unlike RNNs, Transformers process all tokens in a sequence simultaneously, making training faster.

Transformer

Applications:

- NLP: Models like BERT, GPT etc.
- Computer Vision: Vision Transformer (ViT).
- Speech Processing: Speech-to-text, language modeling.

DL Models for Various Problems

- Computer Vision – GANs and CNN
- Natural Language Processing – BERT, Attention, Memory networks, RNN, LSTM, GRU and CNN
- Adversary Attacks Detection – GAN
- Object Detection – YOLO models

DL Models for Various Problems

- Semantic Segmentation – Mask RCNN
- Image Classification – CNNs
- Sequence Problem Prediction – RNN, LSTM, GRU
- Linear Problems Modeling and Analysis – ANN

Tools and Languages for DL Models

- Scilab
- Pycharm
- Matlab
- OpenCV
- Jupyter
- Anaconda
- Popular languages: Python, R

Research on Deep Learning

- To detect and classify
  - objects in images
  - objects in videos
  - emotions in images
  - emotions in audio
- To generate
  - new images from a given set of images.
  - new audio from a given set of audio

Research on Deep Learning

- To detect and classify
  - emotions in text
  - objects in medical images
  - objects in satellite images
  - objects in speech recognition
  - objects in gesture recognition
  - objects in sentiment analysis

Research on Deep Learning

- To detect and classify
  - objects in time series analysis
  - objects in anomaly detection
  - objects in recommender systems
  - objects in medical diagnosis
  - objects in fraud detection

Current research trends in DL

- **Transfer Learning**: Fine-tuning pre-trained models for specific tasks.
- **Explainability**: Making deep learning models interpretable to humans (e.g., SHAP, LIME).
- **Self-Supervised Learning**: Leveraging unlabeled data to pre-train models.
- **AI Ethics and Bias**: Addressing fairness and reducing bias in AI models.
- **Edge AI**: Deploying deep learning models on edge devices (e.g., smartphones, IoT).

Our Latest Research Works

1. **Classification of cotton crop disease using hybrid model and MDFC feature extraction method (Journal of Phytopathology, 2023)** *[P. Nimbhore, R. Tiwari, T. Hazra, M. Yadav]*

   - A novel Modified Deep Fuzzy Clustering (MDFC) based classification model.
   - The features are put through a detection phase, after which the extracted features are trained in the Bidirectional Gated Recurrent Unit (Bi-GRU) model to determine whether or not the cotton crop is infected.
   - Once it is detected to be diseased, the type of disease is classified via an improved Recurrent Neural Network (RNN).
   - MDFC-based classification model outperforms existing models with a specificity of 0.9687

Our Latest Research Works

2. Human face generation from textual description via style mapping and manipulation (Multimedia Tools and Applications, 2022) *[Shantanu Todmal, Ashish Mule, Devang Bhagwat, Tanmoy Hazra, Bhupendra Singh]*

- A Text-to-Face generative model that can produce high quality and high resolution images from a given textual description.
- The model is also able to produce a range of diverse images for a given description.
- Applications of the model: criminal investigation, character generation (video games, movies etc.), manipulating facial attributes according to brief textual description, text based style transfer, text based Image retrieval etc.

Our Latest Research Works

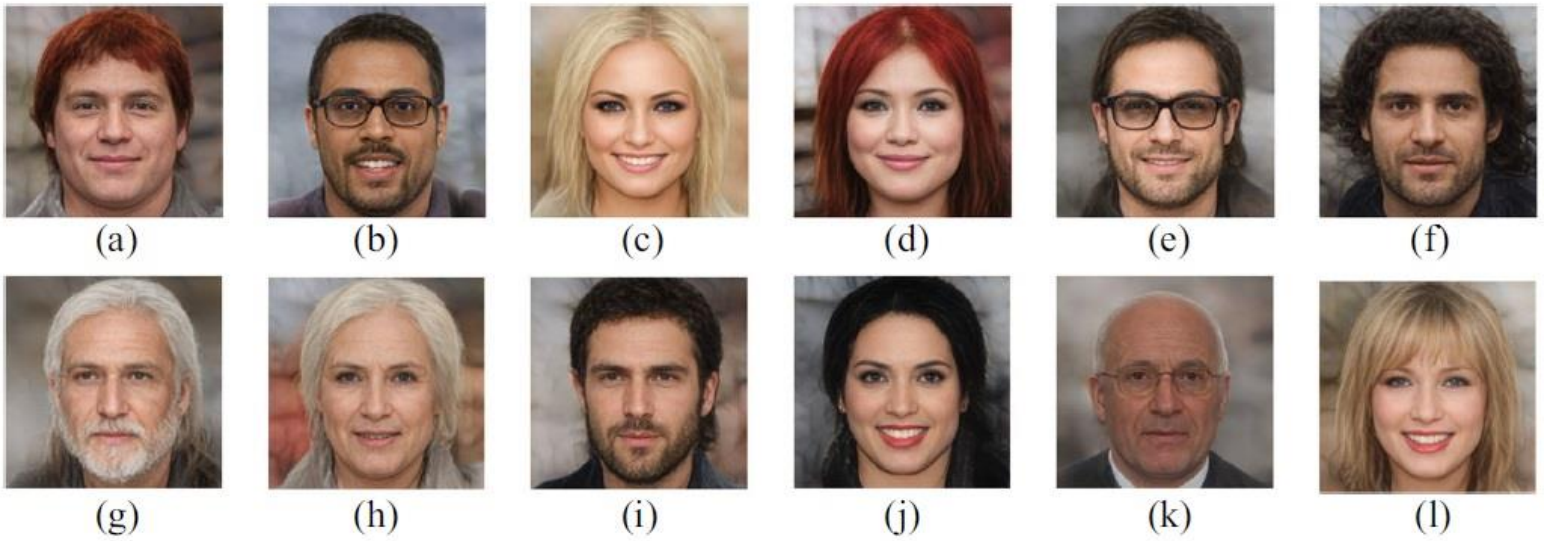


Fig. 9 Images generated from their textual descriptions [Proposed Method 1]. (**a**) Man with double chin having big eyes and bags under eyes with red hair, (**b**) Attractive Young man with short hair and heavy mustache wearing glasses having mouth open, (**c**) Beautiful woman with blonde hair wearing makeup having arched eyebrows smiling, (**d**), Chubby woman with big eyes having red hair wearing makeup smiling, (**e**) Handsome man having beard wearing glasses smiling, (**f**) Man with curly hair having bags under eyes, (**g**) Old man with white hair and beard, (**h**) Old woman with white hair having pointy nose smiling, (**i**)He has mouth slightly open, big nose, and sideburns. He is young. He has beard. (**j**)She wears lip- stick, heavy makeup, has high cheekbones, mouth slightly open, and black hair. She is attractive, (**k**)This person wears eyeglasses, has bags under eyes, and has gray hair. He is bald, (**l**)She wears lip- stick. She has bangs, mouth slightly open, high cheekbones, and blond hair. She is smiling, and attractive

Our Latest Research Works

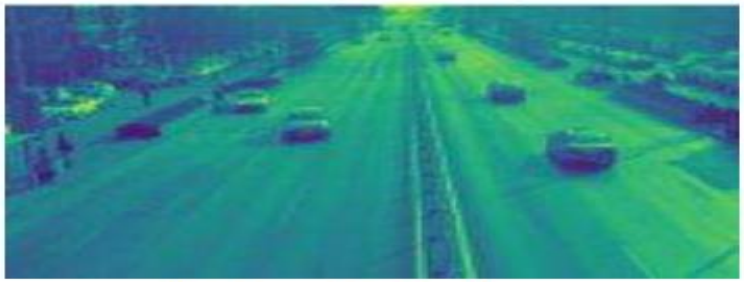
3. Vehicle Detection under Tunnel using Background Subtraction Technique (European Chemical Bulletin, 2022)

[Prerna Rawat, Tanmoy Hazra, Bhupendra Singh]

- Study evaluates various methodologies, such as conventional computer vision techniques and deep learning-based algorithms
- Methodology used bar filter and grate filter-based technique for vehicle detection and shows better outcome than the traditional methods
- Convolution neural networks (CNNs) have demonstrated impressive performance in vehicle detection tasks

Our Latest Research Works

(a) Original Image



(b) Gray Scale Image

(c) Gradient Image

Our Latest Research Works

4. Applications of Game Theory in Deep Learning (Multimedia Tools and Applications, 2021) *[Tanmoy Hazra, Kushal Anjaria]*

- It provides a comprehensive overview of the applications of game theory in deep learning.
- Existing research contributions demonstrate that game theory is a potential approach to improve results in deep learning models.
- The design of deep learning models often involves a game-theoretic approach.
- GAN is a deep learning architecture that is popular in solving complex computer vision problems; the training of the generators and discriminators in GANs is essentially a two-player zero-sum game

Future Directions

- Meta-Learning: Training approaches used for small-scale instances
- fault-tolerant deep learning models for small training data
- Deep learning model execution over mobile devices
- Use of DL in neuroscience
- Applications of deep learning in game theory
- Extracting complete information from partial information using DL
- Deep learning in quantum environment
- GenAI, XAI

Challenges

- Time-consuming process
- Using small-scale data leads to overfitting issue
- Observation-based learning
- Use large-scale data for processing
- Training / Processing of large data is costly
- Hardware constraints and cost

References

1. Todmal, S., Mule, A., Bhagwat, D., Hazra, T., & Singh, B. (2023). Human face generation from textual description via style mapping and manipulation. *Multimedia Tools and Applications*, *82*(9), 13579-13594
2. Hazra, T., & Anjaria, K. (2022). Applications of game theory in deep learning: a survey. *Multimedia Tools and Applications*, *81*(6), 8963-8994
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*
4. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444
5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press
6. Web resources

Thank You