Classification

Mohammadreza Ahmadi Teshnizi
Sajad Sabzi

NAME	SUMMARY	USAGE	YEAR	DIV
BASIC-L	Method for discovering optimization algorithms for deep neural network training, resulting in the creation of an efficient optimization algorithm called Lion, which outperforms other popular optimizers in various tasks.	BASIC-L is a Bayesian algorithm that learns from labeled and unlabeled data. It can handle different types of learning problems, data sets, and classes.	2023	2D
CoCa	A pretrained image-text model that combines contrastive and captioning losses. CoCa achieves state-of-the-art performance on various computer vision tasks, including image classification and crossmodal retrieval.	It can learn rich image-text representations from different types of text labels, using both contrastive and generative objectives, without any convolutional networks.	2022	2D
ViT-L/16	A pure transformer-based architecture for computer vision tasks like image classification. ViT outperforms convolutional networks on various benchmarks while needing fewer computational resources when pretrained on large datasets.	It can scale well with data and model size, benefit from higher resolution inputs, and be easily transferred to other vision tasks.	2020	2D
AlexNet	AlexNet is a deep convolutional neural network (CNN) model that revolutionized image classification by significantly outperforming previous approaches. It consists of multiple convolutional and fully connected layers.	Excellent performance on large-scale image classification tasks, introduced the concept of deep learning for computer vision tasks.	2012	2D
VGGNet	VGGNet is a deep CNN model that uses very small (3x3) convolutional filters throughout the network architecture. It has different versions with varying depths, such as VGG16 and VGG19.	Simplicity and uniformity of the architecture, achieved excellent performance on various image recognition benchmarks.	2014	2D
ResNet	ResNet is a deep CNN model that introduces the concept of residual connections. It addresses the degradation problem encountered with very deep networks by allowing direct information flow from earlier layers to later ones.	Exceptional performance on ultra-deep architectures, prevents the vanishing gradient problem, easy to train.	2015	2D

InceptionNet	InceptionNet (also known as GoogLeNet) is a deep CNN model that features the Inception module, which employs parallel convolutions of different sizes and concatenates their outputs. It reduces computational complexity while maintaining accuracy.	Improved efficiency by reducing the number of parameters, better utilization of computational resources.	2014	2D
Xception	Xception is an extension of InceptionNet that replaces regular convolutions with depthwise separable convolutions. It separates spatial feature extraction and cross-channel information, resulting in a more efficient and accurate model.	Improved performance compared to InceptionNet, reduced computation costs, better generalization.	2016	2D
EfficientNet	This is a convolutional neural network developed by Google AI. It uses a compound coefficient to uniformly scale each dimension of the depth/width/resolution of the network, which enables it to perform better than other networks given the same computational resources.	Provides higher accuracy while requiring less computational resources compared to other models.	2019	3D
Vision Transformer	This model, developed by Google Research, uses Transformer models, traditionally used in NLP tasks, for image recognition tasks. They treat images as a sequence of patches and apply self-attention to this sequence.	This model, developed by Google Research, uses Transformer models, traditionally used in NLP tasks, for image recognition tasks. They treat images as a sequence of patches and apply self-attention to this sequence.	2020	3D
NASNet	It is a Neural Architecture Search Network developed by Google Brain. The Neural Architecture Search (NAS) algorithm for NASNet uses reinforcement learning to find the most efficient network architecture.	Achieves state-of-the-art accuracy on ImageNet while maintaining a smaller computational footprint.	2018	3D
BiT	Introduced by Google Research, BiT is a transfer learning technique for visual tasks. The strategy is to first pretrain models on a large dataset, then fine-tune them on a smaller, task-specific dataset.	Capable of achieving state-of-the-art results across a wide range of tasks and datasets.	2020	3D
Noisy Student	This is a self-training approach that utilizes a teacher-student training strategy where predictions from the initial trained network (teacher) are used to label and train a larger model (student), which can include additional unlabelled data.	Enables the use of unlabeled data in the training task, leading to good performance on several benchmarks.	2020	3D
C3D	C3D (Convolutional 3D) is a deep 3D CNN model specifically designed for video analysis. It captures both spatial and temporal information by performing 3D convolutions on spatiotemporal video volumes.	Excellent performance on video-based action recognition tasks, effectively captures motion patterns and temporal dynamics.	2014	3D

P3D	P3D (Pseudo-3D) is a variant of 3D CNN that adapts 2D CNN models to the 3D domain without much modification. It achieves this by using 2D pre-trained networks and transforming them into pseudo-3D networks.	Takes advantage of pre-trained 2D models for transfer learning, efficient computation compared to full 3D CNNs.	2017	3D
I3D	I3D (Inflated 3D) is another influential 3D CNN model that inflates 2D CNNs into 3D space by replicating the 2D filters across the temporal dimension. It can utilize pre-trained 2D networks and learn spatiotemporal features.	Incorporates both appearance and motion information, achieves competitive performance on various action recognition benchmarks.	2017	3D
R(2+1)D	R(2+1)D is an extension of 3D CNNs that disentangles spatial and temporal convolutions. It performs 2D spatial convolutions followed by 1D temporal convolutions, reducing computational complexity while maintaining accuracy.	Efficiently models spatiotemporal relationships, balances computational cost, and achieves state-of-the-art performance on action recognition tasks.	2018	3D
Slow Fast	SlowFast is a two-stream 3D CNN model that processes video frames both slowly and rapidly. It exploits the dual temporal resolutions to capture both appearance and motion information, achieving better performance.	Highly effective in modeling videos with spatial and temporal cues, provides better accuracy and robustness in action recognition.	2019	3D
PointNet	A network that directly consumes point clouds and generates rich per-point features by independently applying the learnt input map to each point.	Able to preserve the permutation invariance of points in the input representation.	2017	3D
3D-CNN	This is an extension of traditional 2D-CNNs to handle 3D volumetric data. It is used in tasks like video content analysis, medical imaging, and 3D object recognition.	Maintains spatial correlation between data in 3 dimensions.	2016	3D
PointNet++	This is an extension of PointNet which captures local features from neighboring points.	Effective in classification and semantic segmentation.	2017	3D
VoxNet	It is a basic 3D CNN-based model that applies convolution operations on voxelized objects.	It is a basic 3D CNN-based model that applies convolution operations on voxelized objects.	2016	3D
VoteNet	It is a deep learning framework based on PointNet that uses voting to predict 3D object bounding boxes directly from point clouds.	Can accurately predict the layout of 3D objects from raw point clouds without the need to generate proposals.	2019	3D

2D Classification Models:

1. BASIC-L
    - Paper: Symbolic Discovery of Optimization Algorithms
    - GitHub: BASIC-L
2. CoCa:
    - Paper: Contrastive Captioners are Image-Text Foundation Model
    - GitHub: CoCa
3. ViT-L/16:
    - Paper: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale
    - GitHub: ViT-L/16
4. AlexNet:
    - Paper: ImageNet Classification with Deep Convolutional Neural Networks
    - GitHub: AlexNet
5. VGGNet:
    - Paper: Very Deep Convolutional Networks for Large-Scale Image Recognition
    - GitHub: VGGNet
6. ResNet:
    - Paper: Deep Residual Learning for Image Recognition
    - GitHub: ResNet
7. InceptionNet:
    - Paper: Going Deeper with Convolutions
    - GitHub: InceptionNet

8. Xception:
    - Paper: Xception: Deep Learning with Depthwise Separable Convolutions
    - GitHub: Xception
9. EfficientNet
    - Paper: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks (2019)
    - GitHub: EfficientNet GitHub Repository
10. Vision Transformer
    - Paper: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (2020)
    - GitHub: Vision Transformer GitHub Repository
11. NASNet
    - Paper: Learning Transferable Architectures for Scalable Image Recognition (2018)
    - GitHub: NASNet GitHub Repository
12. BiT (Big Transfer)
    - Paper: Big Transfer (BiT): General Visual Representation Learning (2020)
    - GitHub: BiT GitHub Repository
13. Noisy Student
    - Paper: Self-training with Noisy Student improves ImageNet classification (2020)
    - GitHub: Noisy Student GitHub Repository

3D Classification Models:

1. C3D:
    - Paper: Learning Spatiotemporal Features with 3D Convolutional Networks
    - GitHub: C3D
2. P3D:
    - Paper: Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks
    - GitHub: P3D
3. I3D:

- Paper: Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset
- GitHub: I3D

4. R(2+1)D:

- Paper: A Closer Look at Spatiotemporal Convolutions for Action Recognition
- GitHub: R(2+1)D

5. SlowFast:

- Paper: SlowFast Networks for Video Recognition
- GitHub: SlowFast

6. PointNet

- Paper: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation (2017)
- GitHub: PointNet GitHub Repository

7. 3D-CNN

- Paper: Voxnet: A 3d Convolutional Neural Network for real-time object recognition (2016)
- GitHub: 3D-CNN GitHub Repository

8. PointNet++

- Paper: PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space (2017)
- GitHub: PointNet++ GitHub Repository

9. VoxNet

- Paper: Voxnet: A 3d Convolutional Neural Network for real-time object recognition (2016)
- GitHub: VoxNet GitHub Repository

10. VoteNet

- Paper: Deep Hough Voting for 3D Object Detection in Point Clouds (2019)
- GitHub: VoteNet GitHub Repository