Evolution of Image Classification

Model Architecture

Sakura

2025/06/12
bili_sakura@zju.edu.cn

Image Classification: Overview

Overview of image classification.

Image source: Hugging Face tutorial

Background: AlexNet Architecture

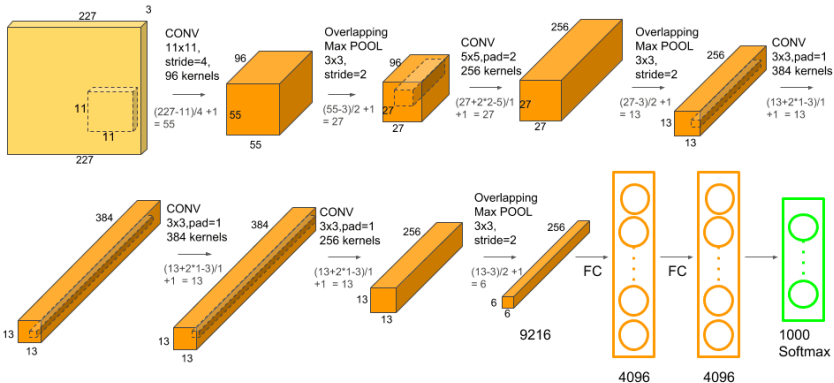

Figure: Architecture of AlexNet (Krizhevsky, Sutskever, and Hinton, 2012). *Image source: Web*

Krizhevsky, et al. ImageNet Classification with Deep Convolutional Neural Networks, NeurIPS, 2012.

AlexNet: ILSVRC-2012 Results

- AlexNet (Krizhevsky, Sutskever, and Hinton, 2012) achieved a winning top-5 test error rate of **15.3%** in the ILSVRC-2012 competition, compared to 26.2% by the second-best entry.

Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
SIFT + FVs	—	—	26.2%
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	**16.4%**
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	**15.3%**

Table: Comparison of error rates on ILSVRC-2012 validation and test sets. *Italics*: best results by others. *: Models pre-trained on ImageNet 2011 Fall release.

Krizhevsky, et al. ImageNet Classification with Deep Convolutional Neural Networks, NeurIPS, 2012.

Background: Image Classification with Deep Learning

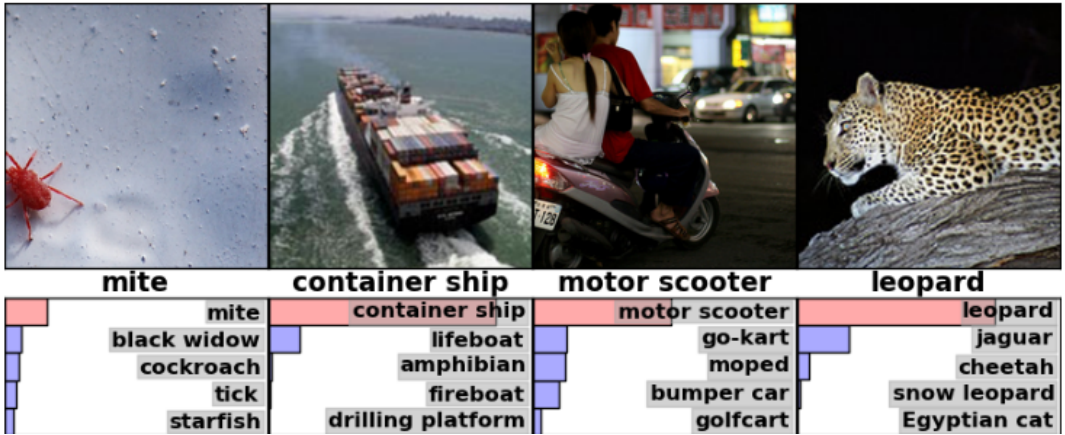

Figure: AlexNet on ILSVRC-2010 (Berg, Deng, and Fei-Fei, 2010).

Berg, et al. Large Scale Visual Recognition Challenge 2010, 2010.

Background: ResNet (2016)

- Key innovation: residual (skip) connections.
- Enabled extremely deep networks (up to 152 layers).
- Achieved state-of-the-art performance in ILSVRC-2015.

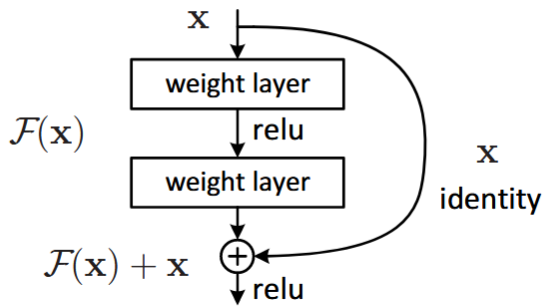

Figure: ResNet block with identity mapping (He et al., 2016).

Transformers for Image Classification

- **Vision Transformer (ViT, 2021):** Applies transformer models from NLP to image patches.

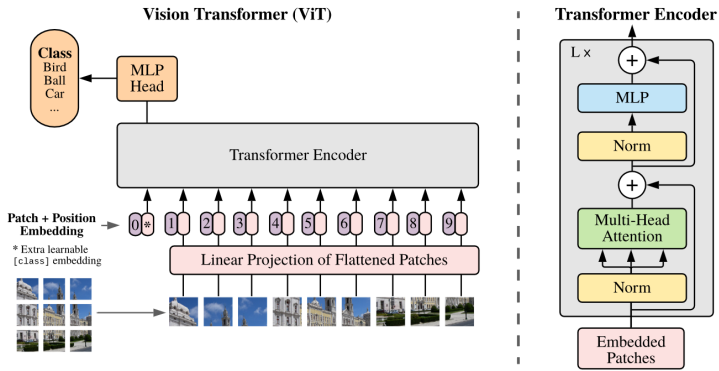

Figure: Vision Transformer overview (Dosovitskiy et al., 2021).