

Introduction

1. Traditional Machine Learning and Prediction

Hand-crafted features, poor scalability.

2. Deep Learning from Scratch and Prediction

Complicated feature engineering, slow convergence of DNN, laborious collection of large-scale, task-specific, and crowd-labelled data.

3. Supervised Pre-training, Fine-tuning and Prediction

Accelerate network convergence, well-performing models with limited task-specific training data, requires large-scale labelled data in pre-training.

4. Unsupervised Pre-training, Fine-tuning & Prediction

Self-supervised learning to learn useful and transferable representations from unlabelled data, e.g., masked image modelling, contrastive learning; still requires a fine-tuning stage.

5. VLM Pre-training and Zero-shot Prediction

Motivated by great success in NLP; matches image-text embeddings; leverages large-scale image-text pairs available online.