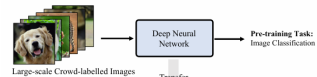


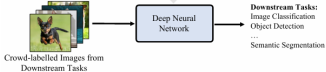
Introduction

(a). Supervised Pre-training, Fine-tuning and Prediction

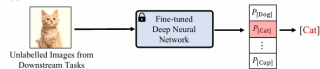
(1) Supervised Pre-training



(2) Task-specific Fine-tuning

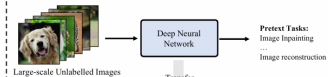


(3) Prediction

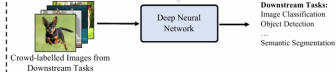


(b). Unsupervised Pre-training, Fine-tuning and Prediction

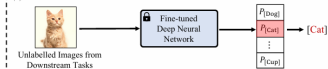
(1) Unsupervised Pre-training



(2) Task-specific Fine-tuning

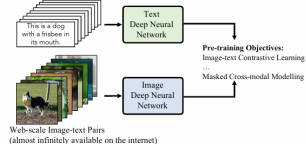


(3) Prediction



(c). Vision-Language Model Pre-training and Zero-shot Prediction

(1) Vision-Language Model Pre-training



(2) Zero-shot Prediction without Fine-tuning

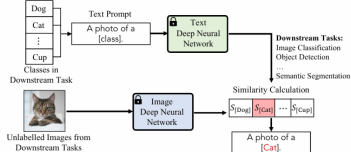


Figure 4: Three DNN training paradigms in visual recognition.