Multi-task learning of objects and parts – Initial Problem Description, Solution and Plan

Hailin Weng

Introduction

This project aims to develop and test a computational model of vision (Convolutional Neural Network model) that is simultaneously trained on object name labels and part labels.

Problem Description

In the realm of computer vision, a primary objective is enabling machines to emulate human vision by interpreting and making decisions based on visual data. Over the years, deep learning methods have revolutionized this field, notably in tasks such as image classification. The ImageNet dataset has been crucial in this evolution, acting as a benchmark for models to recognize and categorize diverse image content. Typically, advanced models like the deep convolutional neural networks (e.g., AlexNet [1], VGG[2]) are trained using labels like EAGLE, CAR, etc. However, a new dimension emerges when considering object properties over labels, coined as property norms.

Who has the problem?

Researchers, developers, and businesses trying to further understand image content beyond just object labels.

What is the problem?

While models like AlexNet excel at categorizing images using predefined labels, they don't capture the intrinsic attributes or parts of objects. This object-centric perspective lacks a deeper, attribute-centric view that considers shared properties across different object categories.

When does the problem occur?

Whenever deep learning models are trained solely on object labels, missing out on the deeper intricate understanding provided by property norms.

Where does the problem manifest itself?

In applications where understanding the properties of objects is crucial, such as advanced image search engines, augmented reality, or robotics.

Historically, the emphasis has been majorly on label-centric training. AlexNet, for instance, was groundbreaking for its time, yet focused on this paradigm [1]. The problem emerges when we need machines to identify shared properties between objects, like recognizing both EAGLES and WOLVES have "EYES". Existing models lack this granularity.

Why other solutions do not work?

Previous solutions, like the influential AlexNet, focused on raw categorization performance. Such models might misclassify or fail to recognize objects in unfamiliar contexts where shared properties might be more indicative than holistic labels.

How might your work be innovative or helpful?

This project seeks to pivot from the traditional label-centric training to property norms. By training a model on these property norms using the CSLB specification [3], we intend to provide a richer, more nuanced understanding of images. It becomes imperative to ask: Can a model trained this way outperform or complement traditional models? And, can such a model identify parts of an object category it hasn't seen?

Research Questions:

1. How does a property norm-trained model's performance compare with one trained solely on object labels?
2. How do internal representations differ in both models?
3. Can a property norm-trained model correctly identify object parts it hasn't been trained on?

Goals and requirements

- Become familiar with the Pytorch framework for building, training and evaluating neural networks.
- Become familiar with influential deep convolutional neural network models for vision (e.g. AlexNet; VGG, ResNet).
- Extend and evaluate a deep convolutional neural network model for vision to predict semantic properties of objects, as well as labels.
- Analyse the internal representations of the model and compare them to the internal representations in a model trained on the labelling task alone.

Success Criteria

1. Implementation: Implement AlexNet using PyTorch with comparable accuracy on the ImageNet dataset.
2. Property Norms Extension: Adapt AlexNet to predict properties based on the CSLB norms, achieving a predefined accuracy on a validation set.
3. Comparative Analysis: Compare performance metrics (e.g., top-1 accuracy) between the traditional and property norms-based models.
4. Representation Insight: Analyze internal representations of both models to understand differences in feature learning.
5. Generalization: Demonstrate the property norms model's ability to identify untrained object parts.
6. Documentation: Offer clear documentation for model replication and evaluation.
7. Robustness: Validate models against varied real-world scenarios and datasets.

Expected Project Development Plan

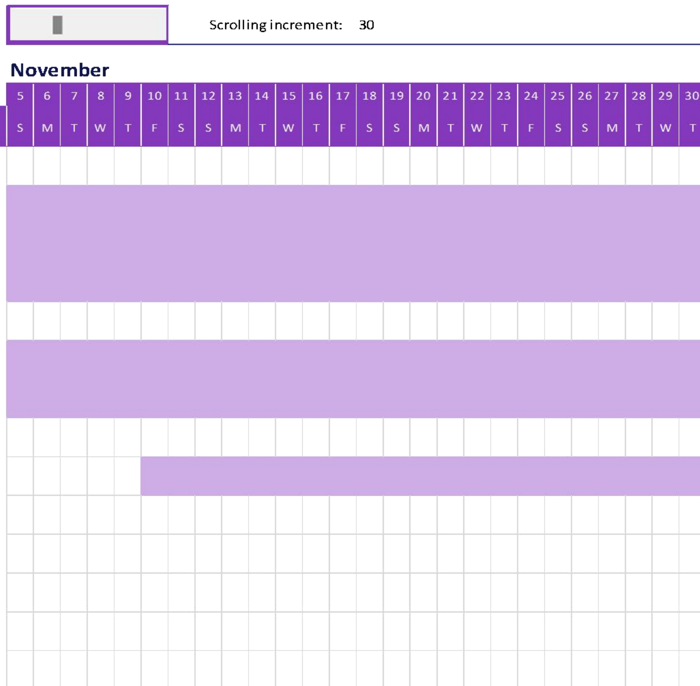
CSC3002: Multi-task learning of objects and parts

Student: Hailin Weng

Supervisor: Barry Devereux

Project start date: 06/10/2023

Milestone marker: 1



Milestone description	Assigned to	Progress	Start	Days
Foundation				
Machine Learning		15%	06/10/2023	97
Deep Learning		10%	09/10/2023	94
Computer Vision		8%	11/10/2023	92
Implementation				
AlexNet (Pytorch)		10%	13/10/2023	90
Property Norms		0%	23/10/2023	80
Develop (Solution)				
My own CNN (Project)			09/11/2023	63
Project Description, Solution Approach and			30/12/2024	12
Demo			01/01/2024	10
Dissertation				
Final Dissertation			13/1/2024	91
Demonstration			15/04/2024	

References

[1] Krizhevsky et al. "ImageNet Classification with Deep Convolutional Neural Networks", NIPS, 2012.

[2] Simonyan & Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv, 2014.

[3] Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). "The Centre for Speech, Language and the Brain (CSLB) concept property norms." Behavior Research Methods, 46(4), 1119-1127.