CENTER FOR BIOSYSTEMS AND BIOTECH DATA SCIENCE
Esla Timothy Anzaku, Arnout Van Messem, Wesley De Neve

IF AN IMAGE IS WORTH A THOUSAND WORDS, WHY DESCRIBE IT WITH A SINGLE WORD?

Image Multi-classification Task

- ➔ Cornerstone task in deep learning for modern computer vision
- ➔ Training AI models to assign images into predefined categories
- ➔ Each image is assigned only to a single defined category

ImageNet 1k Dataset [1]

- ➔ Pivotal dataset fueling AI research and development in computer vision
- ➔ Million-plus images, 1000 categories
- ➔ Spans categories from ‘dogs’ and ‘plants’ to ‘building’ and ‘vehicles’
- ➔ Serves a multitude of purposes

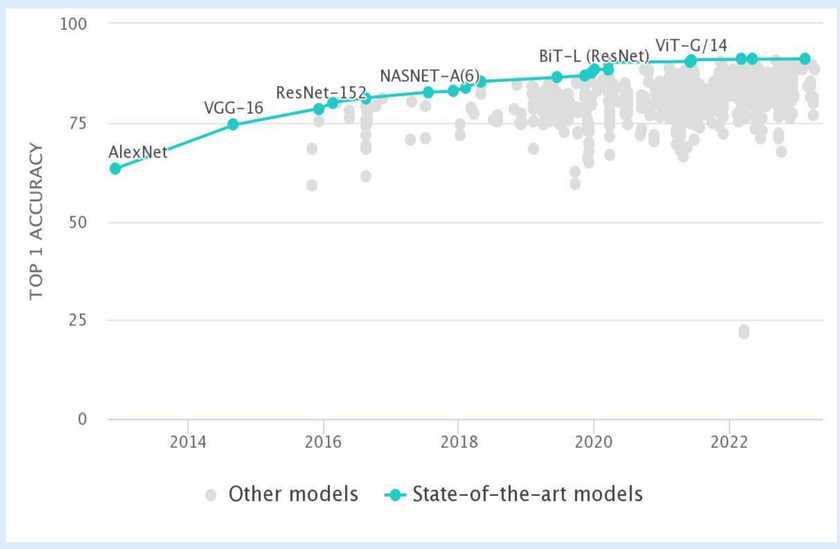
Example ImageNet Images

Example ImageNet-based Applications

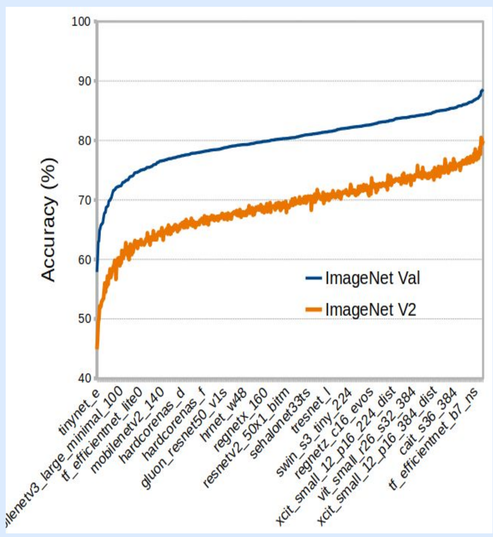
Benchmarking deep learning progress in supervised computer vision	**Benchmarking progress in self-supervised deep learning for computer vision**
Feature extraction for downstream tasks, such as object detection and segmentation	**Fine-tuning models on smaller datasets**

Challenges

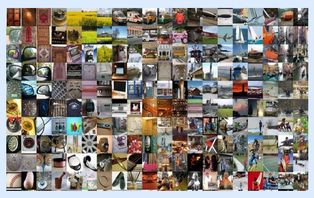
Performance saturation regardless of model architecture, training technique, dataset, and model size [2]



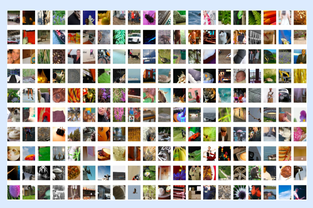
Performance degradation on similar datasets (591 models)



ImageNet Validation (50,000 Images)

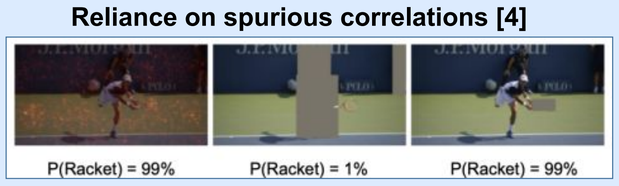


ImageNet V2 [3] (10,000 Images)



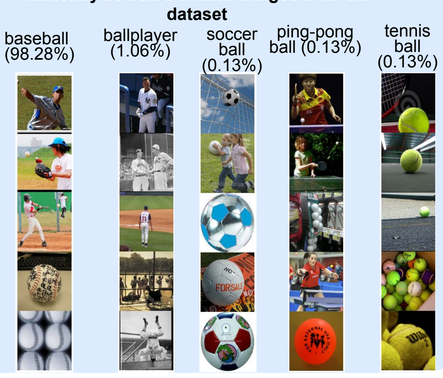
Trustworthiness: Can very confident predictions be wrong?

Reliance on spurious correlations [4]



Original Image Label given by annotators: Baseball Player
Our curiosity: Is the ball the most obvious object?

Top-5 predicts from a ResNet50[5] model and randomly selected similar images from the dataset



Hypothesis

Substantial improvements in ImageNet-based model utility and performance can be achieved by effectively leveraging the dataset’s multi-label nature.

Can leveraging multi-labels help in improving...

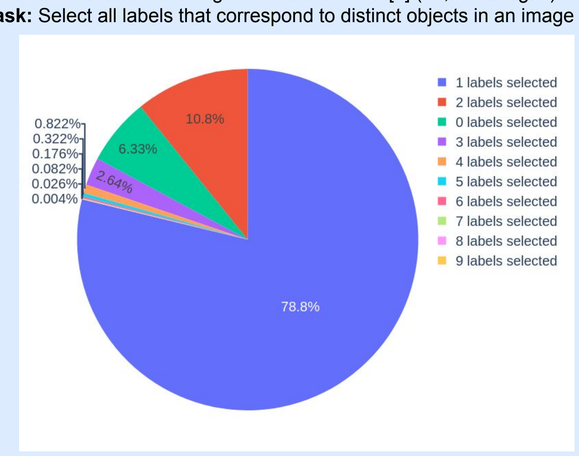
- **➔downstream performance of ImageNet pre-trained models for fine-tuning on smaller datasets, object detection, and object classification (a/o)?**
- **➔the evaluation of model predictive uncertainty?**
- **➔deep learning benchmarking in computer vision?**
- **➔the robustness and trustworthiness of computer vision AI models?**

Research Motivation

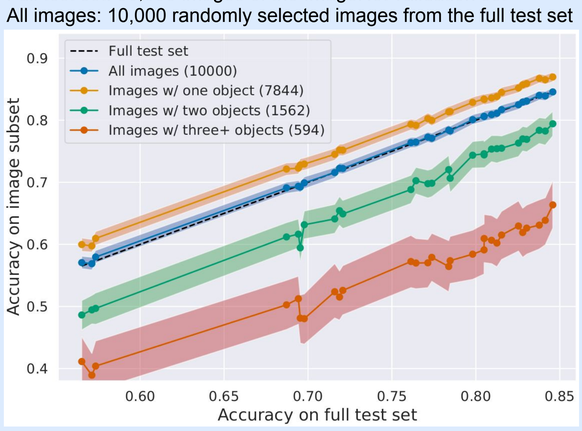
Quantitative Insights

Images with multiple labels are important, thus having impact

Reassessed labels for ImageNet validation set [6] (50,000 images)
Task: Select all labels that correspond to distinct objects in an image



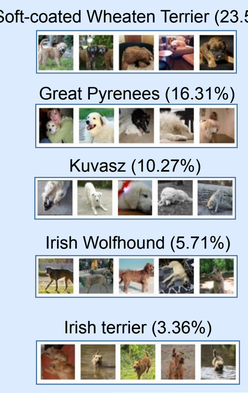
Effect of multi-labels on top-1 accuracy [7]
Five annotators carefully re-labeled the ImageNet validation set
Summary: Accuracy drops by roughly 10% across all models
Full test set: 50,000 images of the ImageNet validation set
All images: 10,000 randomly selected images from the full test set



Qualitative Insights

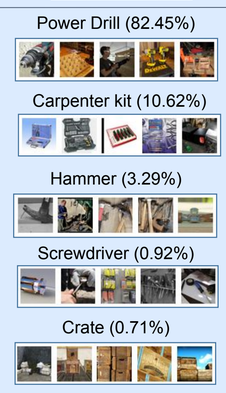
Original Image
Label given by annotators: Soft coated wheaten terrier
Our curiosity: Is this really one dog? Can we confidently say which dog it is?

Top-5 predictions (by a ResNet50 model) and randomly selected images similar to the predicted classes from the dataset
Our curiosity: Aren’t the model’s predictions reasonably better?



Original Image
Label given by annotators: Hammer
Our curiosity: Is it fair to describe this image using only the word “Hammer?”

Top-5 predictions (by a ResNet50 model [7]) and randomly selected images similar to the predicted classes from the dataset
Our curiosity: Aren’t the model’s predictions reasonably better? Top-1 accuracy will penalize this as a false prediction. Is it?



Conclusions

- **Within the field of computer vision, the ImageNet dataset has substantially propelled advancement in deep learning**
- **With models nearing peak performance on this dataset, it is imperative to evaluate the dataset’s limitations and consider the resulting implications on subsequent tasks that utilize these models trained on it**
- **Preliminary evidence suggests that embracing the multi-label nature of the ImageNet dataset could further enhance its utility and efficacy**
- **Our ongoing research efforts are dedicated to exploring this relatively understudied multi-label problem within the ImageNet dataset, with the aim of realizing its full potential**

[1] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Dataset (2009).
[2] https://paperswithcode.com/sota/image-classification-on-imagenet
[3] B. Recht, R. Roelofs, L. Schmidt, V. Shankar, Do ImageNet Classifiers Generalize to ImageNet? (2019).
[4] G. Plumb, M. T. Ribeiro, A. Talwakar, Finding and Fixing Spurious Patterns with Explanations (2022).
[5] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition (2015).
[6] L. Beyer, O. J. Henaff, A. Kolesnikov, X. Zhai, A. van den Oord, Are We Done With ImageNet? (2020).
[7] D. Tsipras, S. Santurkar, L. Engstrom, A. Ilyas, A. Madry, From ImageNet to ImageNet Classification: Contextualizing Progress on Benchmarks (2020).

GHENT UNIVERSITY GLOBAL CAMPUS
Center for Biosystems & Biotech Data Science
LIÈGE université Mathématique
FACULTY OF ENGINEERING AND ARCHITECTURE

For more inquiries and discussion, please contact me at
eslatimothy.anzaku@ghent.ac.kr