Machine Learning with few data

Marcus Liwicki, Prakash Chandra Chhipa
Luleå University of Technology

General LTU.AI Mission

We aim at creating a **strong** and **active ecosystem in applied AI**, an ecosystem that directly connects fundamental research with real-life applications and demonstrations of AI in the industrial sector and beyond, and by that contribute to a safe and measurable strong impact of **AI innovations** in **everyday life**.

Goal: **sustainable flagship** in **applied AI** on national and EU level

Loading….LTU.AI Digital Innovation Hub

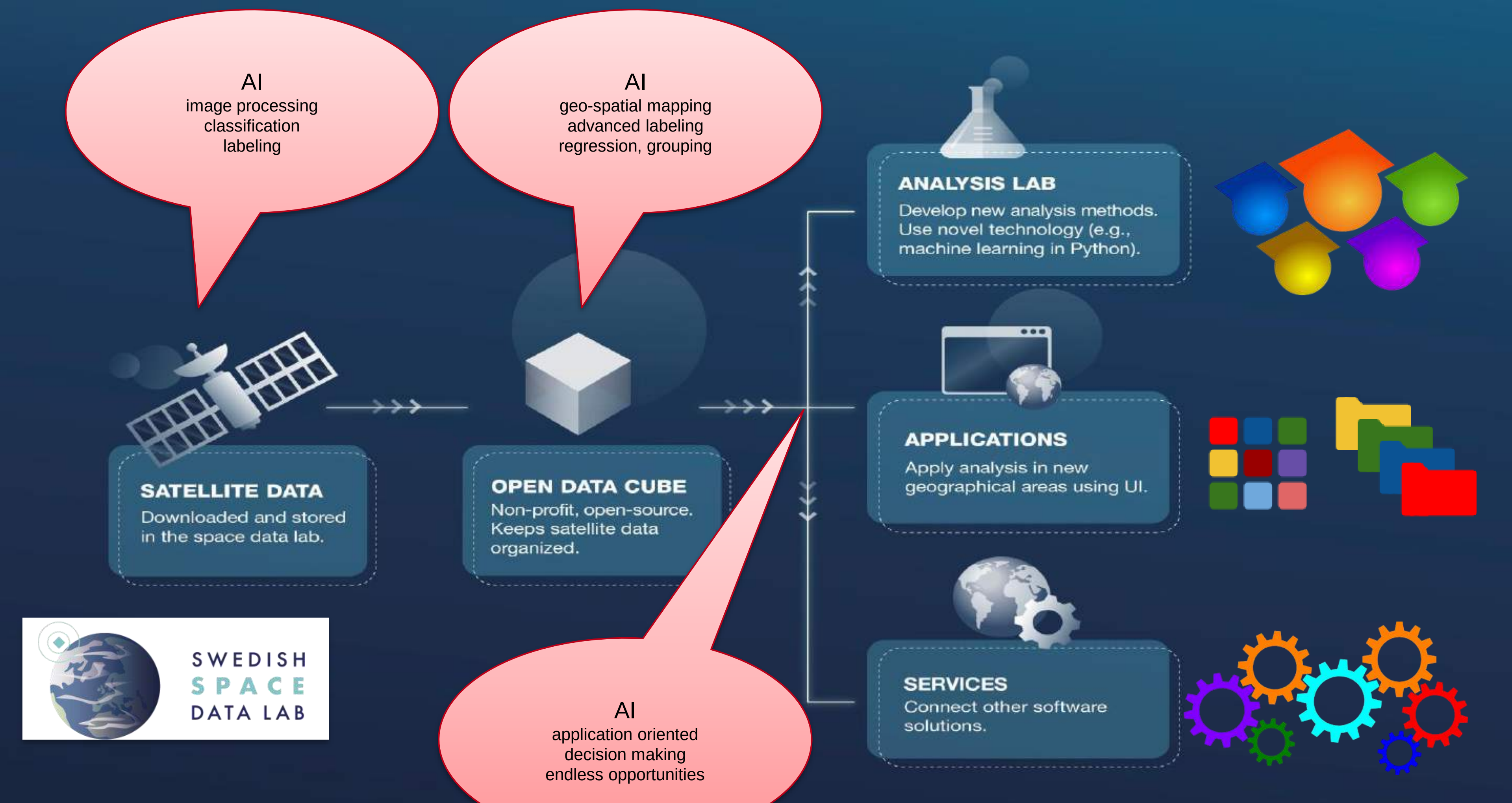
Join our efforts to build together the future of LTU.AI

LULEÅ UNIVERSITY OF TECHNOLOGY

Research @ LTU Machine Learning

- Fundamental: Understanding of Neural Networks & Human Mind
- Application-oriented: Welfare of the Society (education; eHealth; automation; energy-efficiency; agriculture, geology, space data); Reading Systems (NLP+documents); Social Media Analysis, Process-Industry
- Applications: STEAM-education; IT-Security; historical document analysis;

LULEÅ UNIVERSITY OF TECHNOLOGY



Candidate for Swedish European DIH

SWEDISH SPACE DATA LAB

About Project

Pilot project 1 Drought in Mälardalen

Pilot project 2 Water levels Vänern

Pilot project 3 Marine Habitats

Pilot project 4 Coastal Zones

Pilot project 5 AI for Good

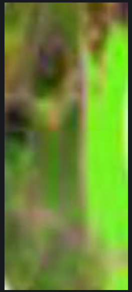
Pilot project 6 API

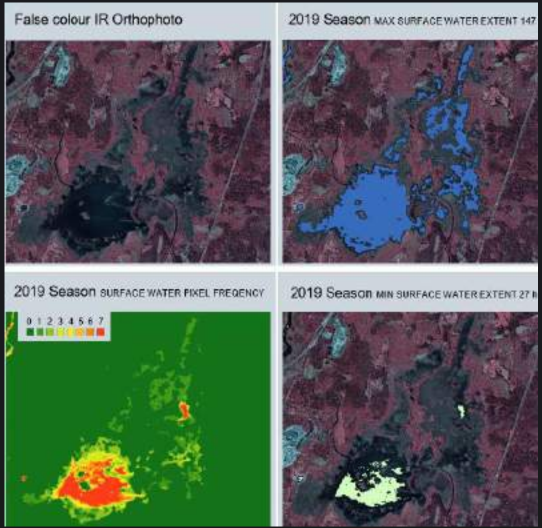
Pilot project 7 Where and when is the Water

Pilot project 8 AI Areal

Pilot candiate 9 AI Shallow Water

Pilot candidate 10 "Grunda bottnar"



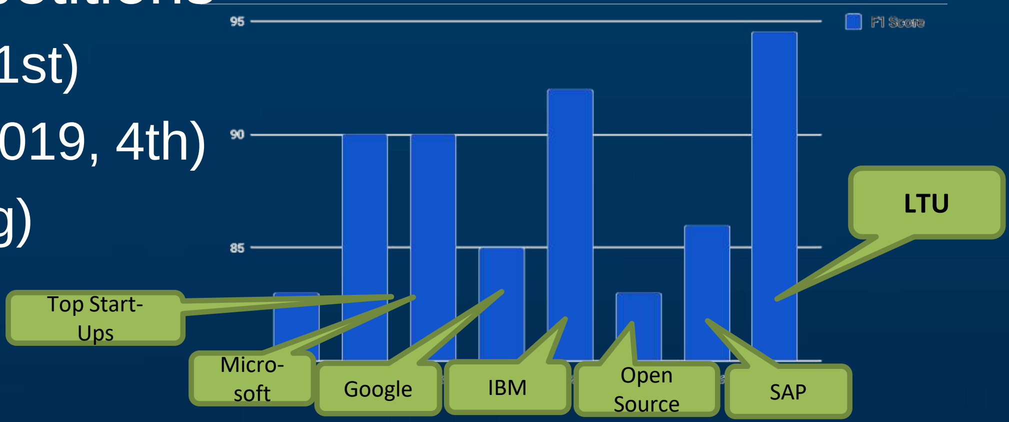


https://rymddatalabbet.se/projects/

PREDICTIVE MOVEMENT VINNER INNOVATIONSTÄVLING

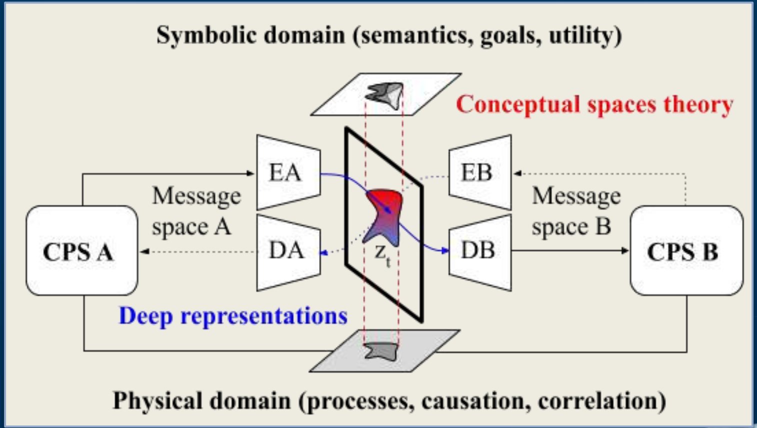
Strength in NLP and Chatbots (10+ publications)

- In the top of several competitions
  - Intent classification (2018, 1st)
  - Hate speech recognition (2019, 4th)
  - Idioms recognition (ongoing)
- Chatbots and application
  - Vinnova Språkdatalabb
  - LTU service bot
  - Ongoing LTU Model chatbot → Swedish AI chatbot assistant
  - ChatPal for mental health (and similar medical chatbots)



Industrial Applications

Arrowhead Tools



Vehicle part recognition

Music

Teaching

- Courses on Machine Learning and Applied AI
- Regular freestanding course (for anyone from EU)
  - Introduction to AI (500+ participants)
  - Advanced Data Mining (100 participants)
  - Natural Language Processing (planned with Örebro U.)
- Customizable lifelong education on AI for Industry
- 2-year international master program Applied AI
- First Applied AI program in Sweden (starting 2021)

LULEÅ UNIVERSITY OF TECHNOLOGY

Equity in the Machine Learning Group!


Marcus


Pedro


Konstantina


Gustav


Fotini


Fredrik


Christian


Priyamvada


György


Saleha


Rajkumar


Oluwatosin


Homam


Mattias


Sana


Ali


András


Richa


Karl


Prakash


Maryam


Jacob

Notice something? Almost 40% **woman**

Machine Learning for the welfare of society

LULEÅ UNIVERSITY OF TECHNOLOGY

Machine Learning with few data

Self-supervised Method

Prakash Chandra Chhipa, Doctoral Student, Machine Learning Group

Supervisor: Prof. Marcus Liwicki

LULEÅ UNIVERSITY OF TECHNOLOGY

Outline

1. Research Interest – Intro & Motivation
2. Goal
3. Prior
4. Current Approaches
5. End to End Learning - Transfer Learning
6. End to End Learning - Clustering
7. Focus – Contrastive Learning (CL)
8. State of the Arts in Contrastive Learning
9. Comparative Summary on SOTA
10. Remarks on Contrastive Learning
11. Research Questions
12. Q&A

LULEÅ UNIVERSITY OF TECHNOLOGY

1. Research Interest – Intro & Motivation

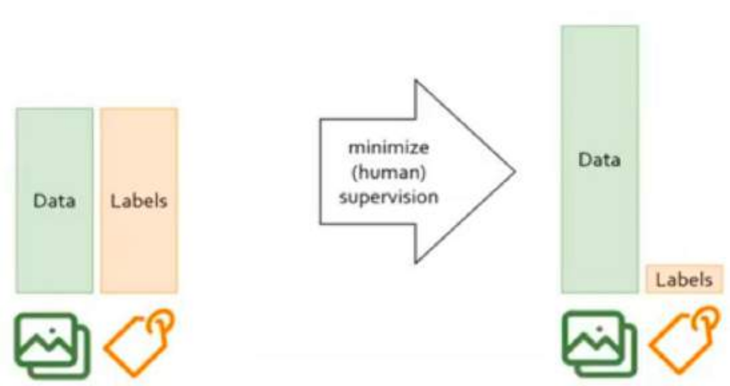
<u>Learning with fewer labelled examples</u>

- How to decrease the requirement of labeled data in supervised training without losing on performance?
  - ✓ With the help of related but unlabeled raw data

- <u>Motivation</u>: Potential to push Applied AI objective for industrial applications
  - Less data annotation means less cost <time & money>
  - Availability of huge amount of raw and unlabeled data <As part of biz process>

LULEÅ UNIVERSITY OF TECHNOLOGY

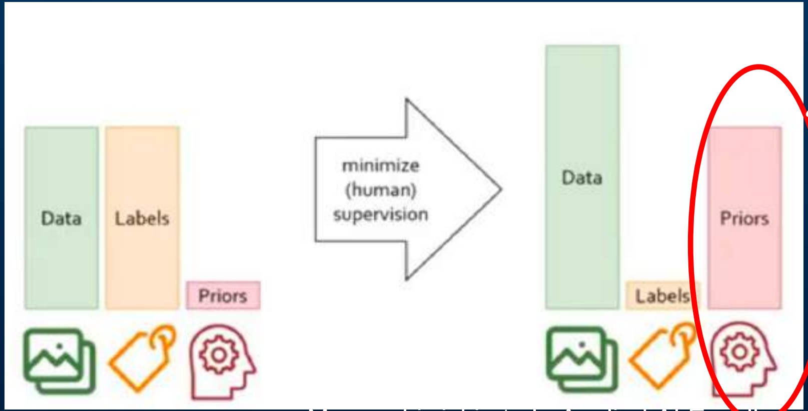
2. Research Interest – Goal



(Ideal)

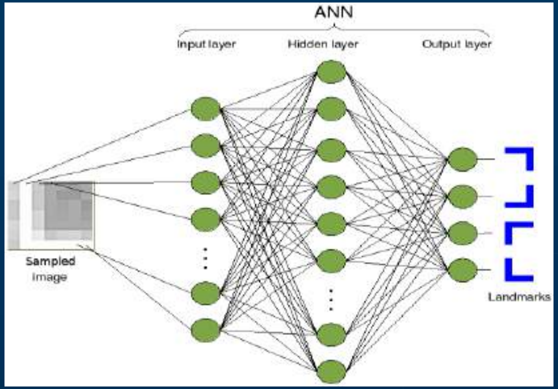
Focus & efforts

(Reality)



3. Prior

- In Terms of Experience
  - Previously trained machine learning model on similar data
    - Proven architectures



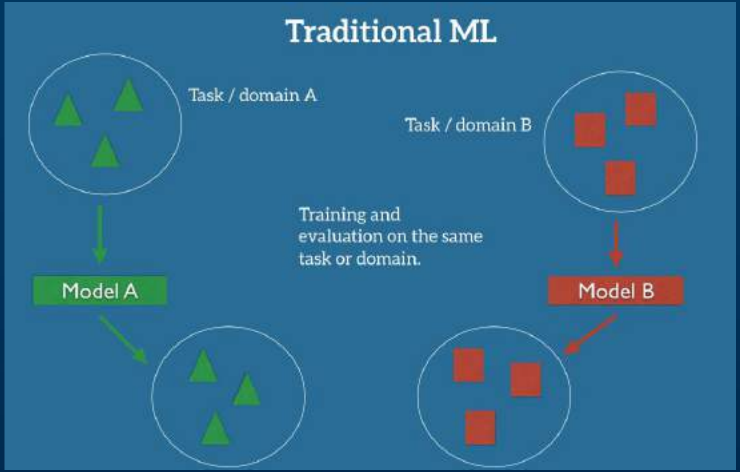
- In Terms of Knowledge
  - Human Reasoning
    - Correlating the given input details and identifying discriminative features

4. Current Approaches

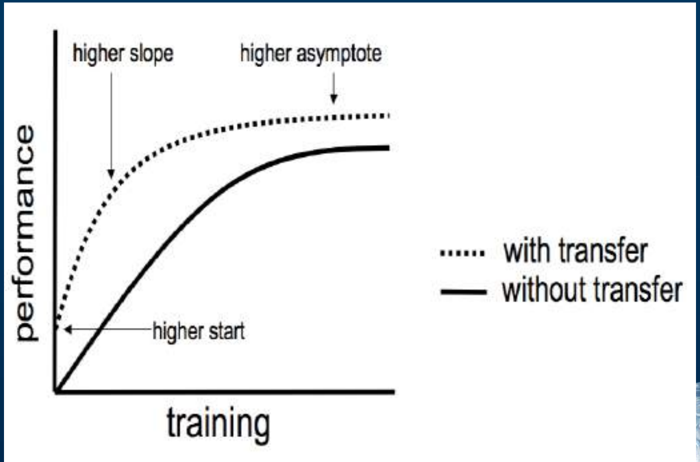
- End to End Learning
  - Transfer Learning *(A Survey on Deep Transfer Learning - 2018)*
    - Utilizing pretrained models and finetuning on application specific data
    - Required less data to fine tune than training it from scratch
  - Clustering – *(Deep Clustering for Unsupervised Learning of Visual Features - 2018)*
    - Labelled data not required
- Representation Learning
  - Contrastive Learning *(SimCLR - July 2020, SwAV – October 2020)*
    - Pretraining mechanism which utilizes application specific unlabeled data
    - Also compute intensive but possibility to scale down

5. End to End Learning - Transfer Learning

- <u>Supervised Method</u>



Source: https://ruder.io/transfer-learning/



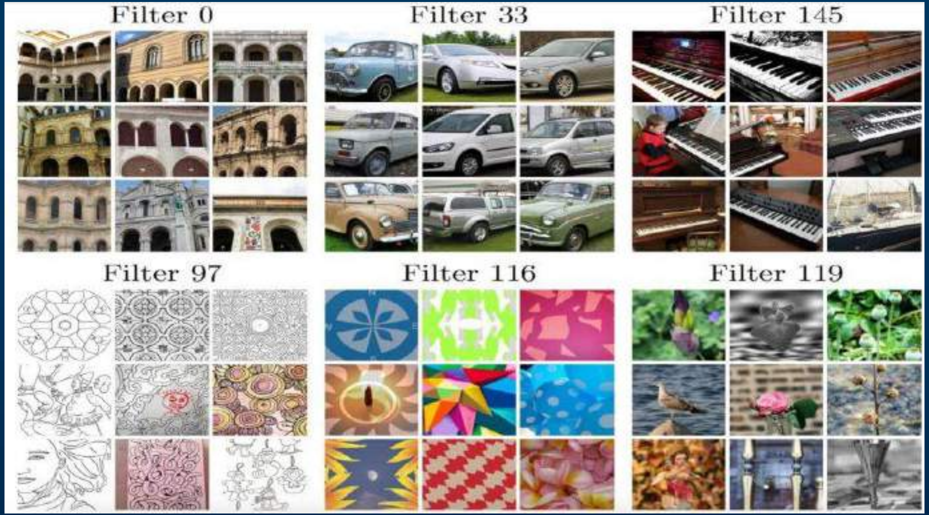
Source: https://machinelearningmastery.com/transfer-learning-for-deep-learning/

- <u>Remarks</u>
  - Successful but only initial layers with low-level features are common & useful across applications
  - No possibility for unlabeled data

LULEÅ UNIVERSITY OF TECHNOLOGY

6. End to End Learning - Clustering

- Un-Supervised Method



Source: https://neurohive.io/en/state-of-the-art/deep-clustering-approach/

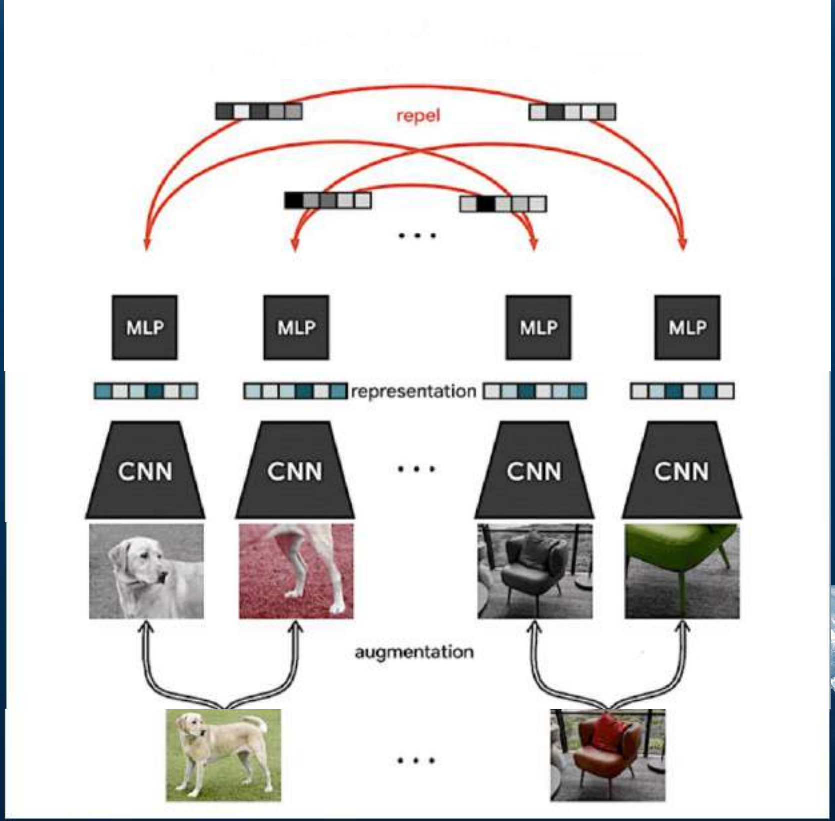
- Remarks
  - Compute intensive when applied on images
  - Non robust feature representation when feature extracted with pretrained models

LULEÅ UNIVERSITY OF TECHNOLOGY

7. Focus – Contrastive Learning (CL)

- Self-Supervised Method: Allows model to learn generic representations on unlabeled data through similarity over pair of input image and its augmented version.

- Method:
  - Learn similarity between augmented representation from same image
  - Learn dissimilarity otherwise



Source: https://ai.googleblog.com/2020/04/advancing-self-supervised-and-semi.html

8. State of the Arts in Contrastive Learning

4.1 Simple Framework for Contrastive Learning (SimCLR)

4.1.1 A Simple Framework for Contrastive Learning of Visual Representations (SimCLR v1), ICML - 2020

4.1.2 Big Self-Supervised Models are Strong Semi-Supervised Learners (SimCLR v2), NeurIPS – 2020

4.2 Momentum Contrast Learning (MOCO)

4.2.1 Momentum Contrast for Unsupervised Visual Representation Learning (MOCO v1), CVPR - Mar 2020

4.2.2 Improved Baselines with Momentum Contrastive Learning (MOCO v2), ?? Arxiv Oct- 2020

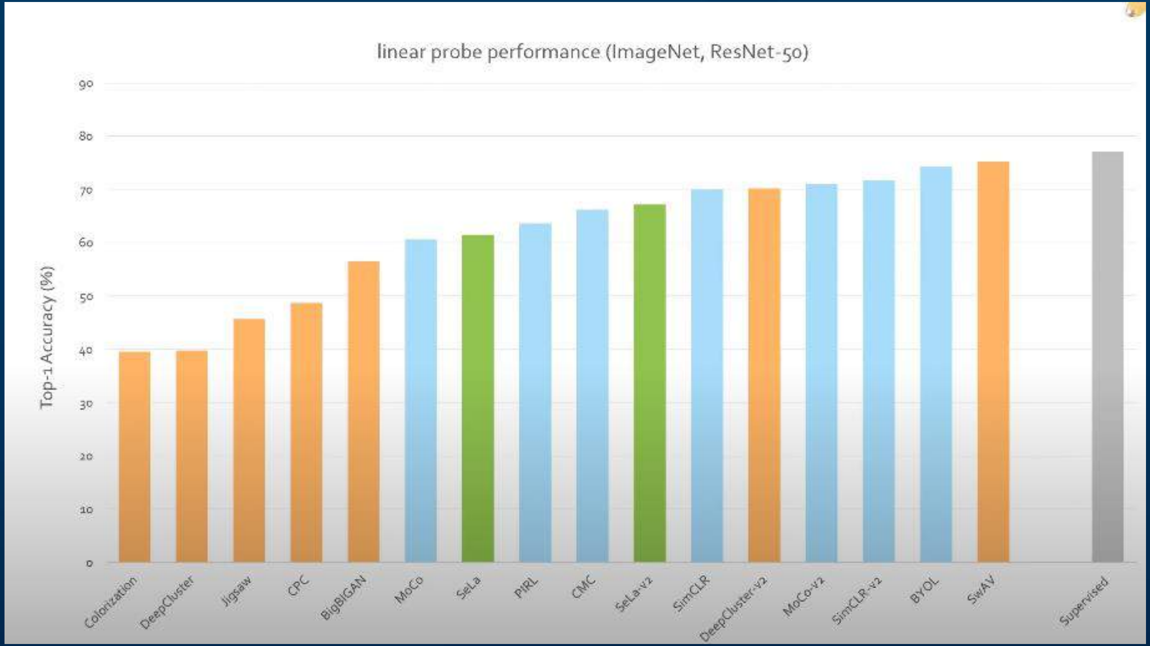
4.2.3 Bootstrap Your Own Latent A New Approach to Self-Supervised Learning, NeurlPS - 2020

4.3 Contrastive Learning with Clustering

4.3.1 Unsupervised Learning of Visual Features by Contrasting Cluster Assignments (SwAE), Arxiv 2020

LULEÅ UNIVERSITY OF TECHNOLOGY

9. Comparative Summary on SOTA



Contrastive Learning

Clustering + Self-supervised

Self-Labelling

Source (IARAI): https://www.youtube.com/watch?v=Bn66HnBxXFM

- Remarks
  - Priors (augmentation mechanism) is more important than learning method
  - Obtains performance approx. equal to supervised methods with 10% labelled data

LULEÅ UNIVERSITY OF TECHNOLOGY

9. Comparative Summary on SOTA …

More Insights

No.	Method	Contrastive Learning Key Factor	Contribution	Limitation
4.1.1	SimCLR V1.0	K1: Similarity learning for positive pairs K2: Dissimilarity learning for negative pairs	Established benchmark performance on unsupervised contrastive learning	1. 'Large batch size' due to positive + negative pair 2. 'Mass gradient computation & backprop issue' due to all (+ve & -ve) pairs
4.1.2	SimCLR V2.0	K1 + K2 on Task agnostic Big n/w which used in distillation for task specific small n/w	+ Added enablement of semi-supervised learning through distillation	Same as SimCLR V1.0 + usage of bigger networks
4.2.1	MOCO V1.0	K1 + K2 over momentum encoder where CL as dynamic dictionary lookup	Revealed unsupervised contrastive learning with smaller batch size and lessor backpropagation of gradients	1. 'Mass gradient computation & backprop issue' due to all (+ve & -ve) *pairs (same as SimCLR because as q-encoder backpropagates)* 2. Overhead of dynamic dictionary queue
4.2.2	MOCO V2.0	MOCO V1.0 + 2-layer MLP projection head	Stronger baseline, outperformed on SimCLR and MOCO v1.0.	1. 'Mass gradient computation & backprop issue' due to all (+ve & -ve) pairs *same as SimCLR because q-encoder and k-encoder both backpropagates* 2. Overhead of dynamic dictionary queue
4.2.3	BYOL	K1+ momentum encoding + two separate networks (online and target)	Achieves self supervised CL without negative pair. Establishes benchmarks in semi-supervised approach. Robust for smaller batch size.	1. Complex pipeline with large number of pruning. Makes it challenging for concept utilization.
4.3.1	SwAE	K1 + Swapped" prediction mechanism + cluster assignment	Achieves self supervised CL without negative pair. Claims state of art in unsupervised image clustering.	1. Relatively complex loss computation due to swapped prediction 2. Additional online cluster assignment swapping

LULEÅ UNIVERSITY OF TECHNOLOGY

10. Remarks on Contrastive Learning

- Currently CL is leading the self-supervision and shows potential push for semi-supervised methods (Industrial Applied AI applications)

- CL in current state is compute intensive (batch size, negative pairs, & gradients) which makes it challenging for direct (as-it-is) applications. Needs (Research Potential) to be tailored for custom and small-scale application requirement.
- Contrastive methods are sensitive to the choice of image augmentations.
- Leveraging utilization of application specific but unlabeled data.

- CL can be benchmarking framework (Different methods for different applications) for semi-supervised and even supervised task.

LULEÅ UNIVERSITY OF TECHNOLOGY

11. Research Questions

Q1: How to devise computationally optimize contrastive learning for application specific context considering semi-supervised scenarios?

Q2: Incorporation of computationally optimize contrastive learning for application specific context considering semi-supervised scenarios?

LULEÅ UNIVERSITY OF TECHNOLOGY

Q&A

LULEÅ
UNIVERSITY
OF TECHNOLOGY