Self-Supervised Learning

The Next Step Toward AI

Presented by
Sonal Kumar
Research Scholar, CSE, IIT Guwahati
Supervised by Prof. Arijit Sur & Dr. Rashmi Dutta Baruah

Technology Innovation Hub
IITG TIDF

Winter School 2025 @ IIT Guwahati
Supported by TIDF IITG
Organased by Multi Media Lab, IITG

What is Learning ?

What is Learning ?

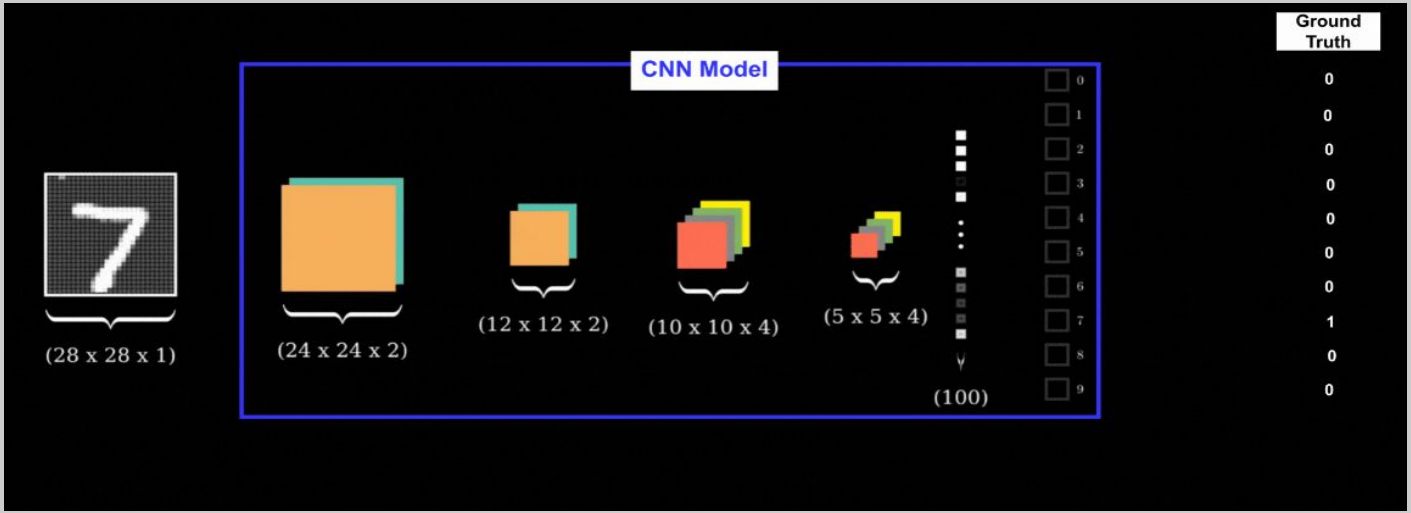
- In General Sense:
  - The process of acquiring knowledge, skills, behaviors, or understanding through **experiences, study, or being taught.**

- From the perspective of artificial intelligence (AI):
  - The process by which an AI system improves its performance on a task over time by **extracting patterns or knowledge from data.**

Can learn with or without the supervision

Supervised Training of a CNN Model

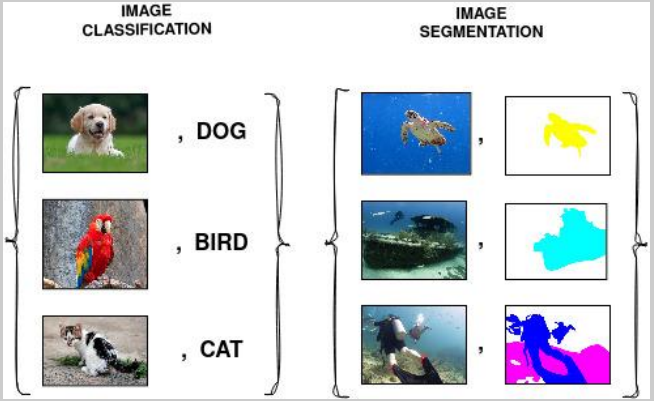
- A deep learning model needs some kind of supervision for efficient training.



Y = WX + B; X (Input data), Y (Labels/Annotations)

What is Supervision ?

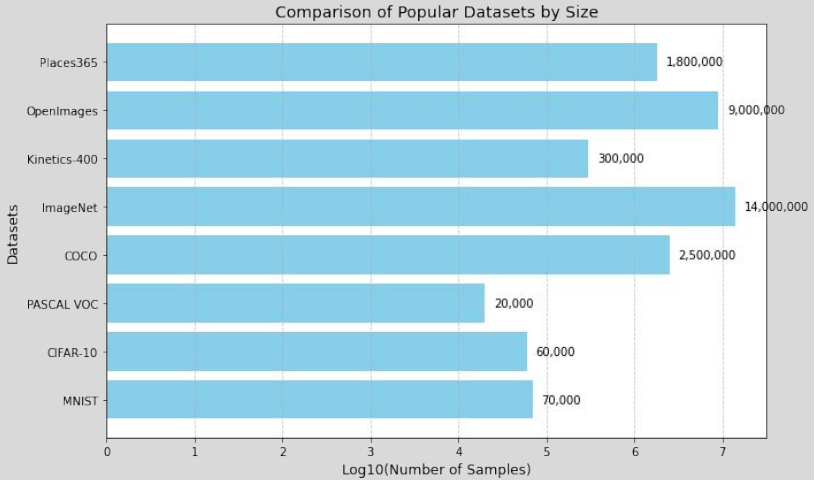
- IMAGE CLASSIFICATION:
  - IMAGE AND CLASS LABEL PAIR
- SEMANTIC SEGMENTATION:
  - IMAGE AND SEGMENTATION MASK PAIR



- Given a task and enough labels, supervised learning can solve it really well.
- Good performance usually requires a decent amount of labeled samples.

Challenges and Limitations of Supervised Learning

- Annotation of large scale dataset is:
  - Time consuming,
  - Cost ineffective, and
  - Hard to be scaled up
- Inflexibility in Adapting to New Data
- Poor Performance in Complex or Unseen Tasks
- High Dependency on Human Expertise

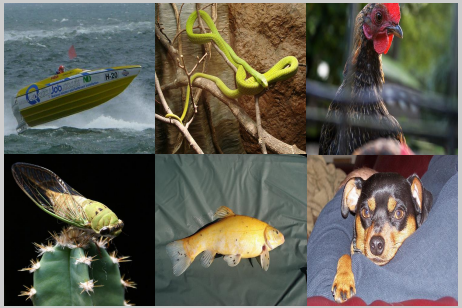


"Supervised learning is a bottleneck for building more intelligent generalist models that can do multiple tasks and acquire new skills without massive amounts of labeled data."

-Yann LeCun, Ishan Misra (Facebook AI)

Unsupervised Learning

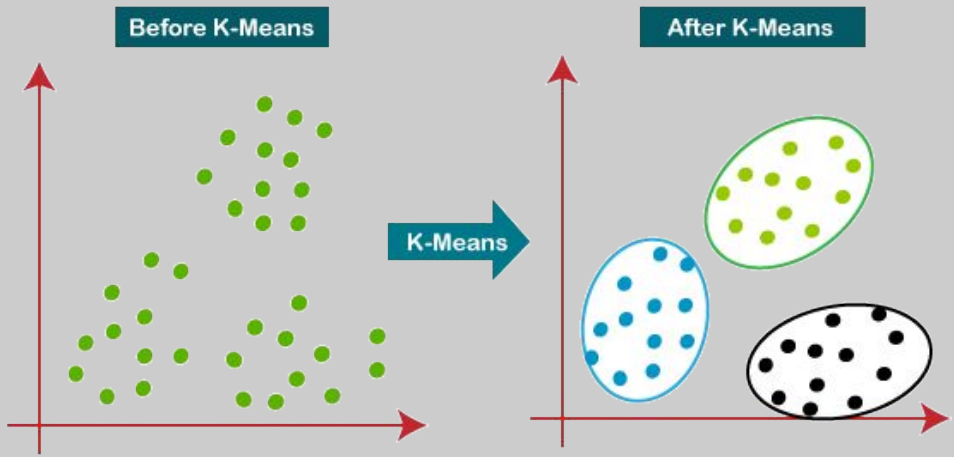
- Learning without the labels or human annotations.
- An example:



unlabeled Image Dataset

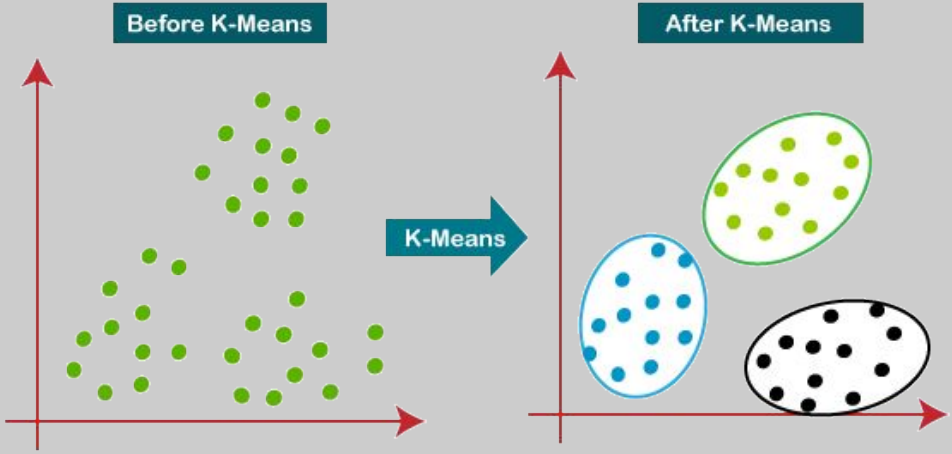
Unsupervised Learning

- Learning without the labels or human annotations.
  - An example is **K-means Clustering (A ML Algorithm to group unlabeled data)**



Unsupervised Learning

- Learning without the labels or human annotations.
  - An example is **K-means Clustering (A ML Algorithm to group unlabeled data)**



NOT A LEARNING METHOD

Unsupervised Training of a DL Model

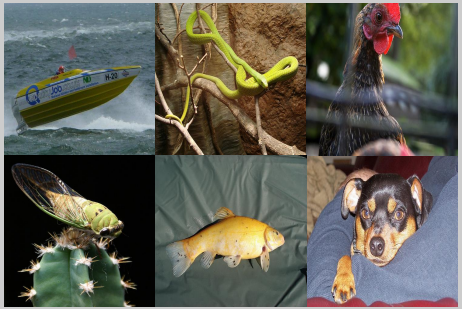
A deep learning model needs some kind of supervision for efficient training.

- No Label —---------> No Supervision —----------> No Training

Unsupervised Training of a DL Model

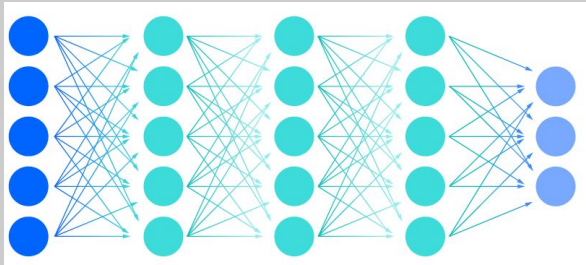
- No Label —---------> No Supervision —-----------> No Training

How to train a deep learning model with unlabeled data ?



unlabeled Image Dataset

&



Deep Learnng Model

Unsupervised Training of a DL Model

- No Label —---------> No Supervision —----------> No Training

How to train a deep learning model in unsupervised setting ?

- **Self-supervised Learning:** An unsupervised way of training a deep learning model.

Generate supervisory signal from the unlabeled dataset itself.

Supervisory signal as one of the properties of the unlabeled dataset.

Self-supervised Learning (SSL)

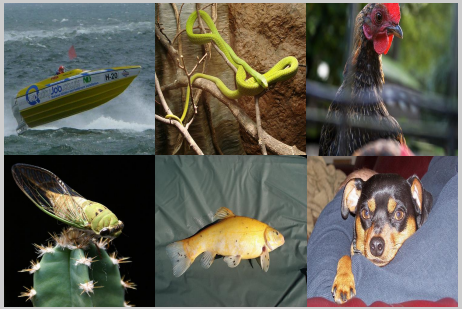
- Leverages **inductive bias** to guide the learning process.

Inductive bias is an assumption or prior about the unlabeled data like color, rotation, spatial or temporal information.

- An example of SSL is **Grayscal Colorization.**
  - **Inductive biased**: Color channels are highly correlated in RGB images, provide complementary information about same object.

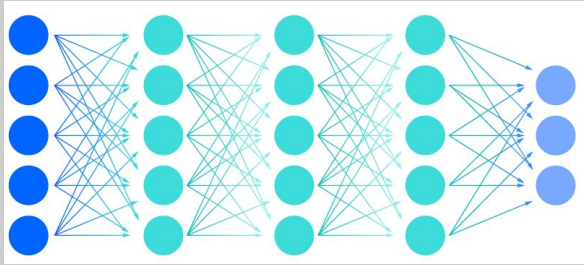
Self-supervised Learning

- **Problem Statement:** Given an unlabeled image dataset and a deep learning model, the aim is to learn the meaning of feature representations without labels.



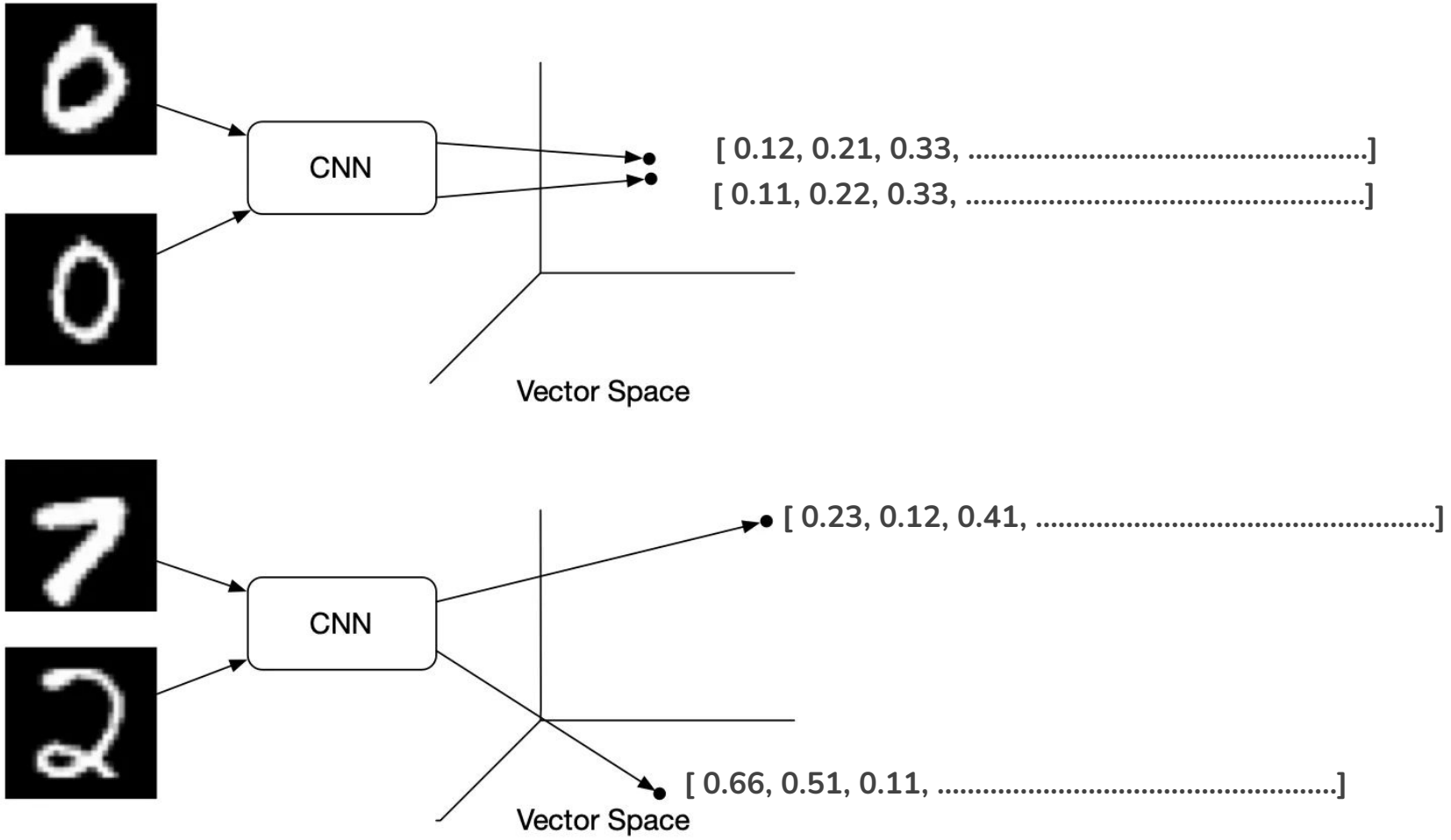
unlabeled Image Dataset

&



Deep Learning Model

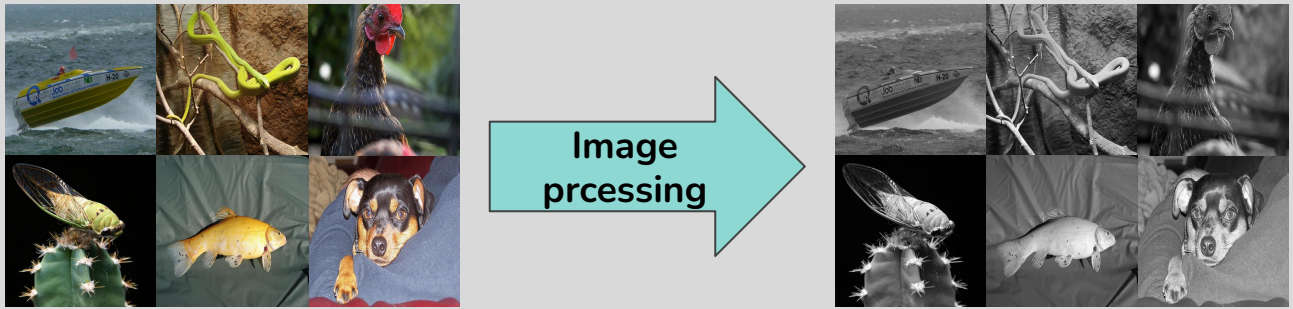
Feature representation of an image is a vector or feature map that captures the essential feature or characteristic of that input image.



Feature representation of an image is a vector or feature map that captures the essential feature or characteristic of that input image.

Grayscale Colorization: A pretext Task
(Self-supervised Training)

- **Step 1:** Obtain the grayscale image of each RGB image in dataset.

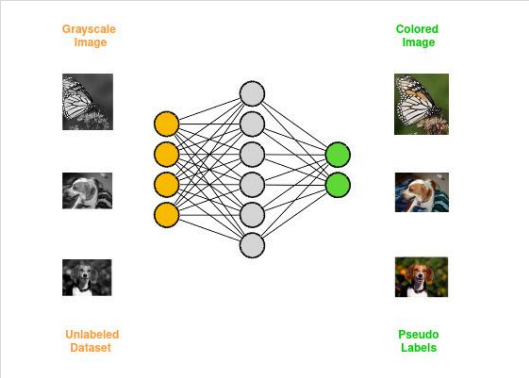


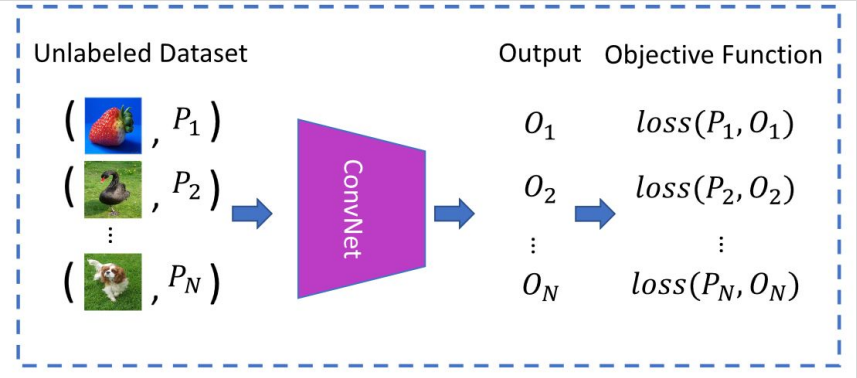
Color Images

Grayscale Images

Grayscale Colorization: A pretext Task

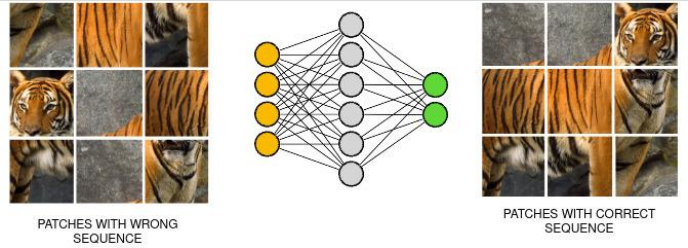
- **Step 2:** Train the deep learning model to color input grayscal image.



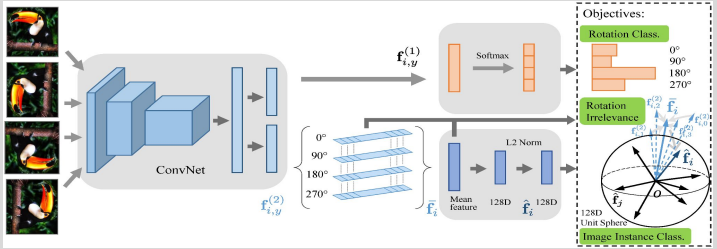


Pretext tasks are pre-designed tasks for networks to solve, and visual features are learned by learning objective functions of pretext tasks.

Pretext Task

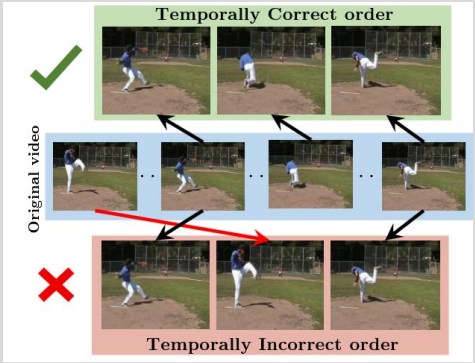


JIGSAW PUZZLE

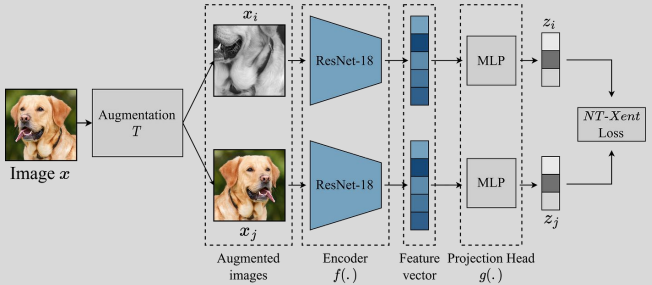


RATATION PREDICTION

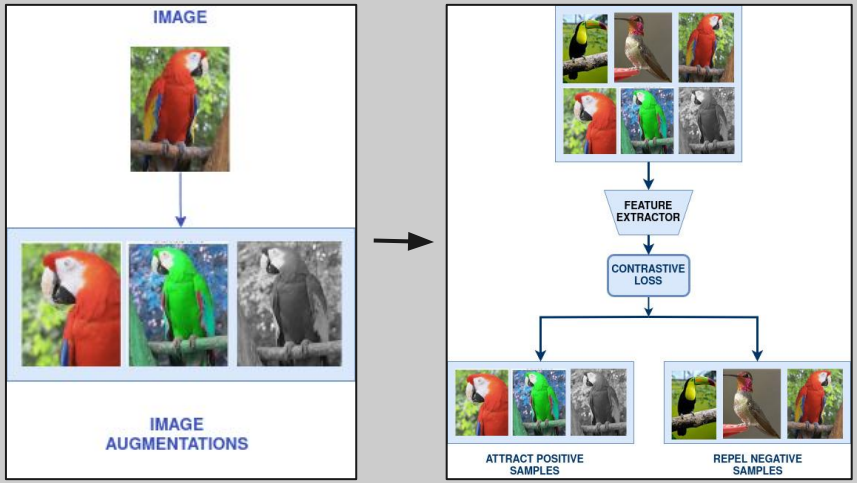
VIDEO FRAME ORDER PREDICTION

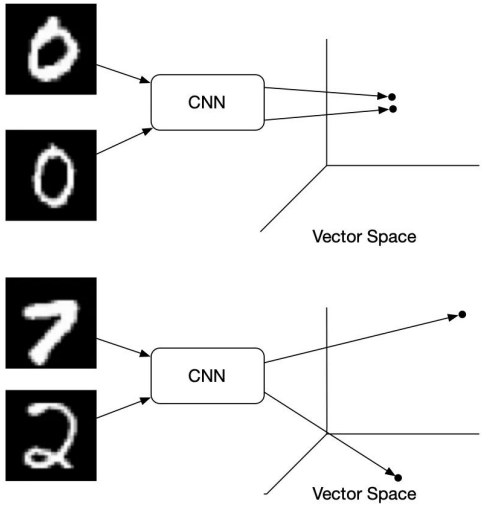


CONTRASTIVE LEARNING



Contrastive Learning





How to use Pretrained SSL Model for Downstream Tasks?

Transfer Learning

- **Downstream tasks are computer vision applications** like:
  - Image Classification,
  - Image Clustering,
  - Semantic Segmentation,
  - Object Localization, etc.
- These applications can **greatly benefit from the pre-trained models** when training data are scarce.

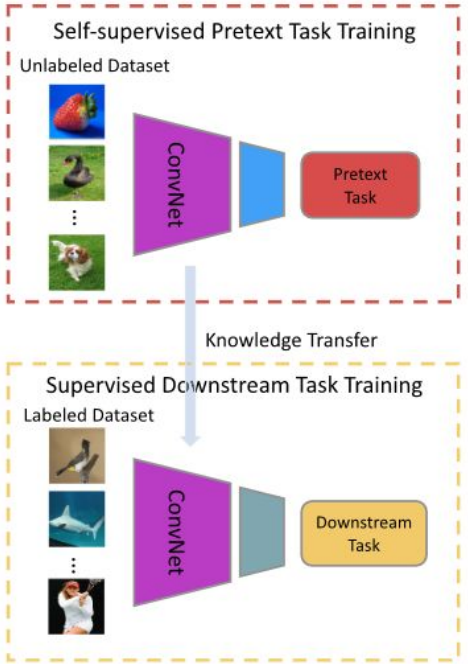
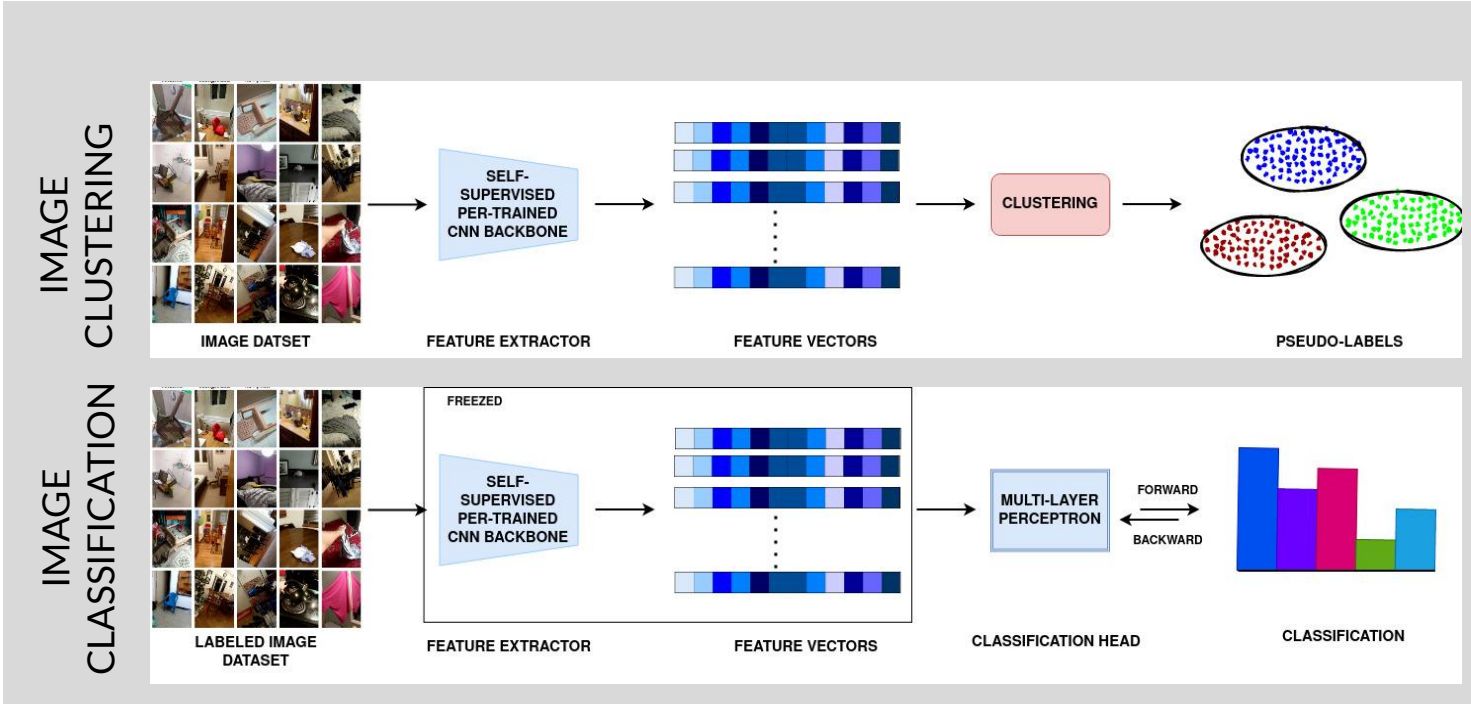
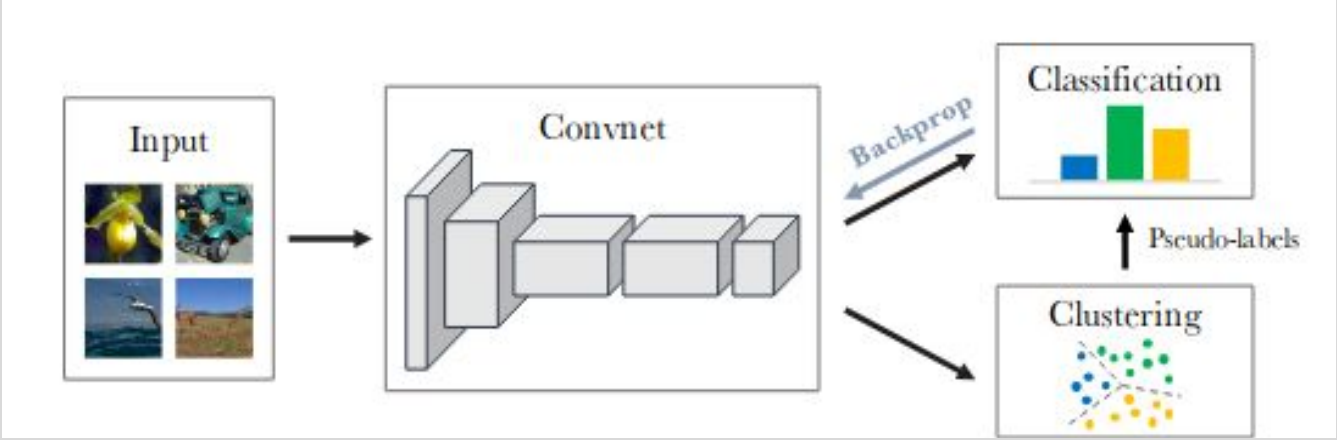


Image Classification & Clustering as Downstream Tasks



End-to-end Self-supervised Training

- Jointly learns the parameters of a neural network and the cluster assignments of the resulting features.



Advantage of Learning with unlabeled Data

- Unlabeled data is widely available in most domains.
- Removes or reduces labeling time and cost.
- Leading to better generalization across various downstream tasks and domain.
- Mitigate human bised introduced during labeling by learning directly from raw data.
- Unlabeled data can reveal hidden patterns, clusters, or relationships in the data.
- Can make use of both labeled and unlabeled portion of data.
- Better performance on rare or Imbalanced Classes.
- Improve learning across multiple modalities (e.g., text, image, audio).
- Good for continual learning setup where the model adapts to new data over time.

Real time applications of Self-supervised Models

- Conversational AI tools: **ChatGPT**
- Personal assistants: **Siri, Alexa, Google Assistant**
- Code auto-completion tools: **GitHub Copilot**
- Text Summarization tools: **Notion, Grammarly**
- Facial Recognition and Biometric Systems: **Apple ID**
- Noise suppression in video calls: **Zoom, Microsoft Teams**
- Delivery drones: **Amazon Prime Air**
- Visual search tools: **Google Lens**
- AI art generation: **DALL-E**
- Content moderation and creation tools: **Canva AI**

Conclusion

- Supervised learning is a bottleneck for building more intelligent generalist models.
- SSL unlocked the true potential of artificial intelligence by learning with raw data.
- It bridges the gap between **unsupervised learning** and **supervised learning.**
- SSL is reshaping machine learning by making it more **efficient, scalable, and adaptable**

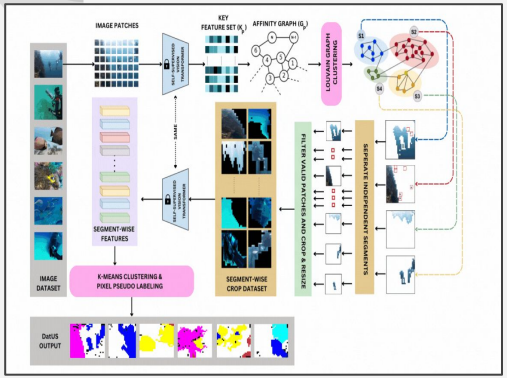
"Machines that learn from unlabeled data are the key to scaling intelligence in the real world."

Referances

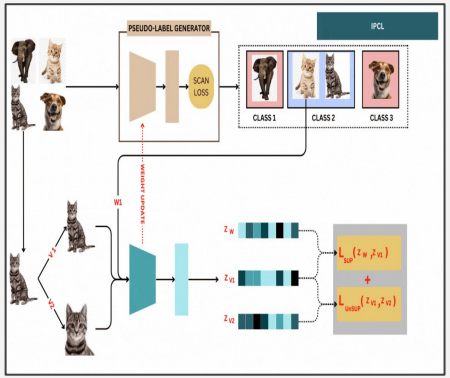
- Jing, L. and Tian, Y., 2020. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *43*(11), pp.4037-4058.
- Caron, M., Bojanowski, P., Joulin, A. and Douze, M., 2018. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 132-149).
- Self-supervised learning: The dark matter of intelligence
- Weng, L. (2019, November 10). *Self-supervised learning*. Retrieved from https://lilianweng.github.io/posts/2019-11-10-self-supervised/
- Hands on SSL: https://colab.research.google.com/drive/1Z9Zax5hTxgp1FtwOOwOiJ6ShEFW5zf4A?usp=share_link#scrollTo=M9jpzeHA6Hh1

Our Research @ Self-supervised Learning

DatUS
IEEE TRANSACTION



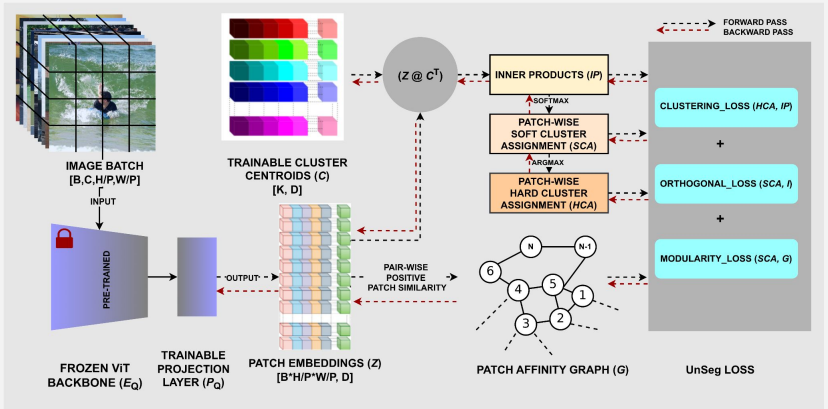
IPCL
ICASSP 2024



LinkedIn

GitHub

GraPix
ICPR 2024



THANK YOU
ANY QUESTIONS ?

LinkedIn

THANK YOU
ANY QUESTIONS ?