IIT Gandhinagar
ACM Summer School 2025

Self Supervised Learning on Vision

Tutor: Rishabh Mondal, Diya Thakor

10-06-2025

SUSTAINABILITY LAB

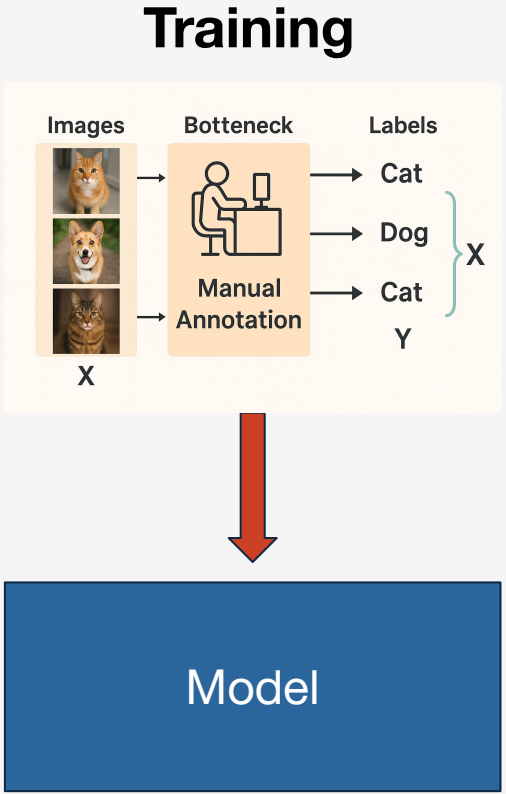
Supervised Learning

Training:

- Input X → Label Y
- Model learns to map X → Y using labeled data

Testing:

- Given new X, model predicts corresponding Y



Supervised Learning

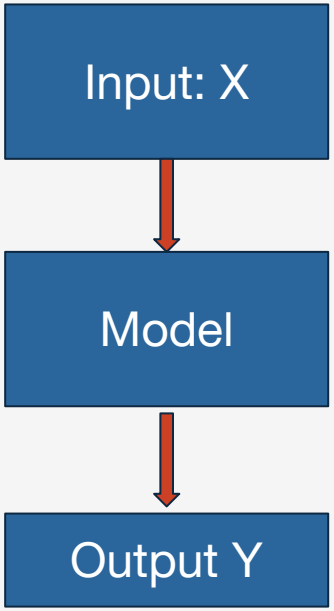
Training:

- Input X → Label Y
- Model learns to map X → Y using labeled data

Testing:

- Given new X, model predicts corresponding Y

Testing



Why Is That an Issue?

- We need a **huge** amount of **labeled data for training**

- Not all data is labeled
- **Labels** can be **expensive or hard** to get (e.g., medical images, satellite data)

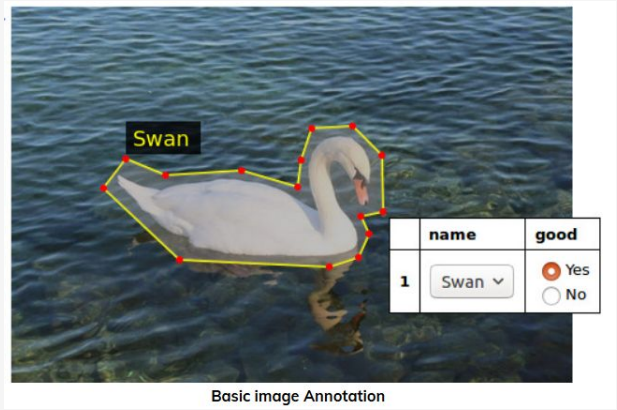
Images in your mobile

Why Is That an Issue?

How many images are there in mobile gallery?

Give label to each image

- Not all data is labeled
- **Labels** can be **expensive or hard** to get (e.g., medical images, satellite data)



Basic image Annotation

Manual effort, Domain expertise, Time

SUSTAINABILITY LAB

What is image embedding?

- A numerical vector representation of an image capturing its semantic content.
- Deep neural networks (e.g., CNN, ViT), usually from intermediate or final layers.
- Similar images will have similar embeddings.

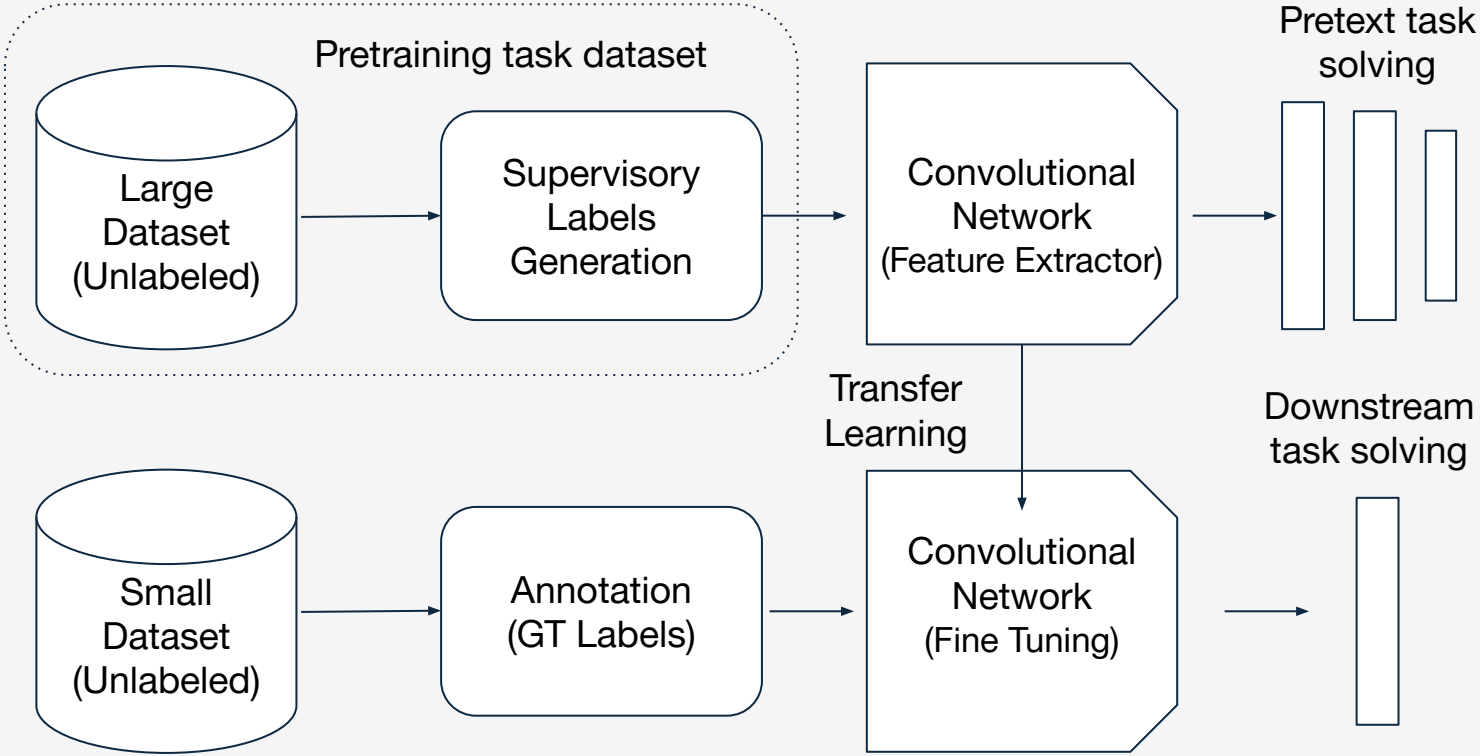
$$\begin{matrix} 0.12 \\ 0.51 \\ \cdots \quad 0.40 \\ \vdots \\ 0.38 \end{matrix}$$

Solution: Self Supervised Learning

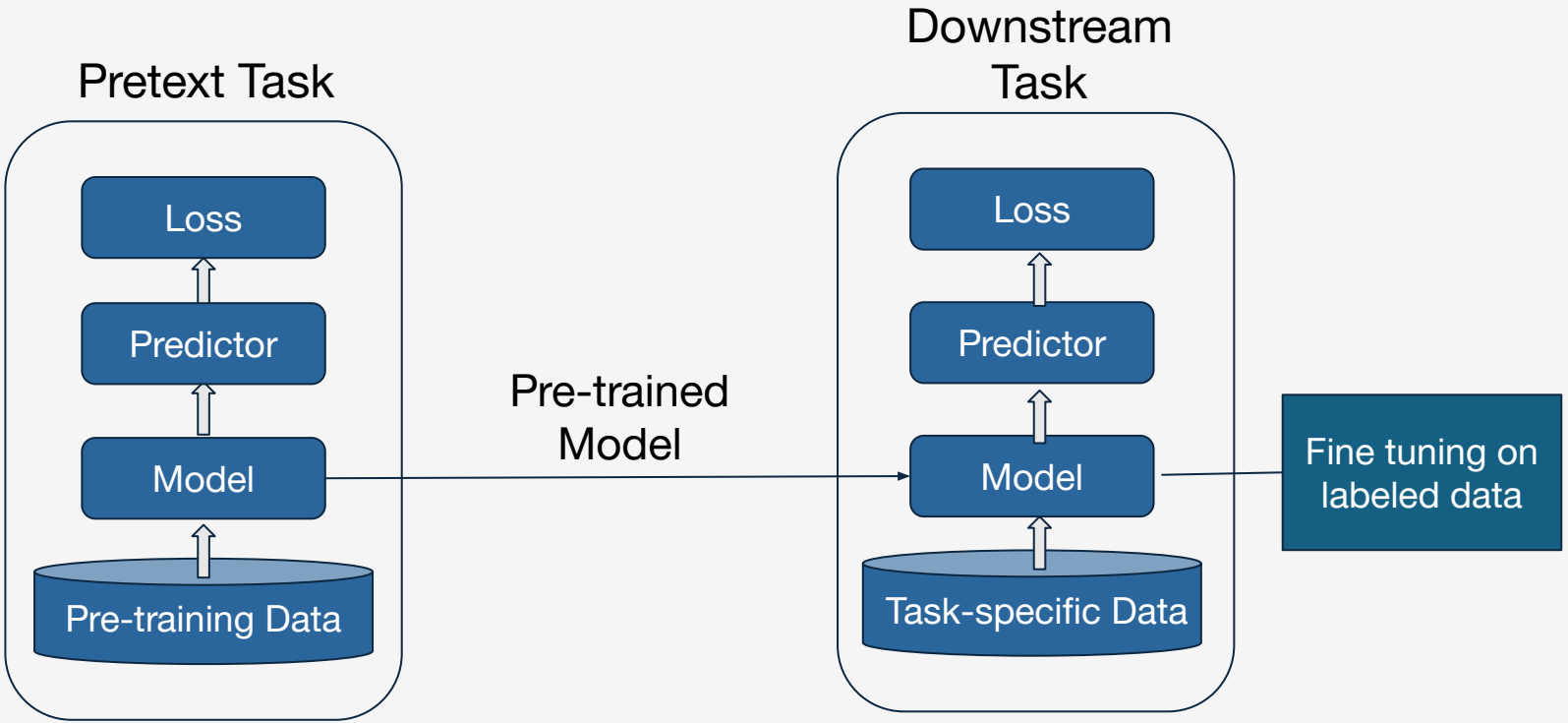
Self Supervised Learning

- The model **supervises itself** using **patterns** in the data.
- **No** need for **human-labeled data** — it generates its own learning representations.

Flow of Self Supervised Learning



Self-Supervised Learning Pipeline



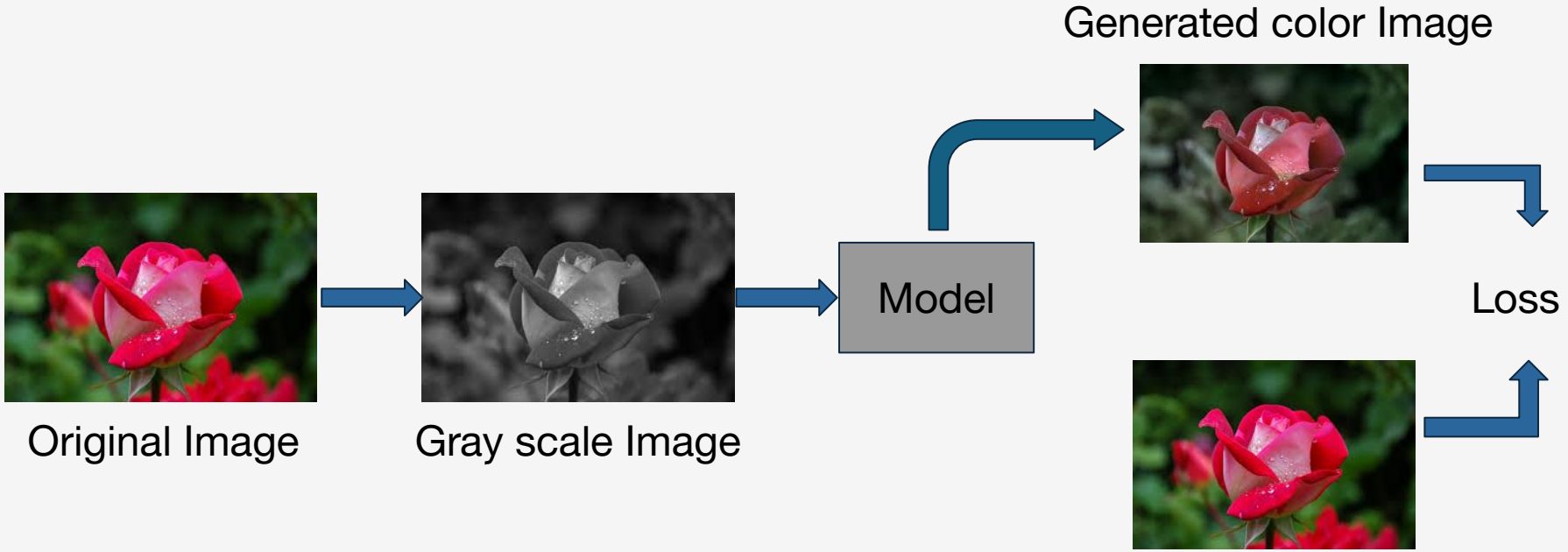
Different pretraing tasks of SSL

A task designed **using unlabeled data** to help the model **learn** general-purpose **features**

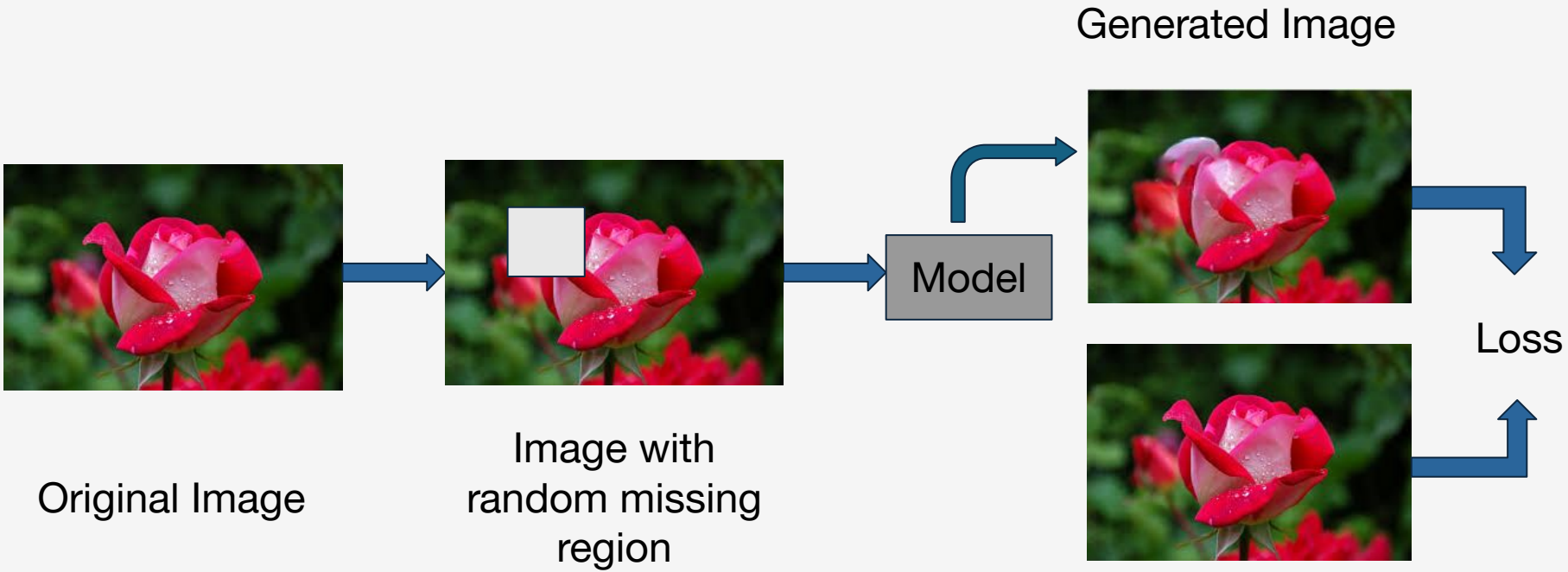
- Image Colorization
- Image Inpainting
- Geometric Transformation Recognition
- Jigsaw puzzle
- DINO
- SimCLR

https://www.jmlr.org/papers/volume23/21-1155/21-1155.pdf

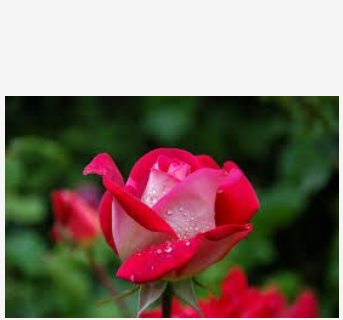
Pre Training Task: Image Colorization



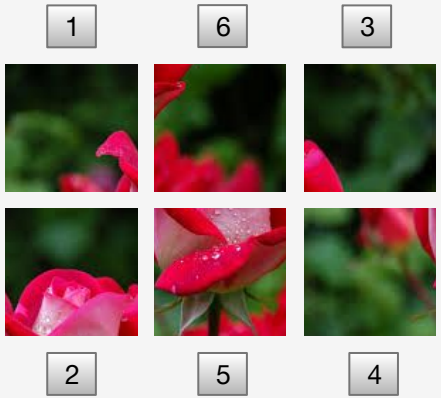
Pre Training Task: Image Inpainting



Pre Training Task: Jigsaw Puzzle



Original Image



Jigsaw Puzzle

Model

Predicted Sequence

Loss

Original Sequence

Pre Training Task: Geometric Transformation Recognition

Original Image

Rotated Image

Model

Predicted rotation (classification):
0˚ , 90˚ , 180˚ , 270˚

4 classes

Loss

Actual rotation: 180˚

Pre Training Task: SimCLR

Image 1

Image 2

Pre Training Task: SimCLR

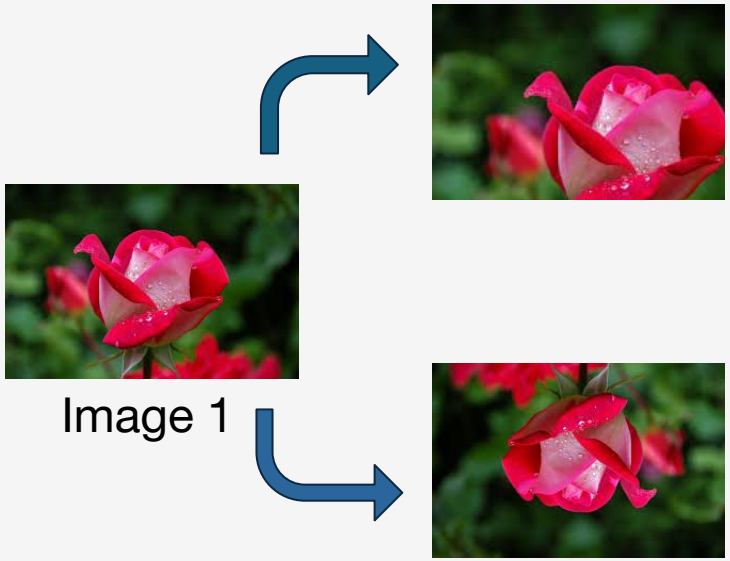
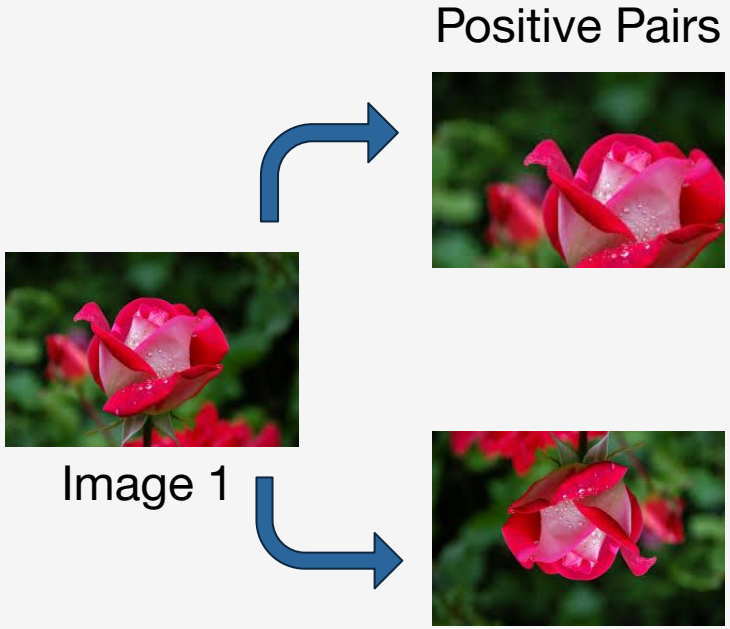


Image 1

Transformation 1:
Cropped and
resized image

Transformation 2:
Rotated Image

Pre Training Task: SimCLR



- Both images are generated **from same image** using transformations like rotation, crop, etc.
- So their **embeddings must be similar**
- By forcing model to generate same embedding we are forcing it to learn the pattern of data

Pre Training Task: SimCLR

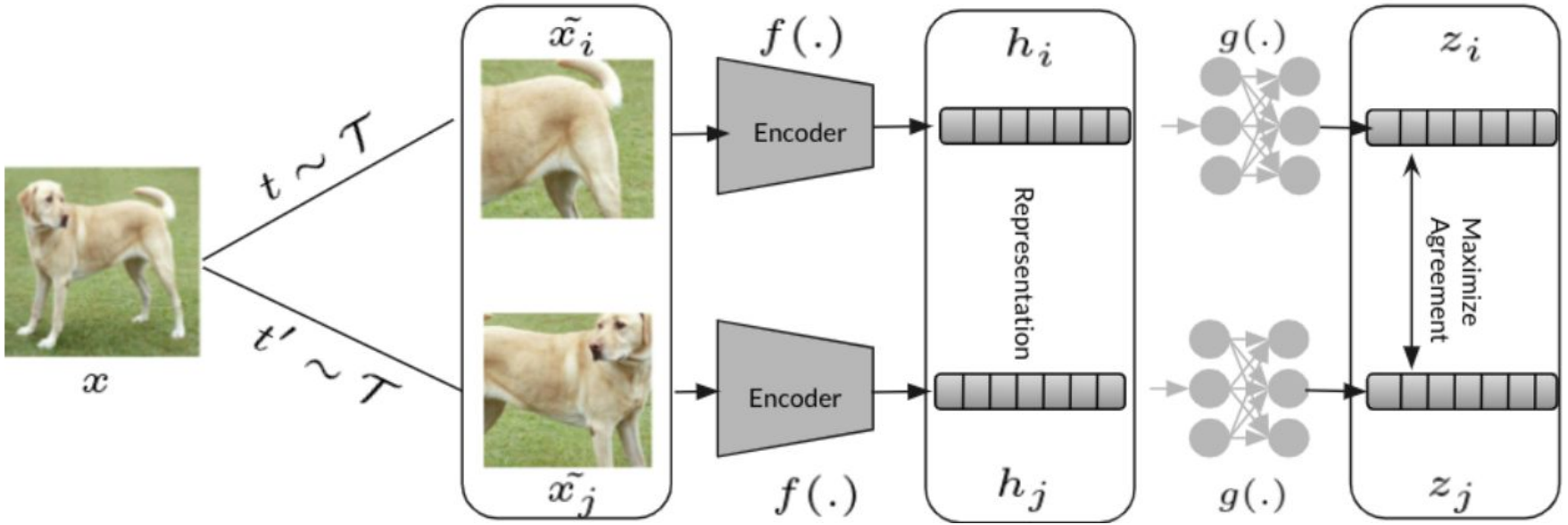
Negative Pairs

Image 1

Image 2

- Both transformations are generated **from different images**
- So here their **embedding must be very different**
- This is the approach of SimCLR (Simple **Contrastive Learning** of Visual Representations)

Pre Training Task: SimCLR



https://vasudev-sharma.github.io/posts/2022/03/SimCLR-visually-explained/

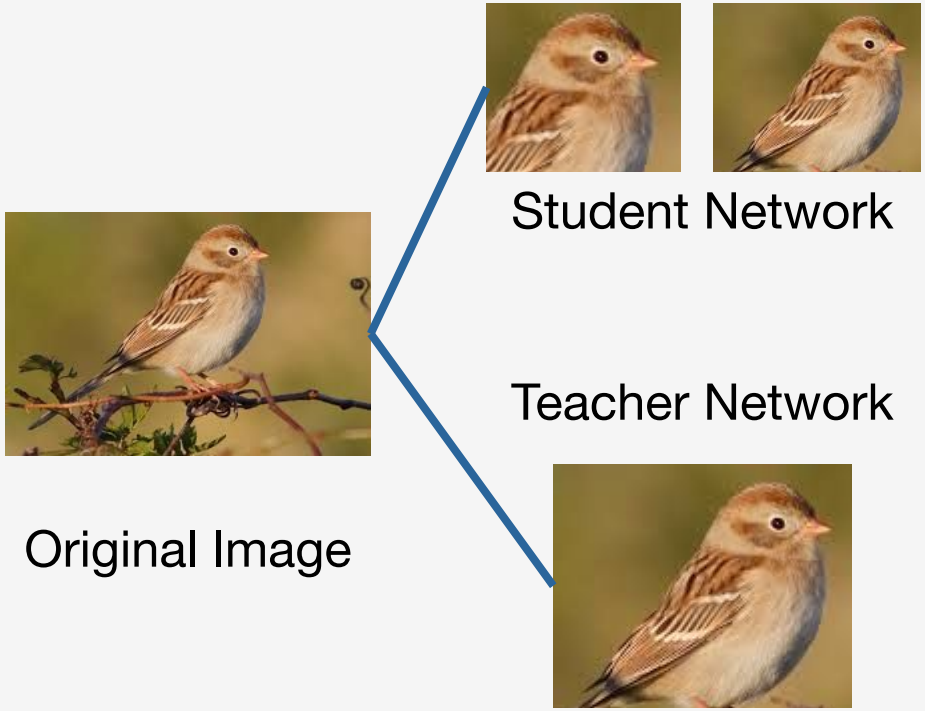
Pre Training Task: DINO

Original Image

Local View:
Small part of original image is cropped (<50%)

Global View:
Larger part of image is kept in cropped image (>50%)

Pre Training Task: DINO



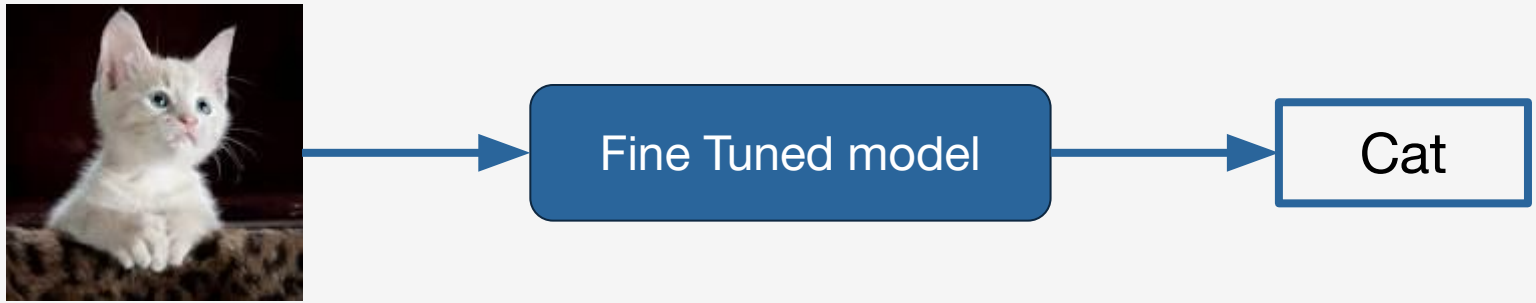
- **Student:** local + global views
- **Teacher:** only global views
- Teacher gives stable targets
- Student learns by matching teacher's output

Downstream Task:

- We have huge data but from that very less 10-20% data is labeled data and other is unlabeled.
- A task with **labeled data** used to **evaluate or fine-tune** the model after pretraining
- Tasks:
  - Classification
  - Segmentation
  - Detection

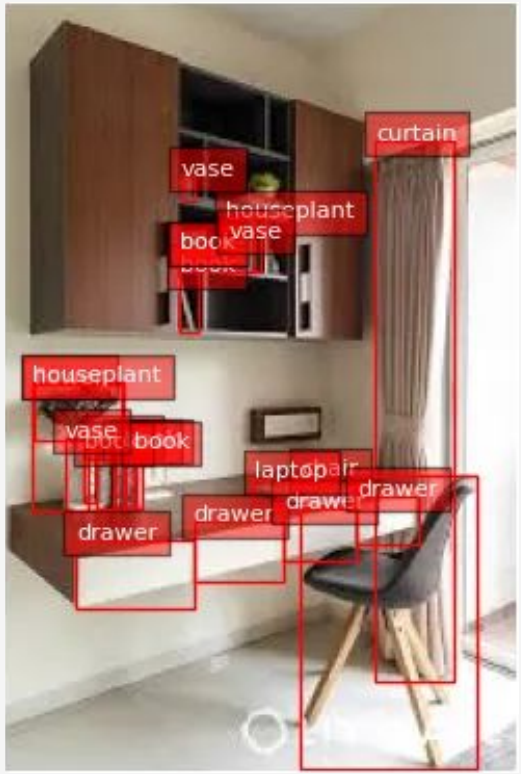
Downstream Task: Classification

Assign label to input image



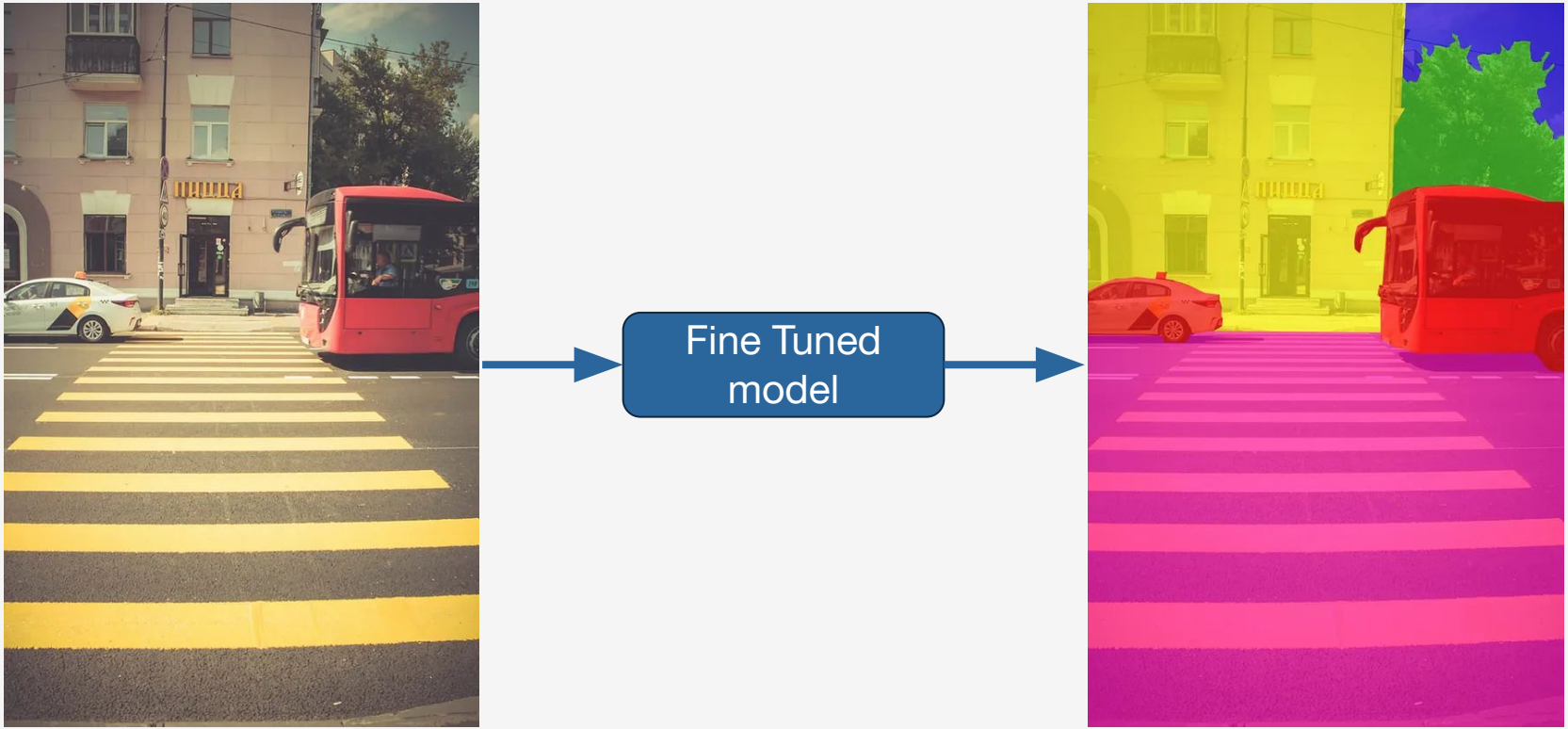
Downstream Task: Object Detection

Fine Tuned model



SUSTAINABILITY LAB

Downstream Task: Segmentation



Solo - learn library