A Graph-Theoretic Framework for Understanding Open-World Semi-Supervised Learning

NeurIPS 2023 (Spotlight)

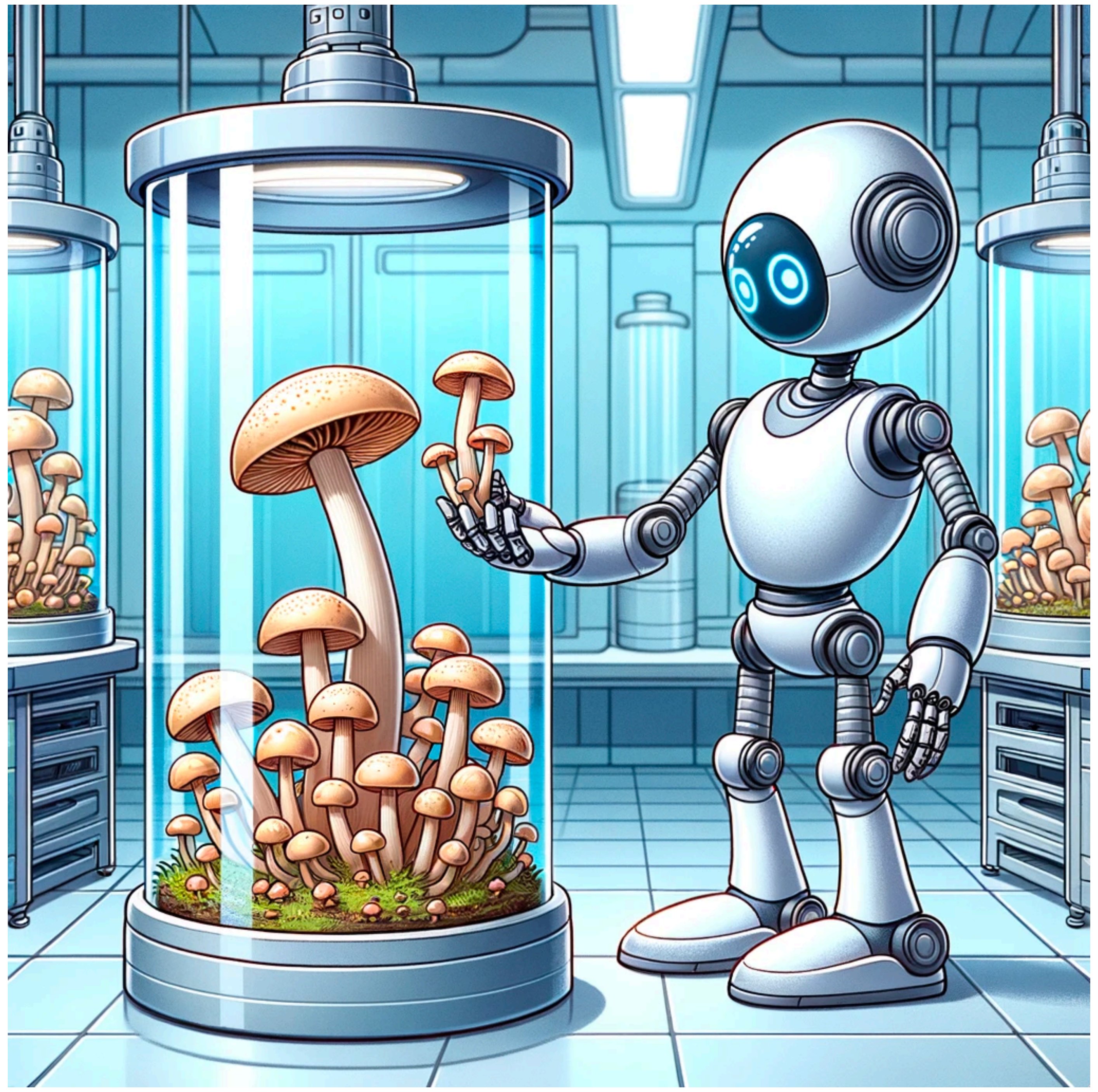
Yiyou Sun
UW-Madison

Zhenmei Shi
UW-Madison

Yixuan Li
UW-Madison

A Paradigm Shift from Closed-world to the Open-world

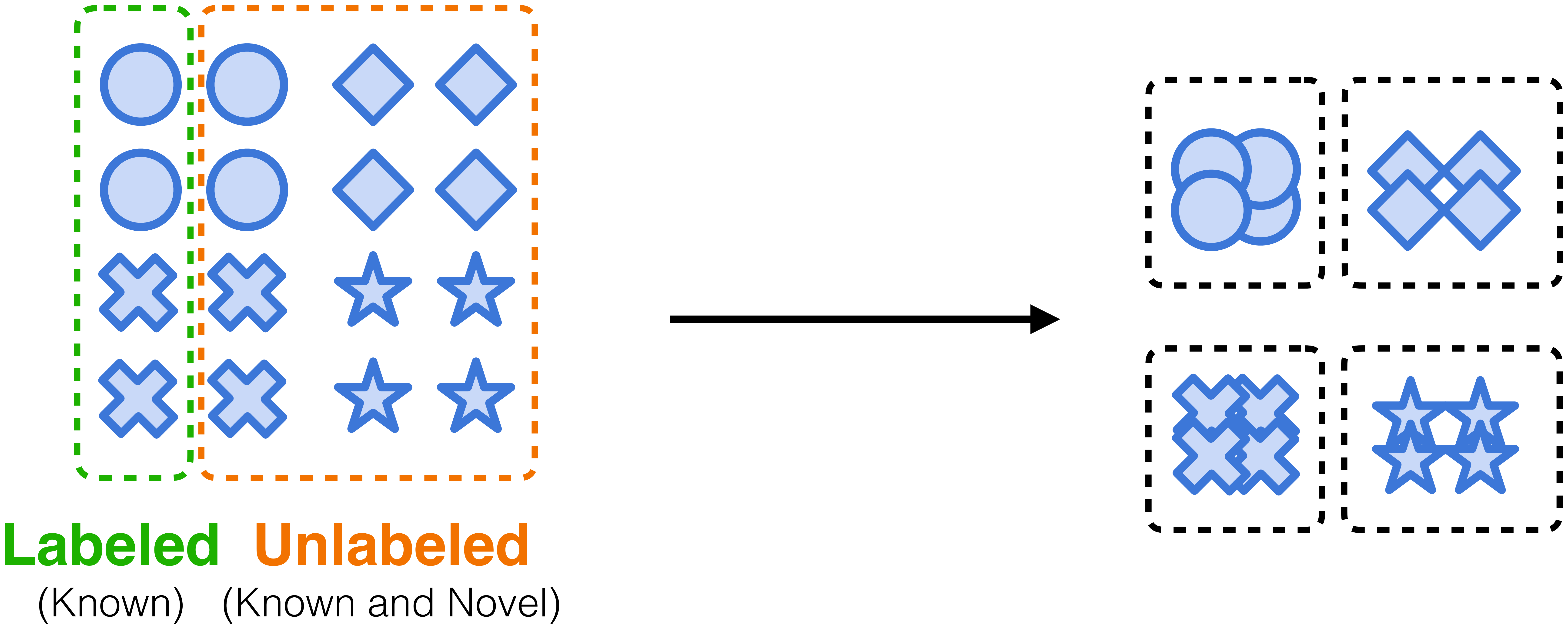
Closed-world ML:
Handle data with the **known** classes



Open-world ML:
Handle data with both **novel and known** classes

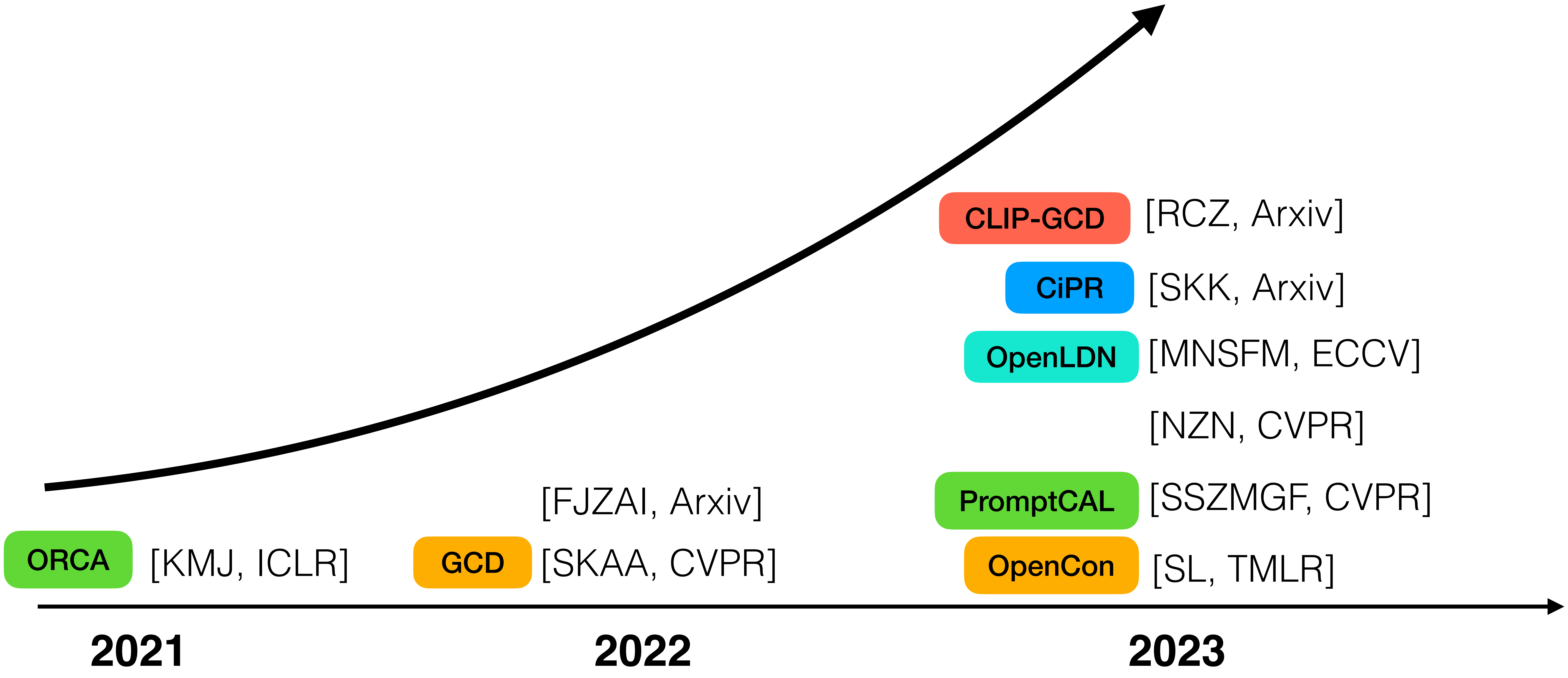
(Figures are powered by GPT-4V)

Open-world Semi-Supervised Learning

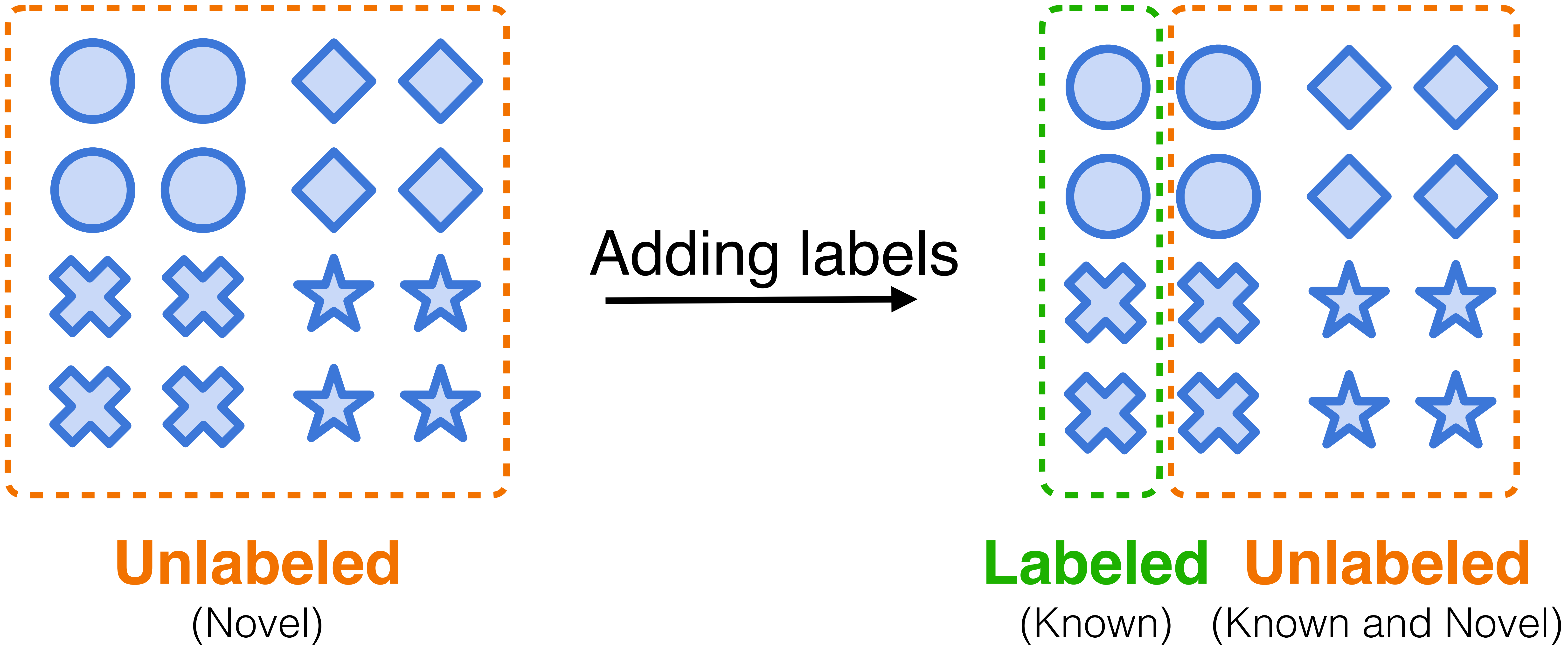


Goal: correctly classify known and cluster novel classes.

This research area starts to gain attention!



An Open Research Question

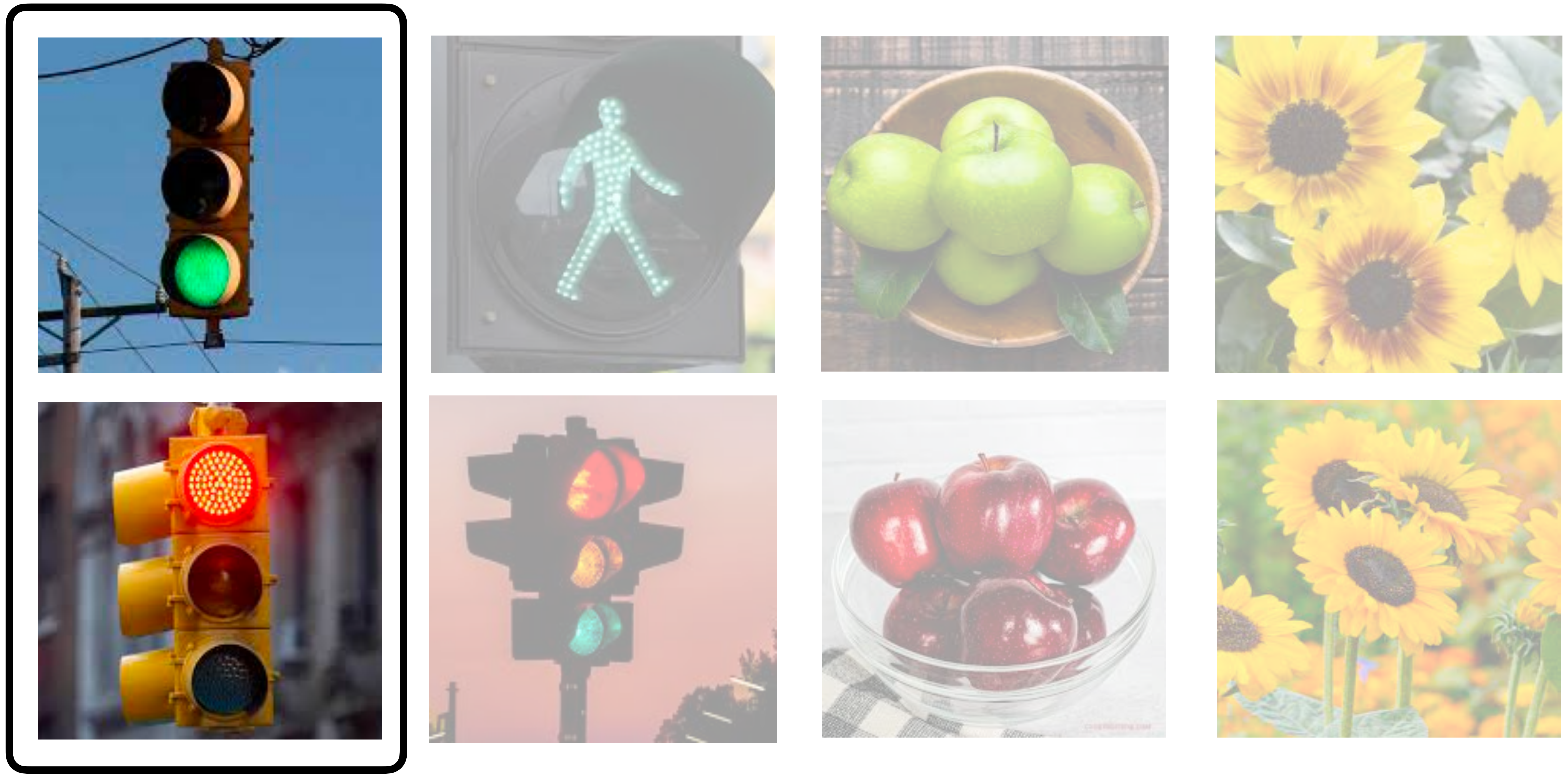


"what is the role of the label information in shaping representations for both known and novel classes?"

An Intuitive Example

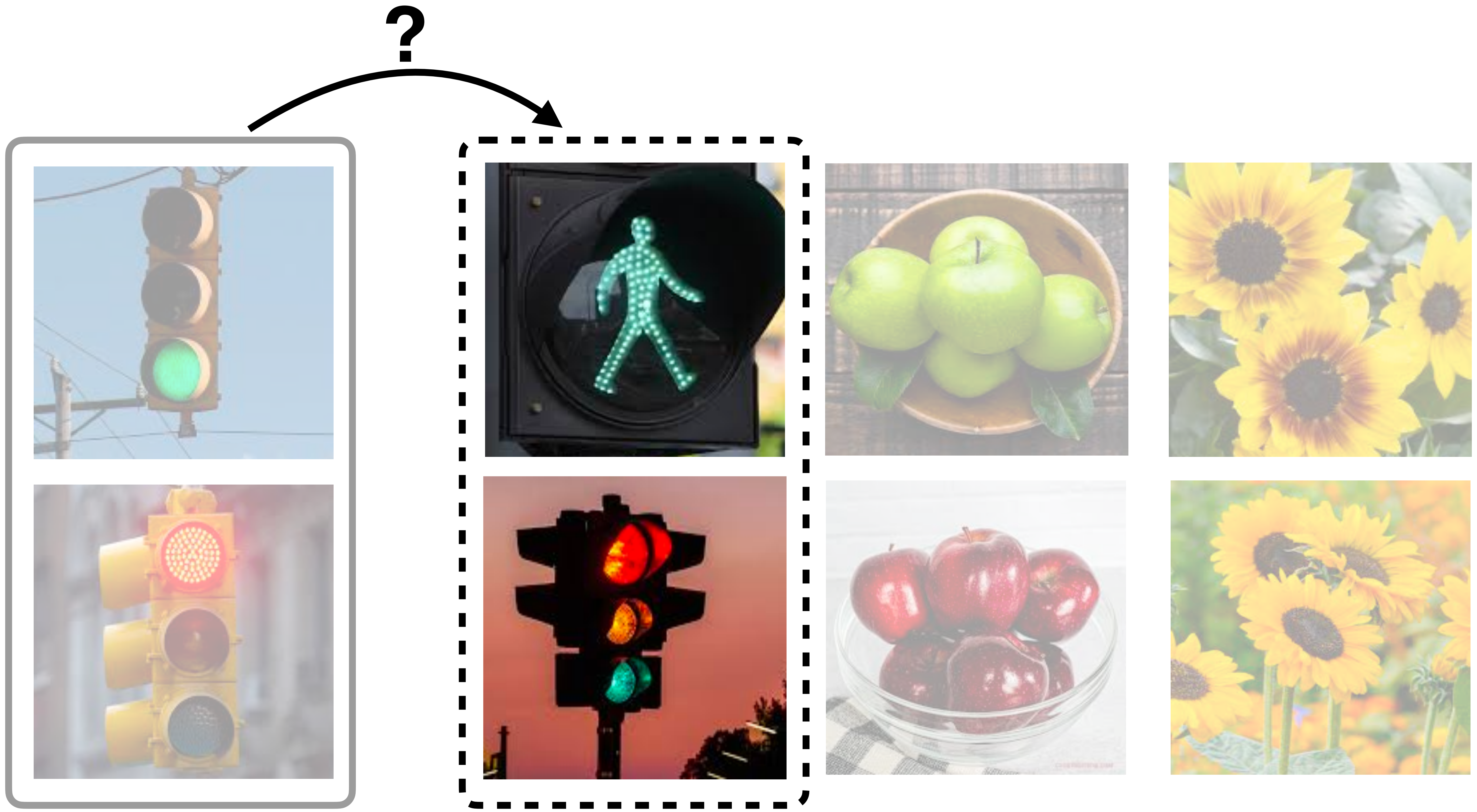
Starting Point: All **Unlabeled** Samples

An Intuitive Example



We label the first two images as "traffic lights"…

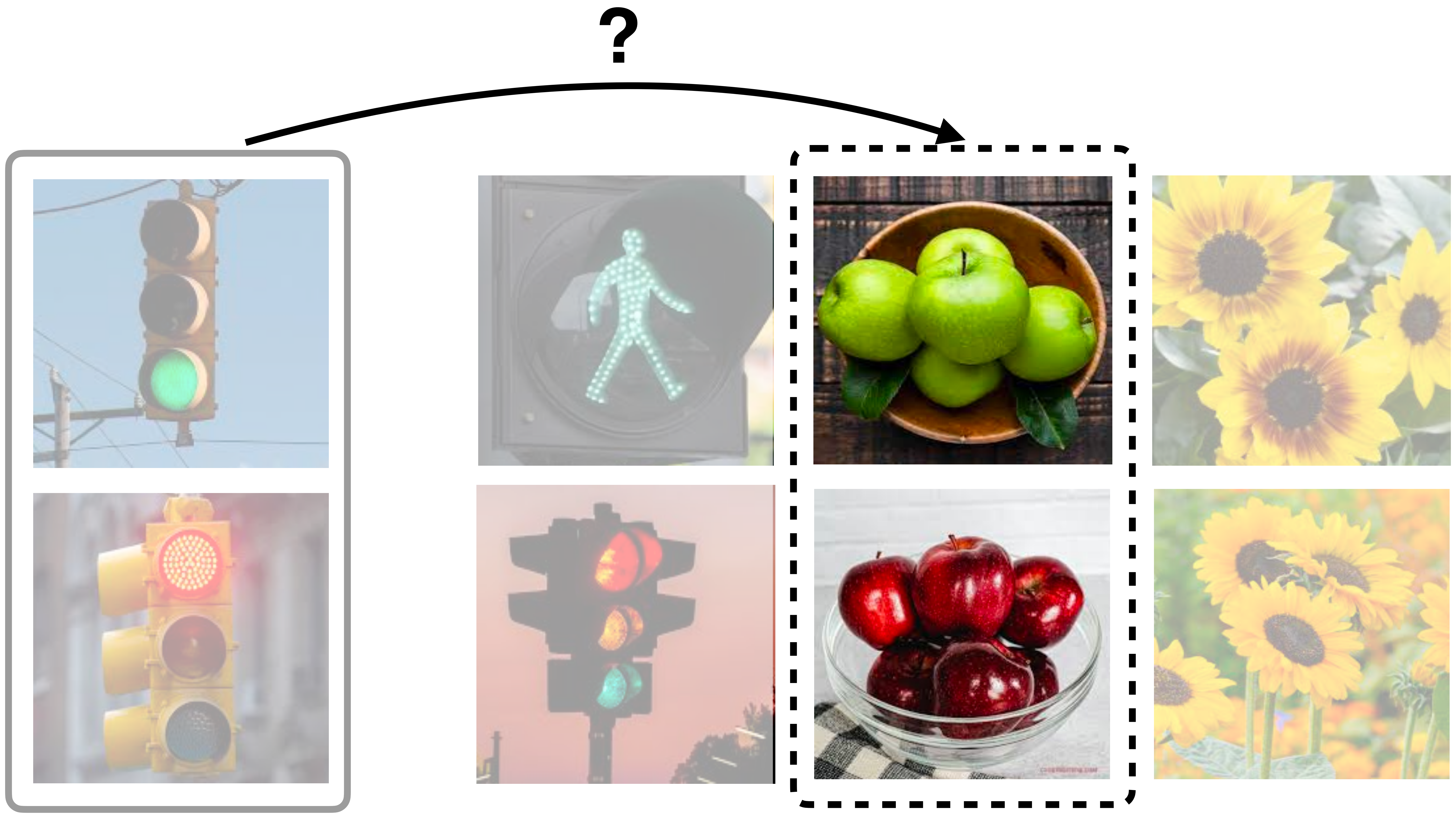
An Intuitive Example



Known Class

Question: Will other “traffic light” samples get closer to each other?

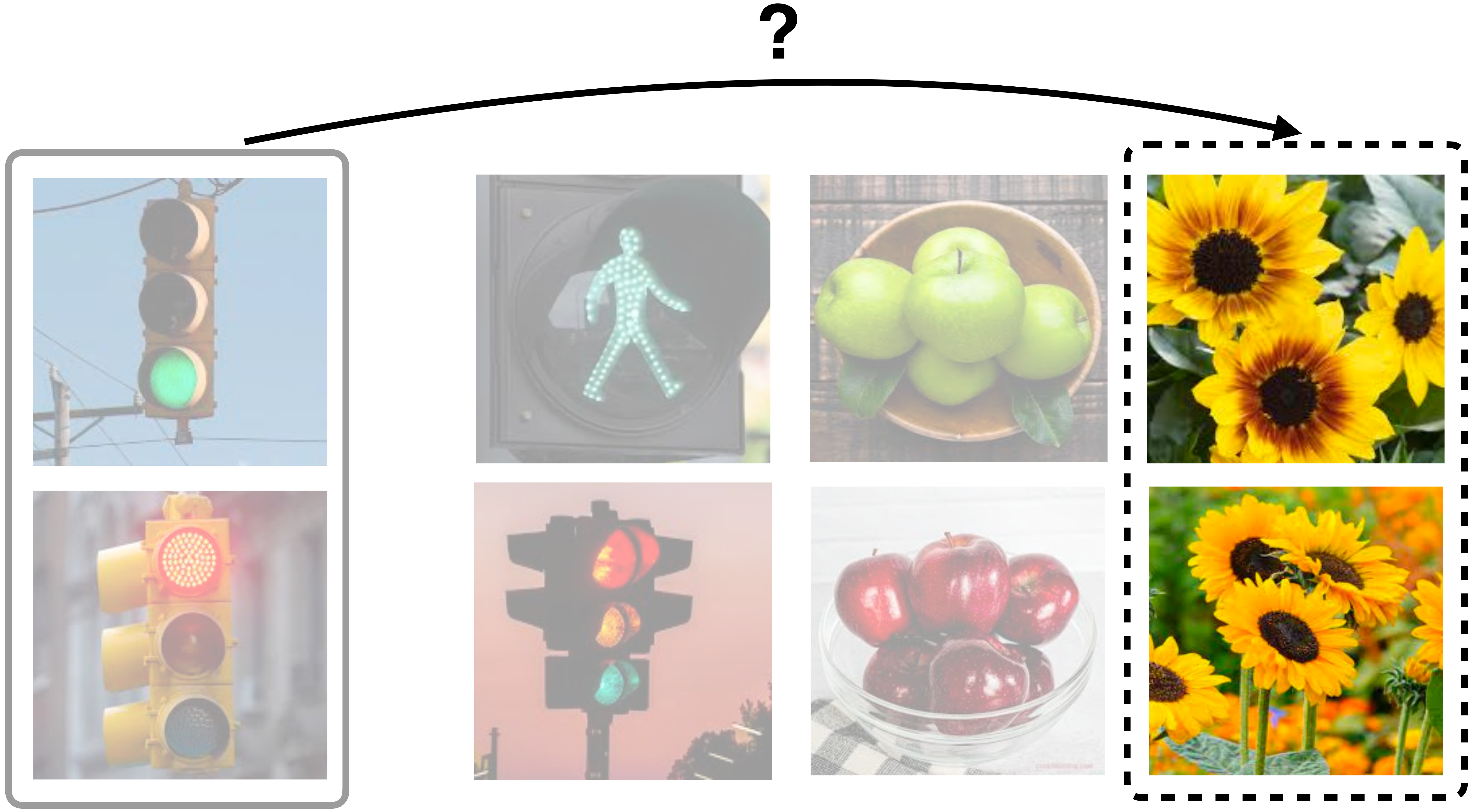
An Intuitive Example



Novel Class
(Strong relationship)

Question: Will other “green” samples get closer to “red” samples?

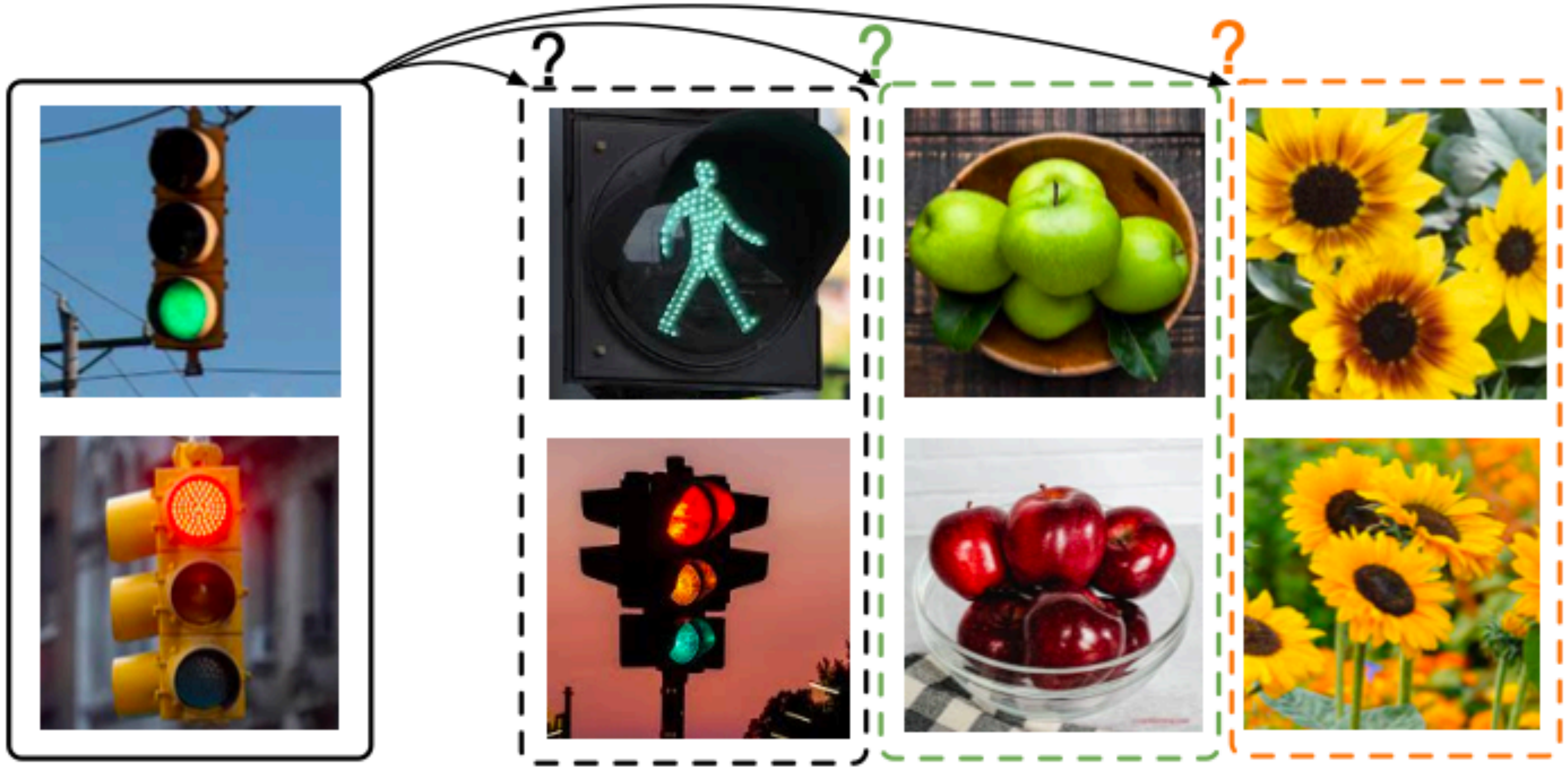
An Intuitive Example



Novel Class
(Weak relationship)

Question: Will unrelated novel class be affected?

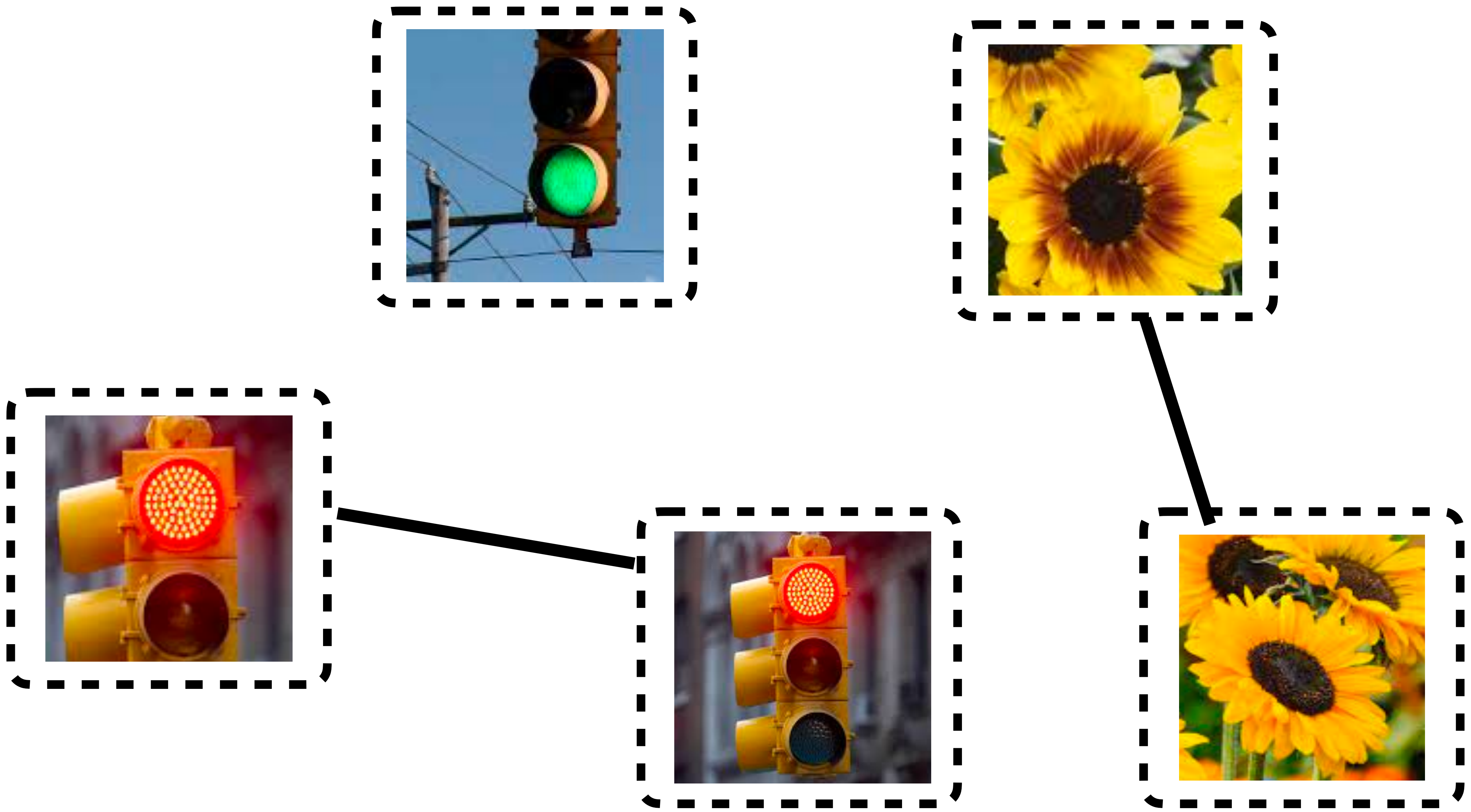
An Intuitive Example



A formal understanding is needed!

Methodology

Augmentation Graph

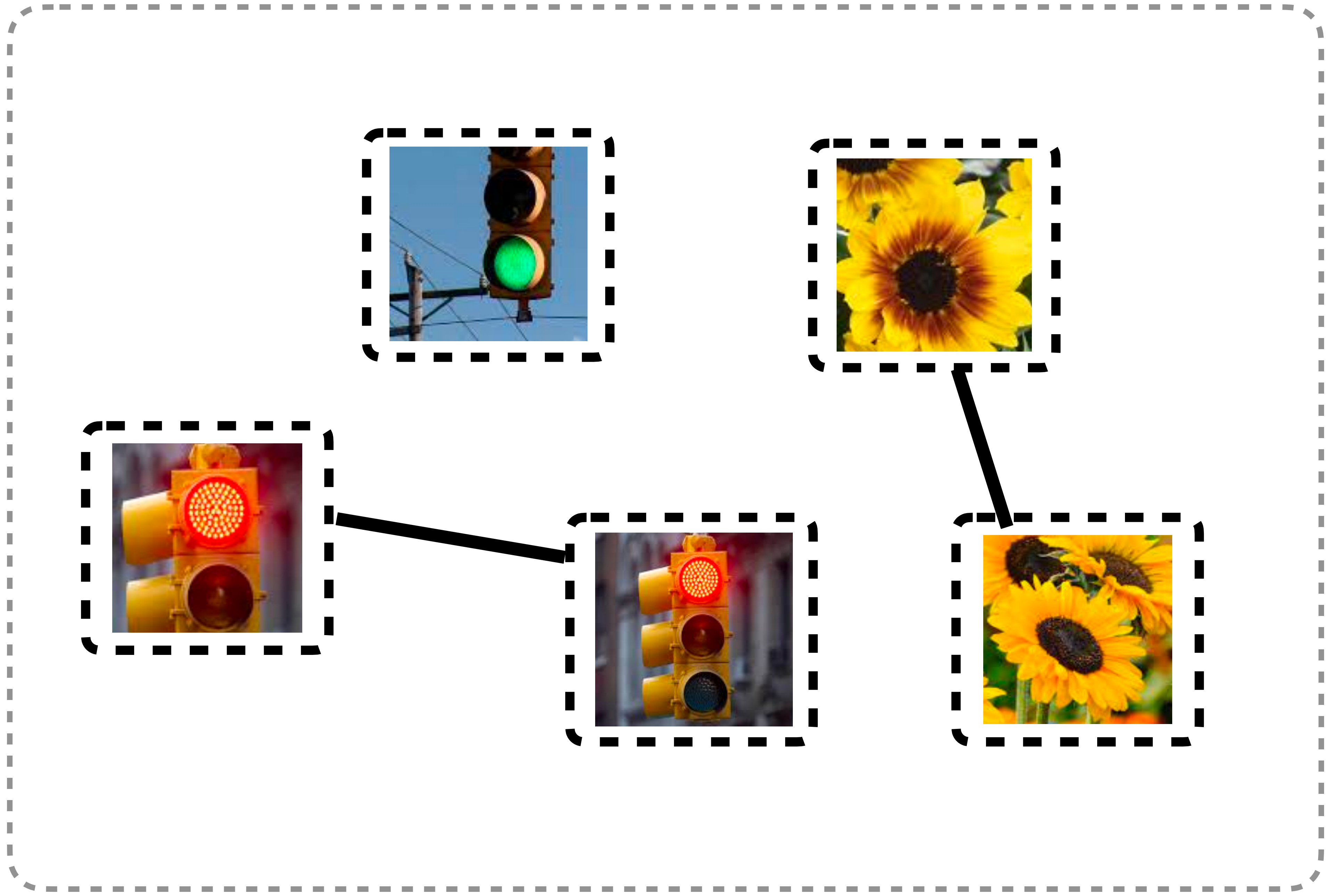


Node: Augmented Images.

Edge Weight: Probability of two images are considered as **positive pair**.

Label Perturbation

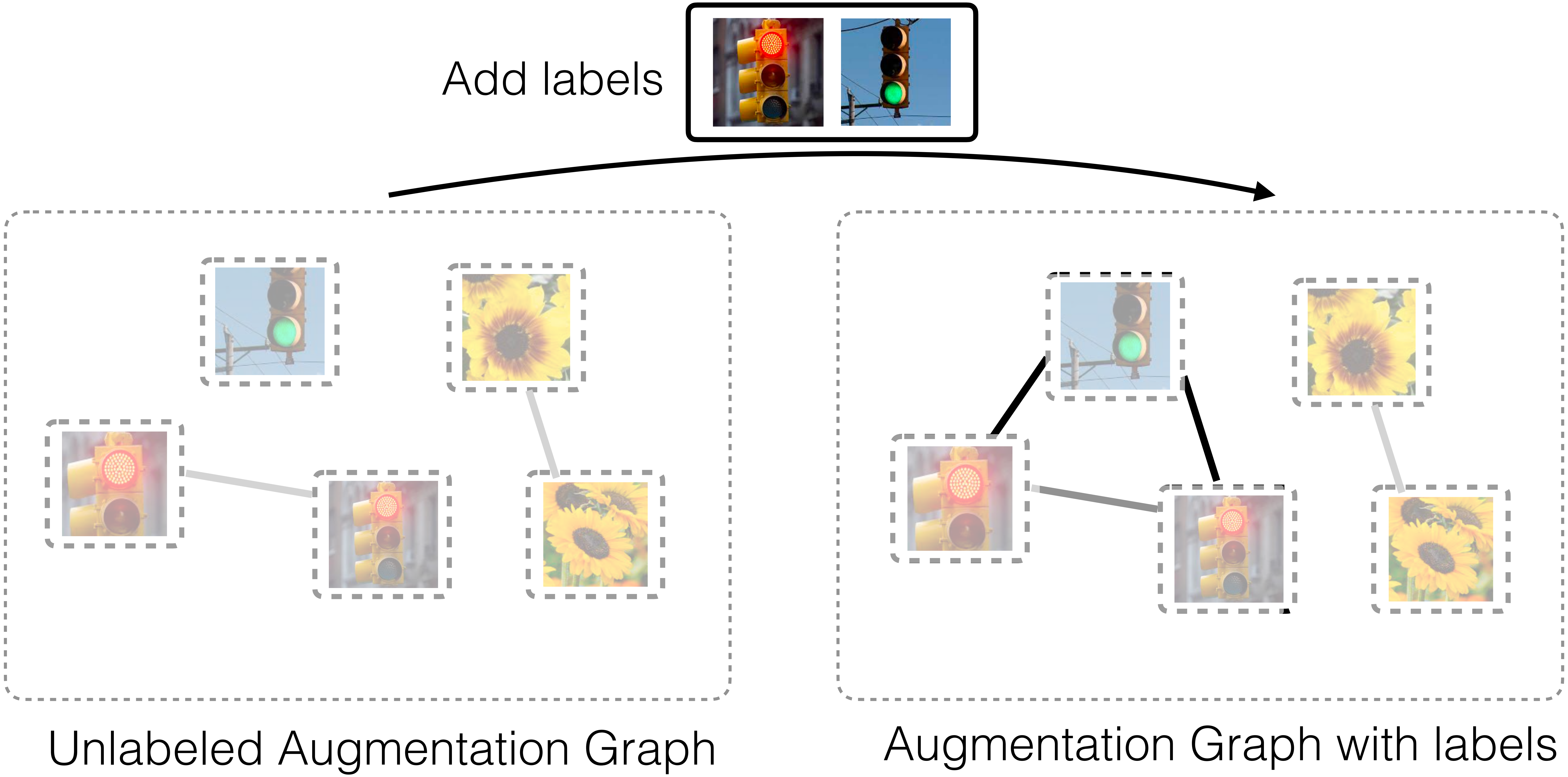
Adding labels changes the graph structure.



Unlabeled Augmentation Graph

Label Perturbation

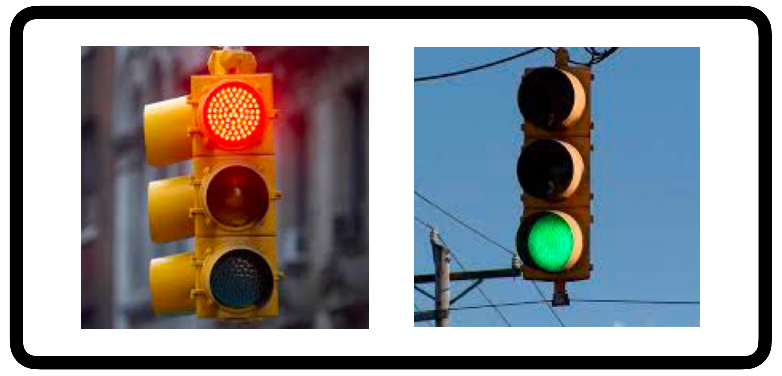
Adding labels perturbs the graph structure.



Label Perturbation

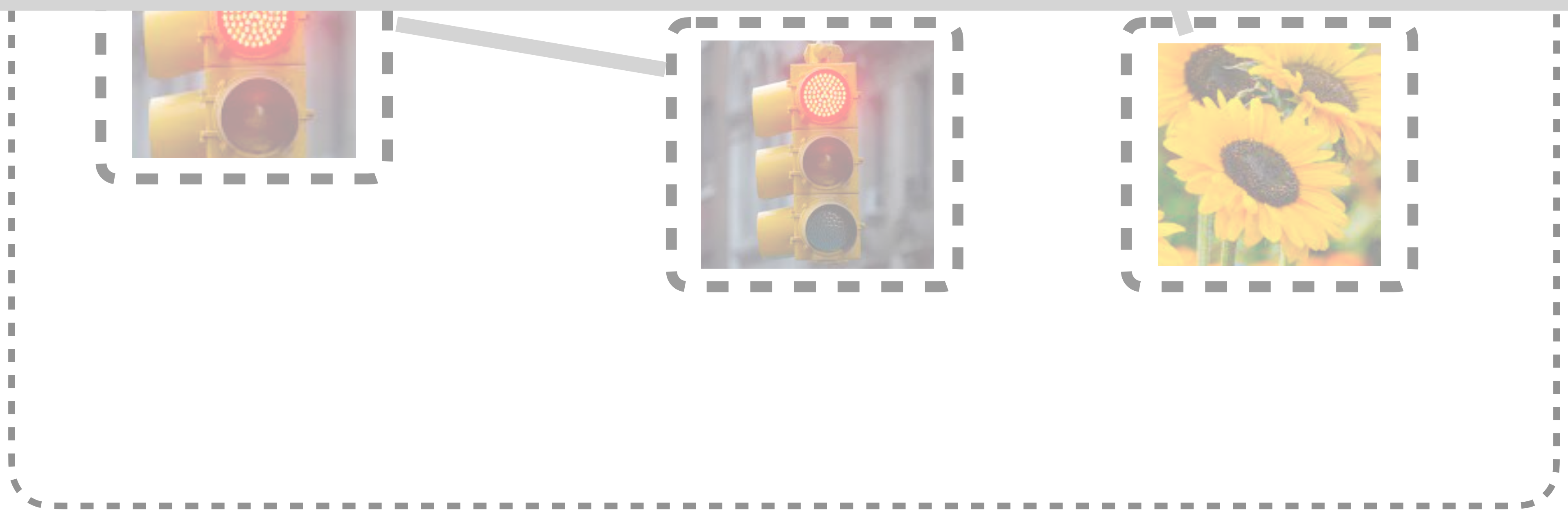
Adding labels changes the graph structure.

Add labels

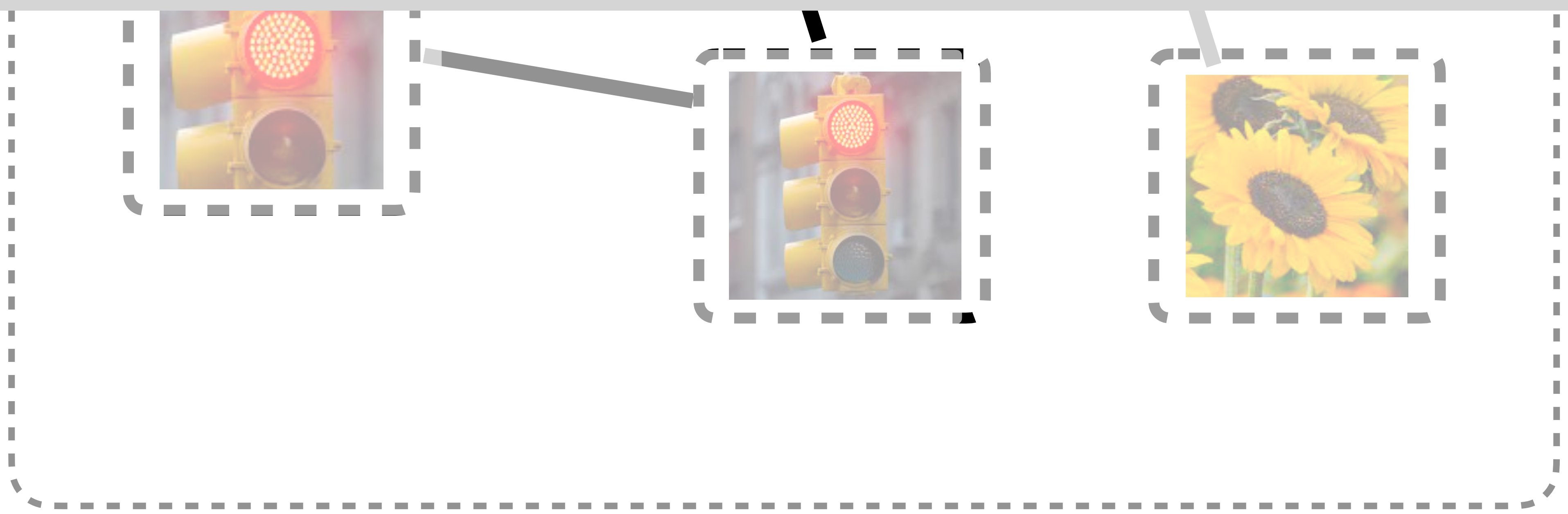


How do representations change?

How do cluster results change?

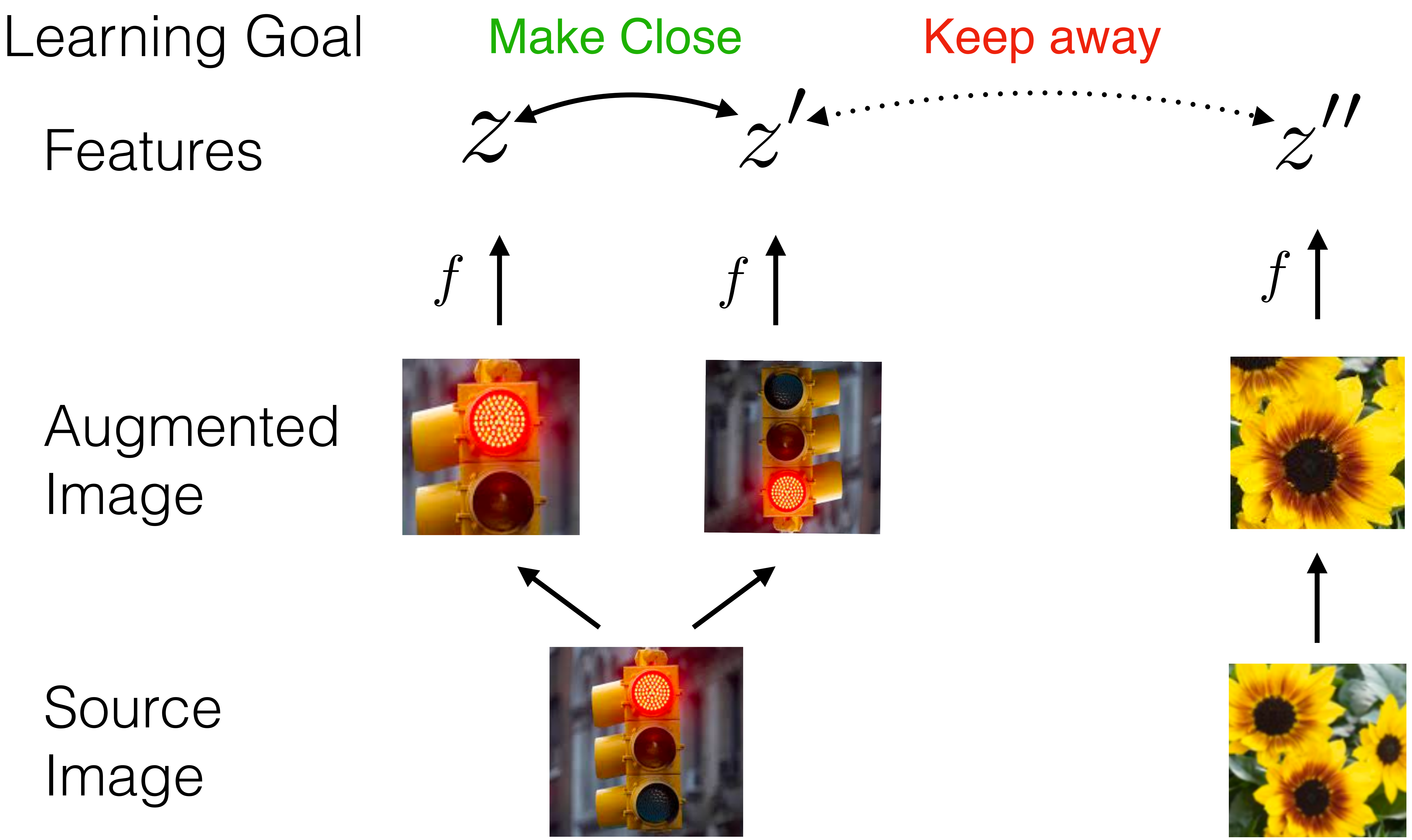


Unlabeled Augmentation Graph



Augmentation Graph with labels

Contrastive Learning learns the augmentation graph.



Spectral Open-world Representation Learning (SORL)

Contrastive loss derived from Matrix Factorization

$$\mathcal{L}_{\mathrm{mf}}(F, A) = \left\| normalize(A) - FF^\top \right\|_F^2$$

↓

$$\mathcal{L}_{sorl}(f) \triangleq -2\alpha\mathcal{L}_1(f) - 2\beta\mathcal{L}_2(f) + \alpha^2\mathcal{L}_3(f) + 2\alpha\beta\mathcal{L}_4(f) + \beta^2\mathcal{L}_5(f)$$

Make Close
Positive Pairs

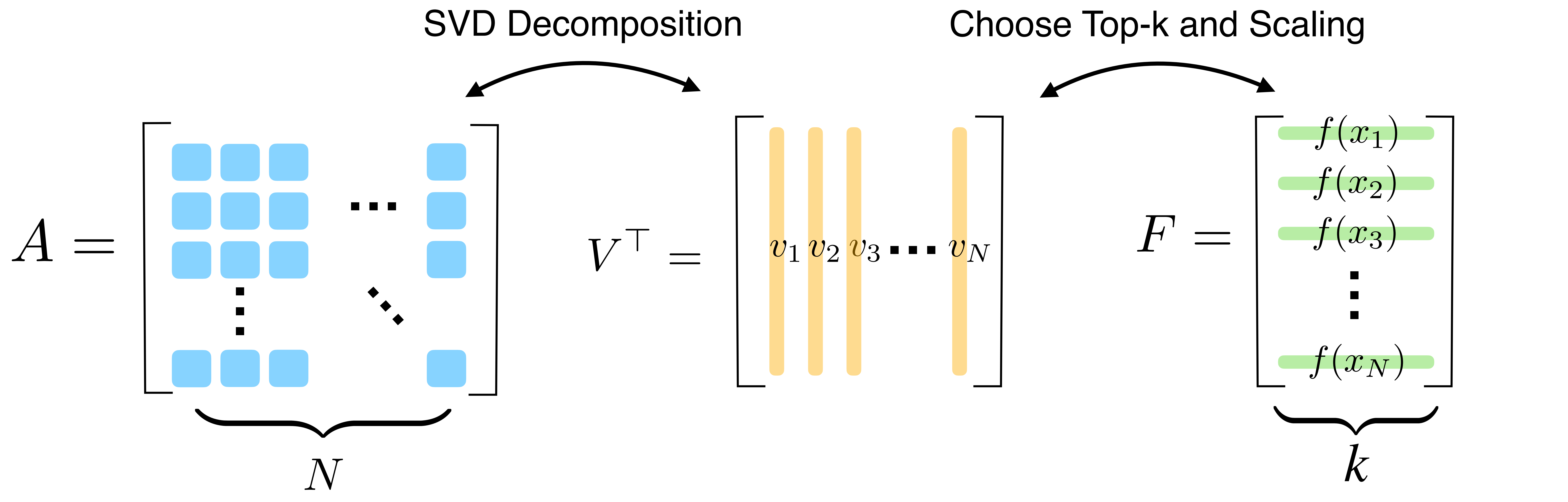
Keep away
Negative Pairs

See more details in paper!

SORL has the closed-form solution.

$$\mathcal{L}_{\mathrm{mf}}(F, A) = \left\| normalize(A) - FF^\top \right\|_F^2$$

Optimal Solution *(Eckart–Young–Mirsky Theorem)*

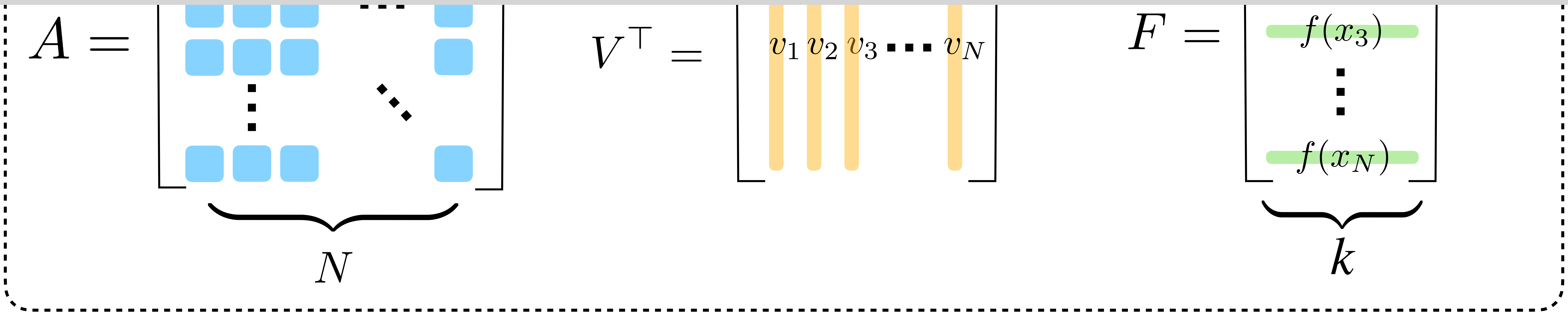


The closed-form solution is known!

$$\mathcal{L}_{\mathrm{mf}}(F, A) = \left\| normalize(A) - FF^{\top} \right\|_F^2$$

Optimal Solution *(Eckart-Young-Mirsky Theorem)*

Good! We can analyze the feature space with **spectral analysis** of the adjacency matrix!



Theory

Main Intuition of the Theorem

Cluster Performance Gain by adding labels for Class c.

Connection from class c to the labeled data.

$$\Delta_{\pi_c}(\delta) = \left(\mathfrak{l}_{\pi_c} - \frac{1}{N}\right) - 2\left(1 - \frac{|\pi_c|}{N}\right)\left(\mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\in\pi_c}\mathbf{z}_i^\top \mathbf{z}_j - \mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\notin\pi_c}\mathbf{z}_i^\top \mathbf{z}_j\right).$$

Intra-class similarity

Inter-class similarity

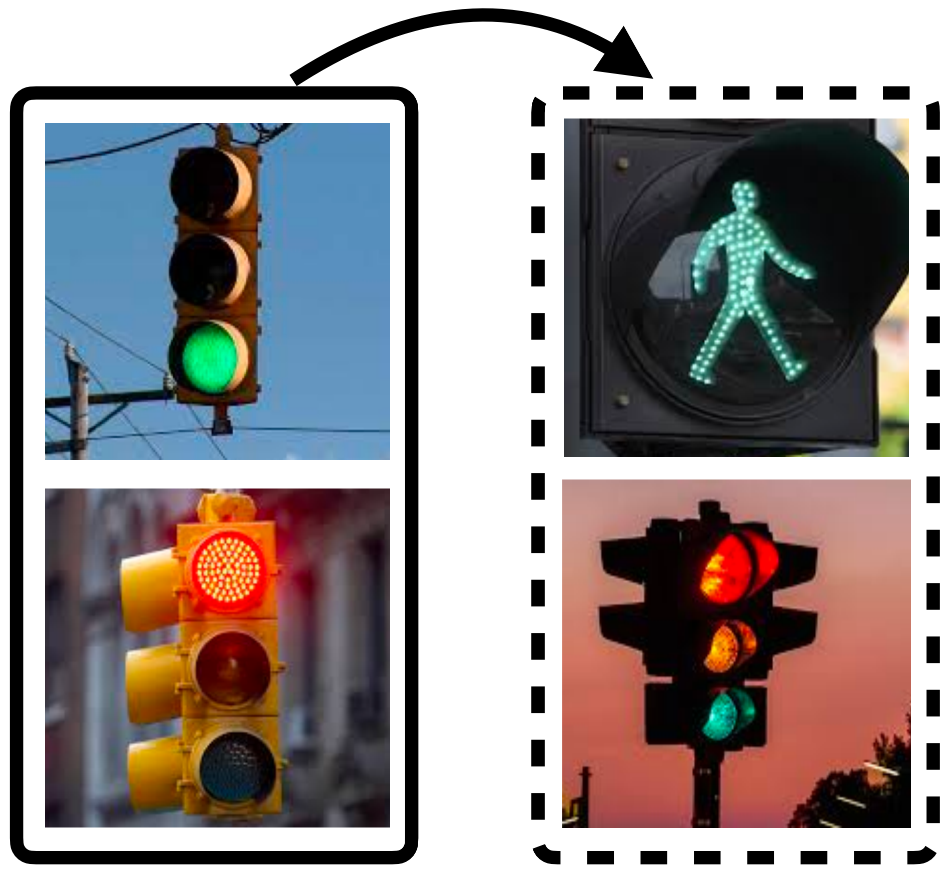
Main Theorem (Case Study)

Connection from class c to the labeled data.

$$\Delta_{\pi_c}(\delta) = \left(\mathfrak{l}_{\pi_c} - \frac{1}{N}\right) - 2\left(1 - \frac{|\pi_c|}{N}\right)\left(\mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\in\pi_c}\mathbf{z}_i^\top \mathbf{z}_j - \mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\notin\pi_c}\mathbf{z}_i^\top \mathbf{z}_j \right).$$

Intra-class similarity ↑   ↑*Inter-class similarity*

Case Study 1 (unlabeled data from known class)**:**



connection >> (intra-sim - inter-sim)

(very large) (…) (…)

Conclusion: Unlabeled traffic lights will be better clustered!

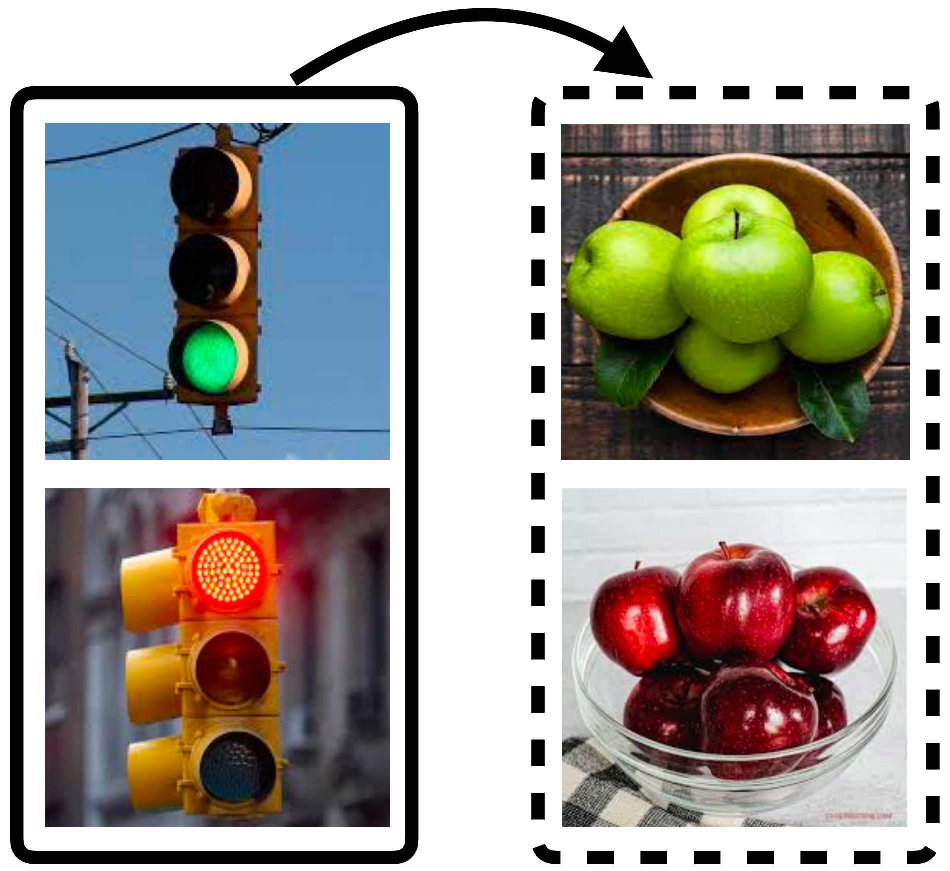
Main Theorem (Case Study)

Connection from class c to the labeled data.

$$\Delta_{\pi_c}(\delta) = (\mathfrak{l}_{\pi_c} - \frac{1}{N}) - 2(1 - \frac{|\pi_c|}{N})(\mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\in\pi_c}\mathbf{z}_i^\top \mathbf{z}_j - \mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\notin\pi_c}\mathbf{z}_i^\top \mathbf{z}_j).$$

Intra-class similarity *Inter-class similarity*

Case Study 2 (novel class with *strong* connection):



connection > (intra-sim - inter-sim)

(large) (Low) (…)

Conclusion: Green and red apple will be close to each other!

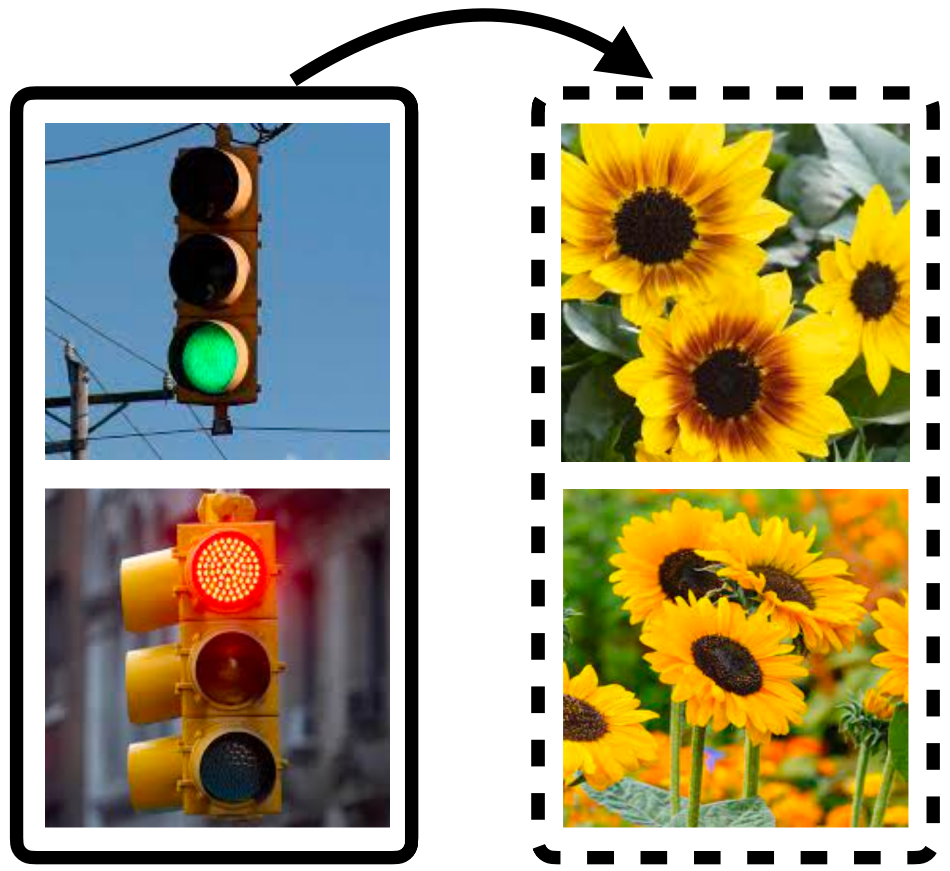
Main Theorem (Case Study)

Connection from class c to the labeled data.

$$\Delta_{\pi_c}(\delta) = \left(\mathfrak{l}_{\pi_c} - \frac{1}{N}\right) - 2\left(1 - \frac{|\pi_c|}{N}\right)\left(\mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\in\pi_c}\mathbf{z}_i^\top\mathbf{z}_j - \mathbb{E}_{i\in\pi_c}\mathbb{E}_{j\notin\pi_c}\mathbf{z}_i^\top\mathbf{z}_j\right).$$

Intra-class similarity ↑ *Inter-class similarity*

Case Study 3 (novel class with *weak* connection):

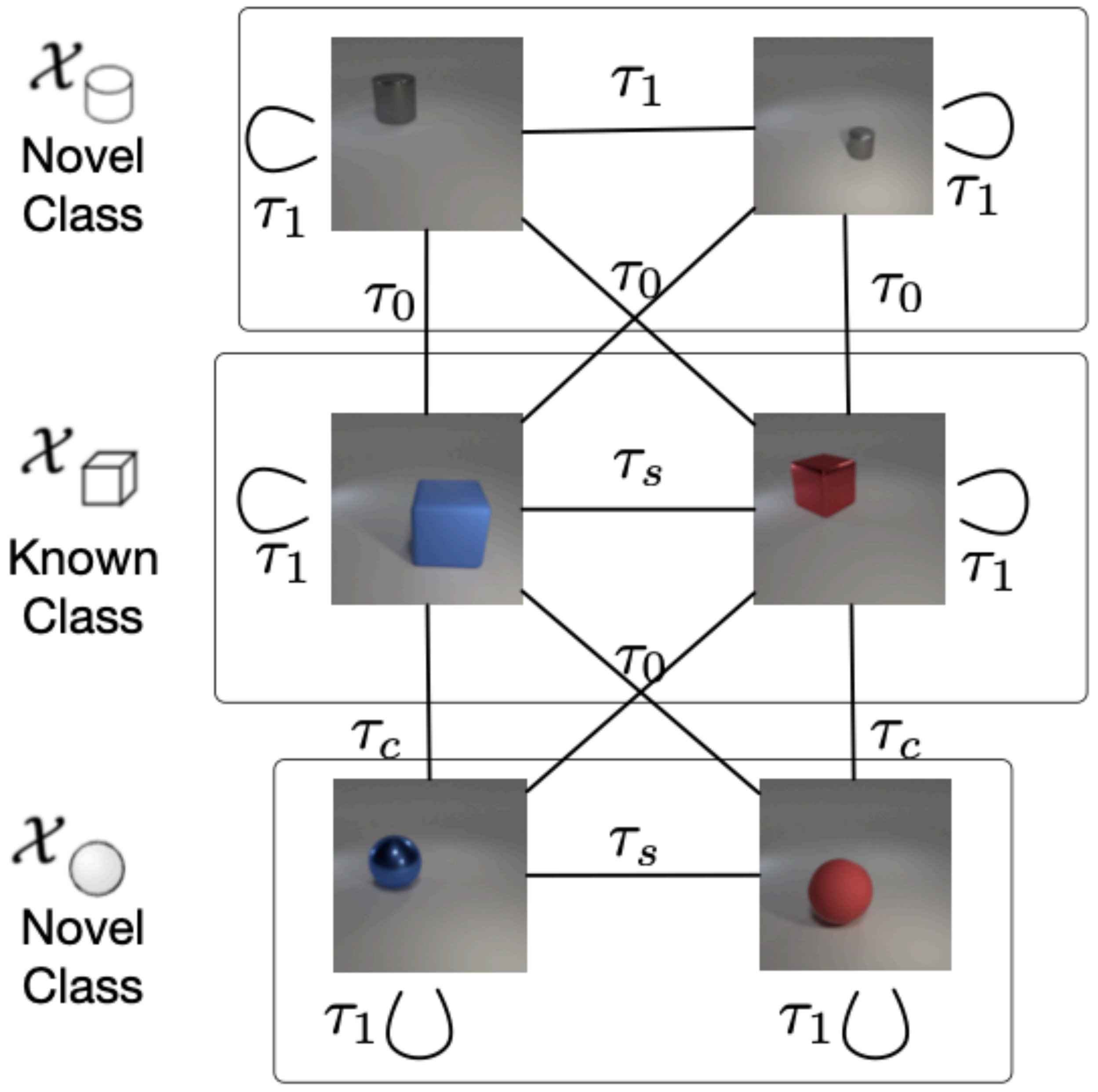


connection < (intra-sim - inter-sim)

(Low) (High) (…)

Conclusion: Add labels may not be beneficial to flower class.

A Toy Example



By Eq. (1,2,3)

$$\eta_u A^{(u)} = \begin{bmatrix} \tau_1^2+\tau_s^2+\tau_c^2 & 2\tau_1\tau_s & 2\tau_1\tau_c & 2\tau_c\tau_s & 0 & 0 \\ 2\tau_1\tau_s & \tau_1^2+\tau_s^2+\tau_c^2 & 2\tau_c\tau_s & 2\tau_1\tau_c & 0 & 0 \\ 2\tau_1\tau_c & 2\tau_c\tau_s & \tau_1^2+\tau_s^2+\tau_c^2 & 2\tau_1\tau_s & 0 & 0 \\ 2\tau_c\tau_s & 2\tau_1\tau_c & 2\tau_1\tau_s & \tau_1^2+\tau_s^2+\tau_c^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\tau_1^2 & 2\tau_1^2 \\ 0 & 0 & 0 & 0 & 2\tau_1^2 & 2\tau_1^2 \end{bmatrix}$$

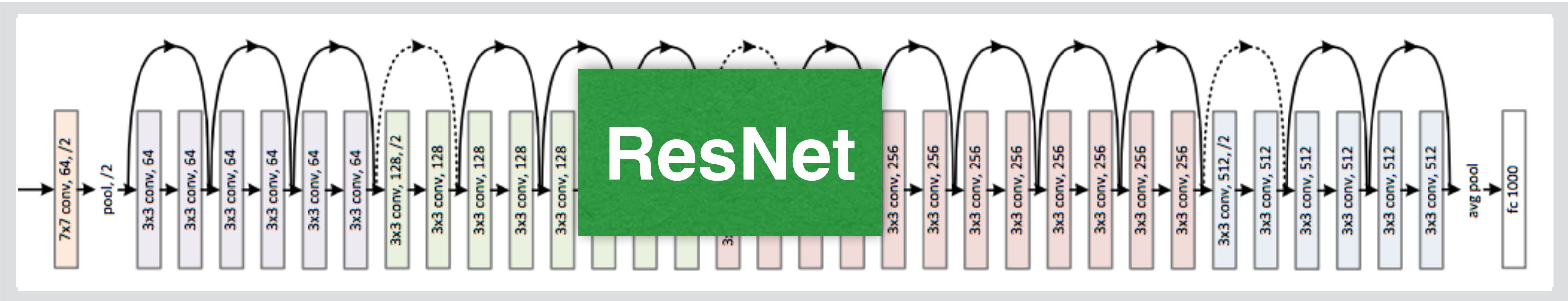
$$A = \eta_u A^{(u)} + \begin{bmatrix} (\tau_1+\tau_s)^2 & (\tau_1+\tau_s)^2 & \tau_c(\tau_1+\tau_s) & \tau_c(\tau_1+\tau_s) & 0 & 0 \\ (\tau_1+\tau_s)^2 & (\tau_1+\tau_s)^2 & \tau_c(\tau_1+\tau_s) & \tau_c(\tau_1+\tau_s) & 0 & 0 \\ \tau_c(\tau_1+\tau_s) & \tau_c(\tau_1+\tau_s) & \tau_c^2 & \tau_c^2 & 0 & 0 \\ \tau_c(\tau_1+\tau_s) & \tau_c(\tau_1+\tau_s) & \tau_c^2 & \tau_c^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

See more details in paper!

Experiment

Set Up

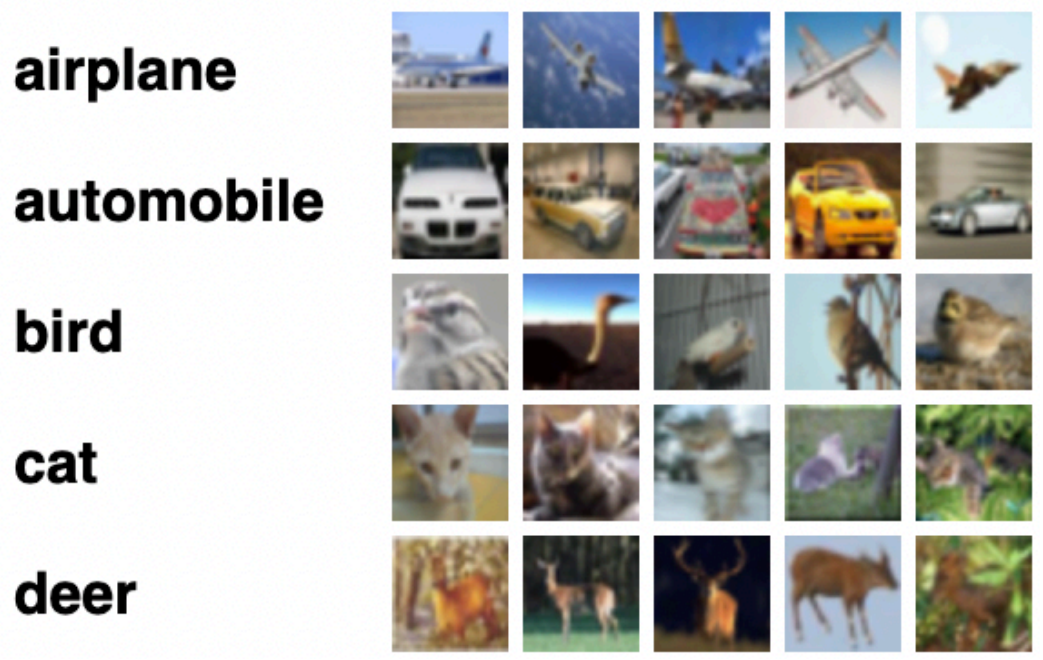
Model



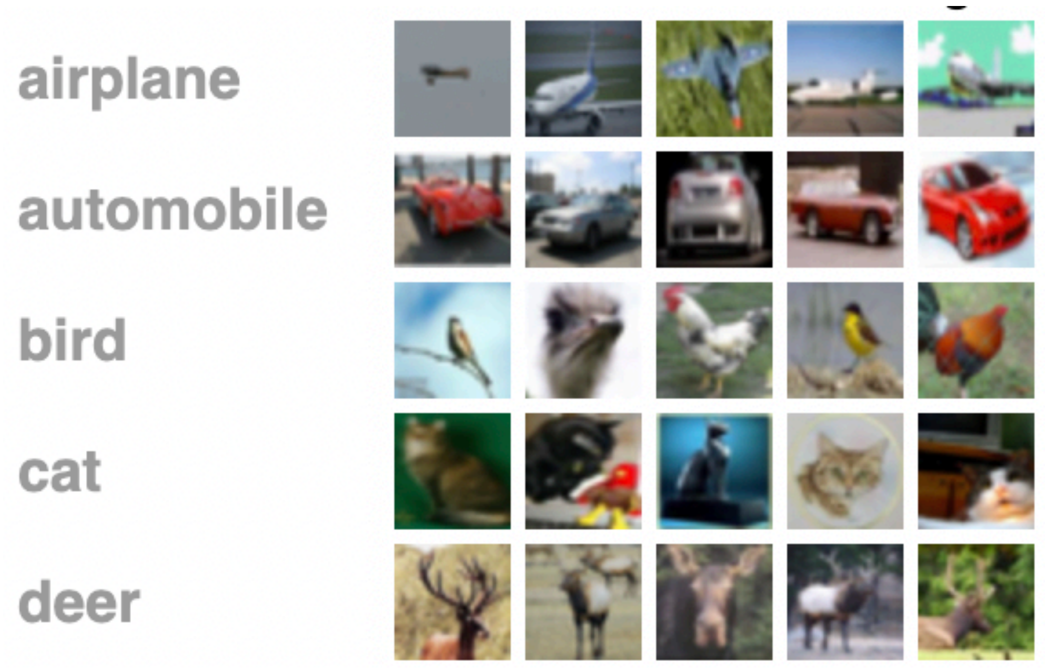
Dataset (CIFAR-10/100)

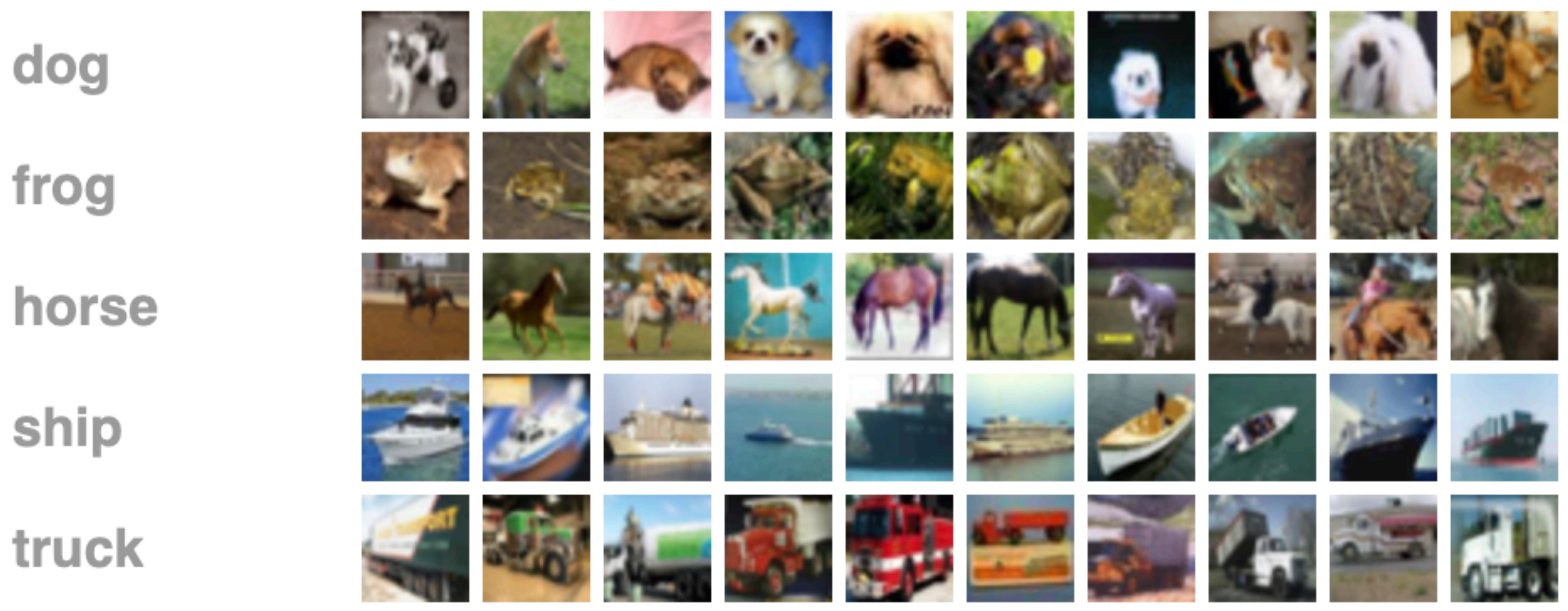
1. Separate all classes into 50% known and 50% novel.
2. Divide known-class samples into 50% labeled and 50% unlabeled.

CIFAR-10 (Labeled Data)

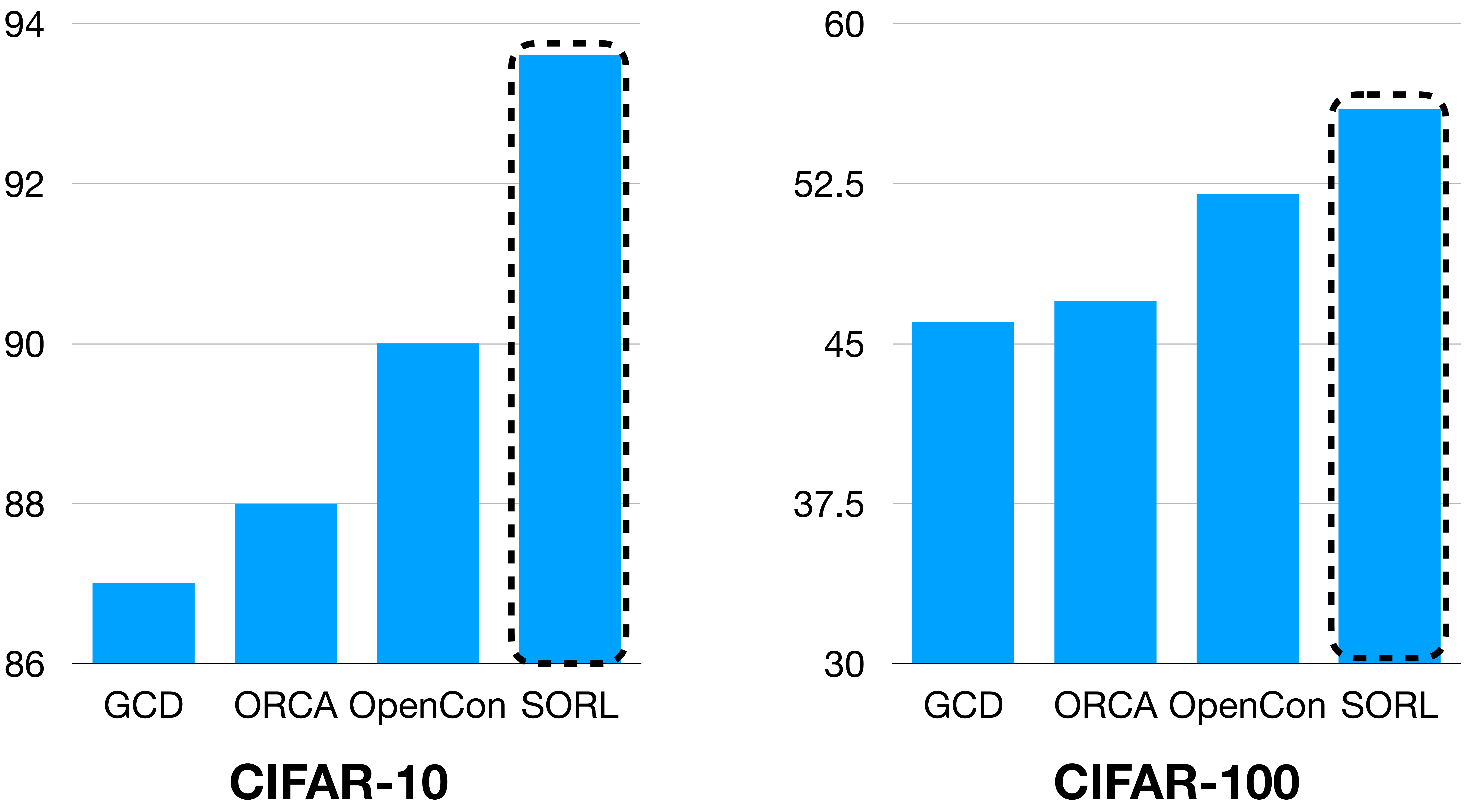


CIFAR-10 (Unlabeled Data)





SORL is also appealing for practical usage!



Thank you!

Our code is available at

https://github.com/deeplearning-wisc/SORL.