Image Clustering Conditioned on Text Criteria

Sehyun Kwon, Jaeseung Park, Minkyu Kim, Jaewoong Cho, Ernest K. Ryu and Kangwook Lee

ICLR 2024

KRAFTON

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

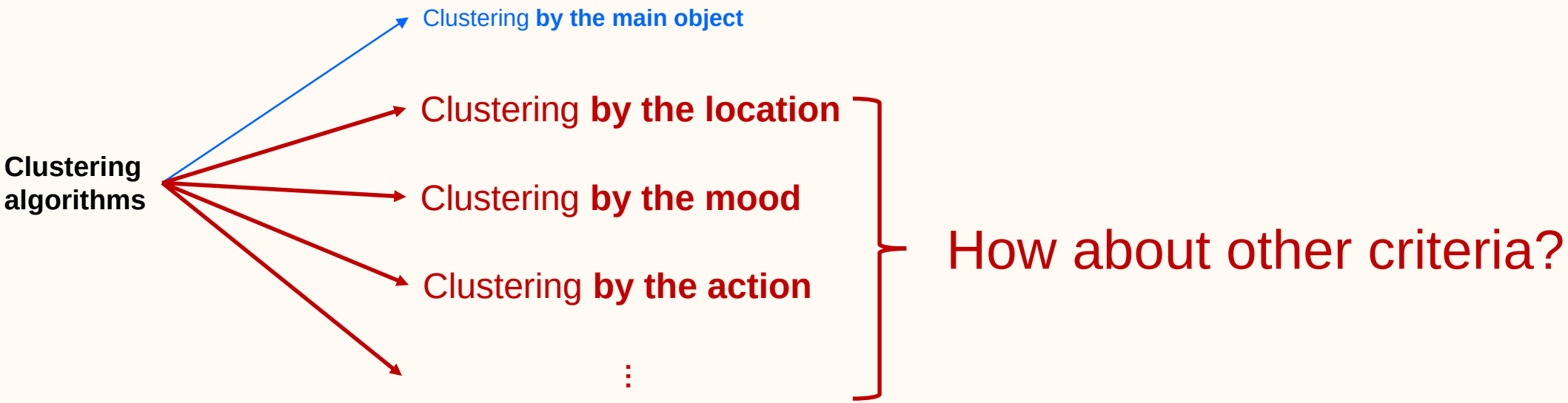
Motivation

Ex) CIFAR-10, CIFAR-100, STL-10, ImageNet, …

Recent image clustering algorithms work well on datasets similar to ImageNet, most of which are labeled based on **the main object** in the image, such as dogs 🐕, cats 🐈 , cars 🚗 , airplanes ✈️, etc.

Clustering algorithms → Work well !! → Clustering **by the main object**

Motivation



Unfortunately, all existing image clustering algorithms share a **limitation** - you **can't directly specify** the clustering criterion. Instead, you **indirectly specify** it by choosing a *feature extractor*, *distance metric*, etc.

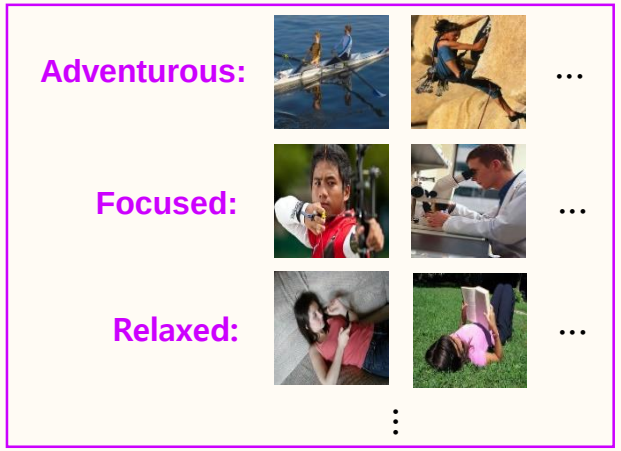
Ex) In contrastive learning, carefully designed data augmentation is needed. And this is very non-trivial.

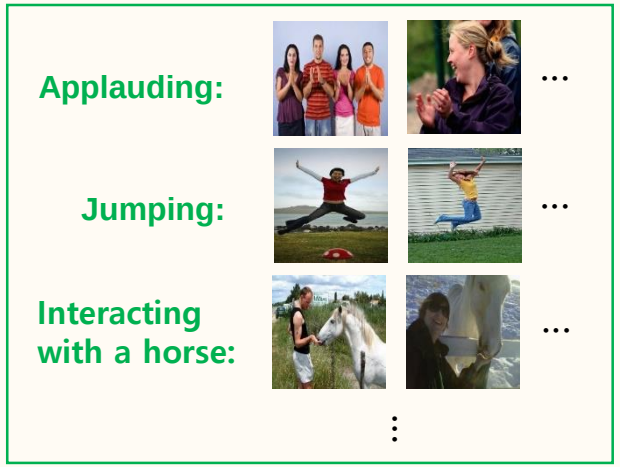
IC|TC: Image Clustering Conditioned on Text Criteria

IC|TC can specify the clustering criterion by ***words***!

*"Cluster these images by based on the **mood**"*

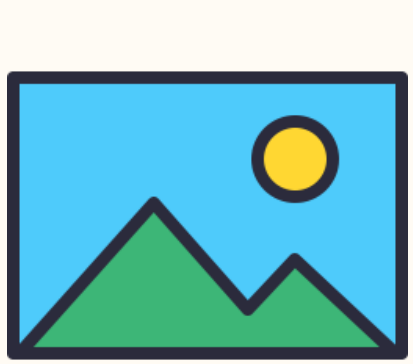
*"Cluster these images by based on the **action**"*





Method (Step 1)

Given clustering criterion, transforming images to image descriptions!



Image

"Describe the image based on {criterion}"

VLM

Image description

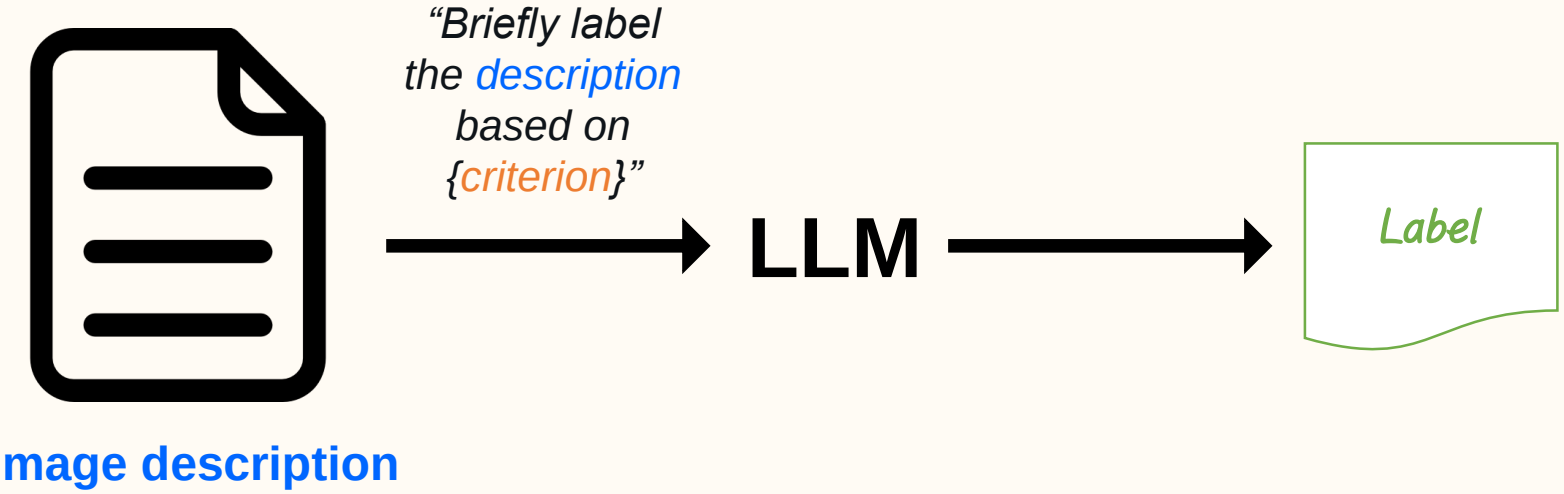
Example:

"Describe the image based on {Action}"

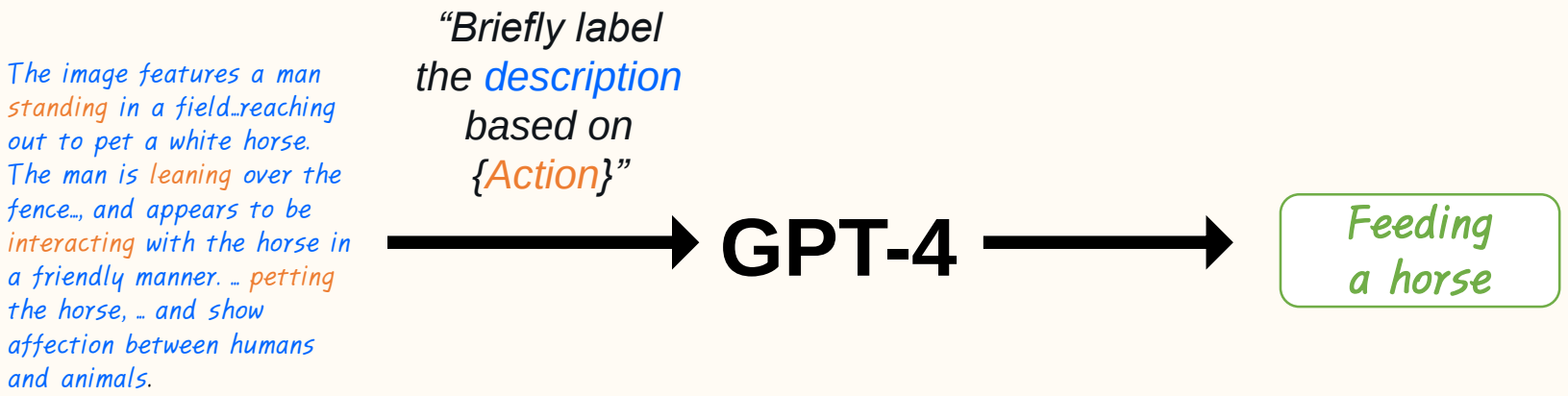
LLaVA

The image features a man standing in a field…reaching out to pet a white horse. The man is leaning over the fence…, and appears to be interacting with the horse in a friendly manner. … petting the horse, … and show affection between humans and animals.

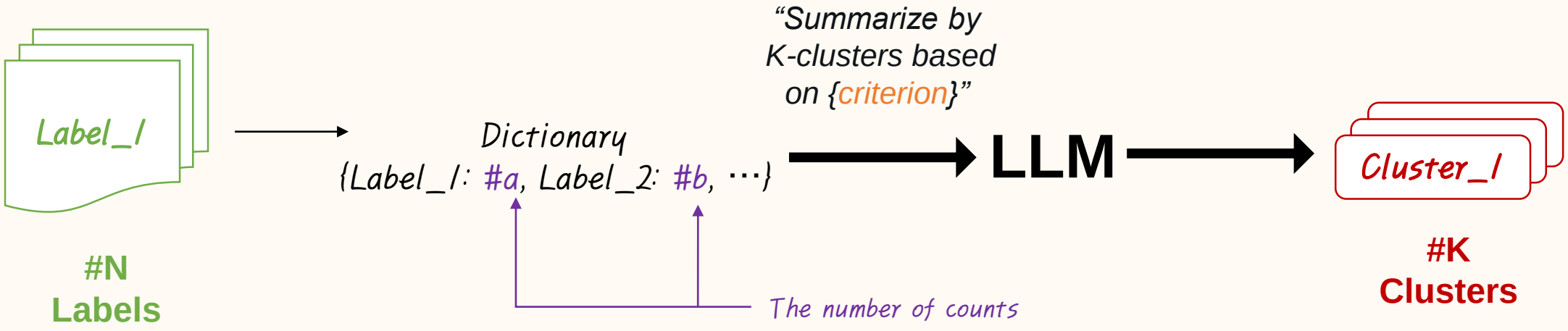
Method (Step 2a) Reducing the amount of description through raw labeling!



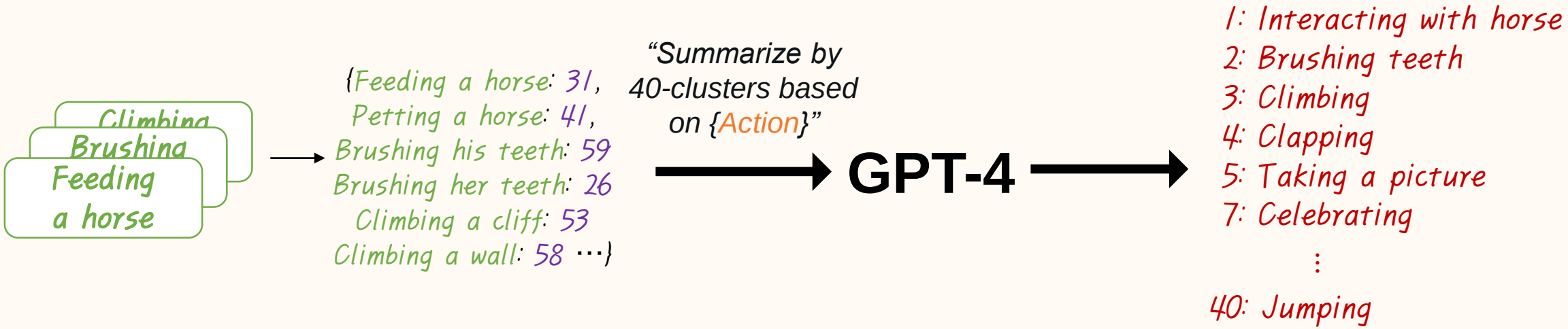
Example:



Method (Step 2b) Obtaining cluster names by LLM!

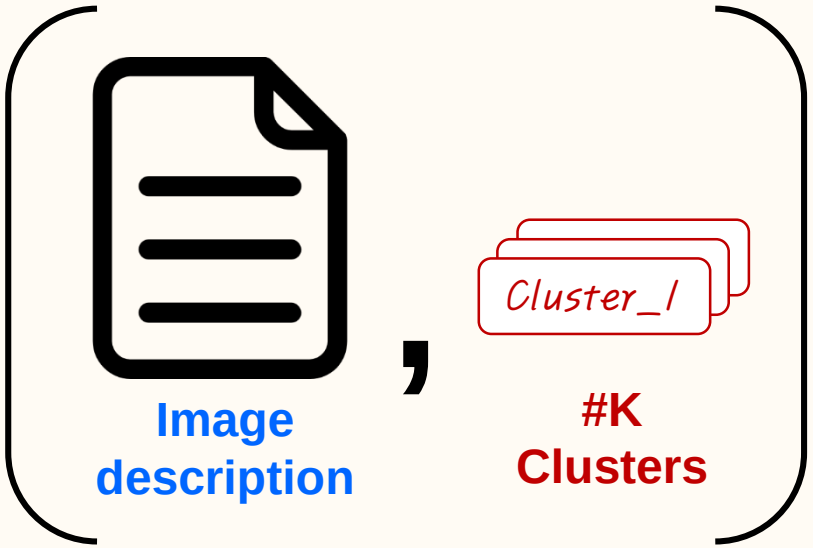


Example:



Method (Step 3)

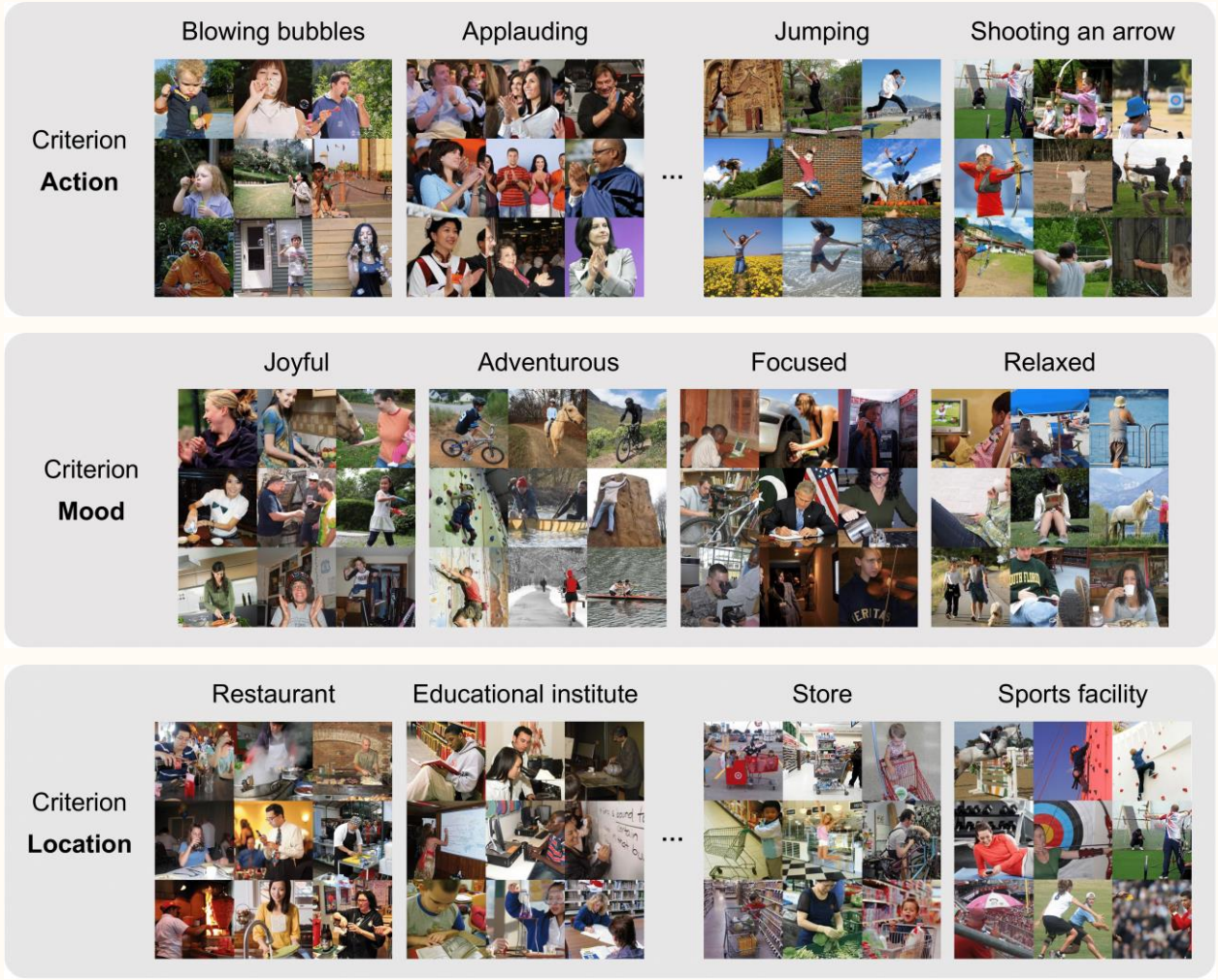
Classifying image description to appropriate cluster name by LLM!



Example:

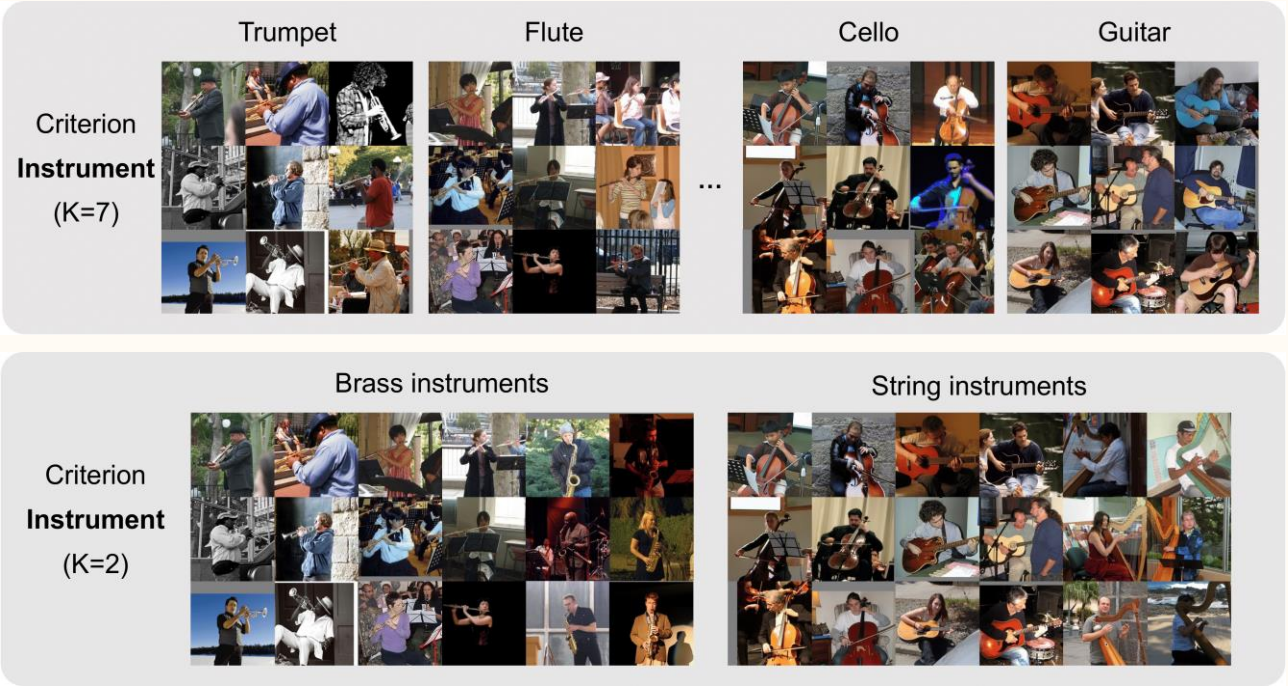
Experiments - Clustering by varying criteria

IC|TC can do the clustering by varying clustering criterion at Stanford-40-Actions dataset.



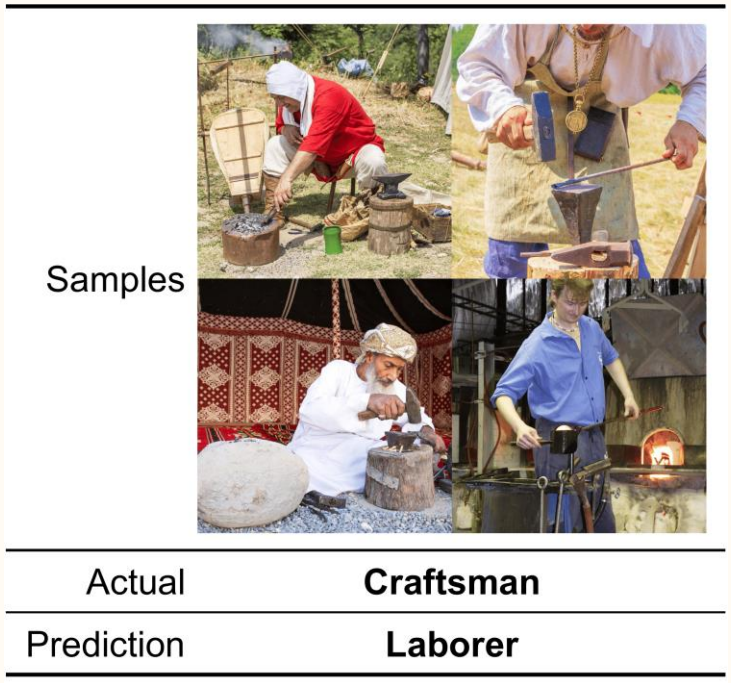
Experiments - Clustering by varying the number of clusters

IC|TC automatically find the hierarchical clustering by varying the number of clusters at People-Playing-Musical-Instrument (PPMI) dataset.

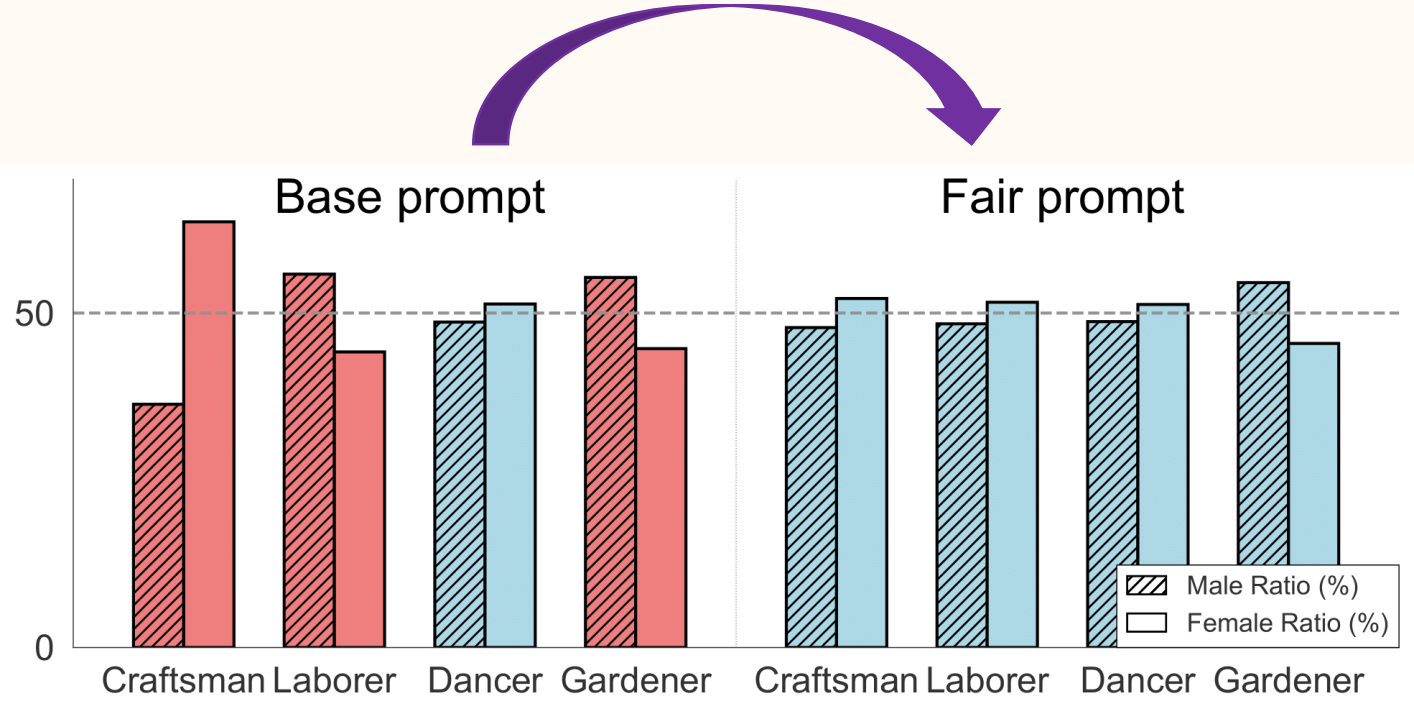


Experiments – Fair clustering by text criterion refinement

IC|TC can easily edit the clustering result when user don't like it.
And this becomes easily achievable through text criterion refinement.



Simply adding a prompt: *"Do not consider gender"*



Other results

For more interesting results and details, please check our paper!

- Comparison between existing clustering algorithms.
- Effect from the choice of VLMs and LLMs.
- Prompt design.
- Large scale experiment.
- Etc.