COMP5121

Data Mining and Data Warehousing Applications

Week 10: Clustering with Deep Learning Models

Dr. Fengmei Jin

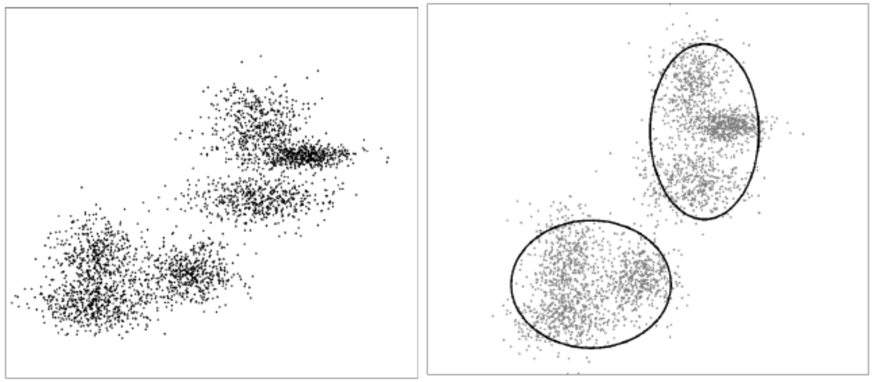
- Email: fengmei.jin@polyu.edu.hk
- Office: PQ747 (+852 3400 3327)
- Consultation Hours: 2.30-4.30pm every Thursday

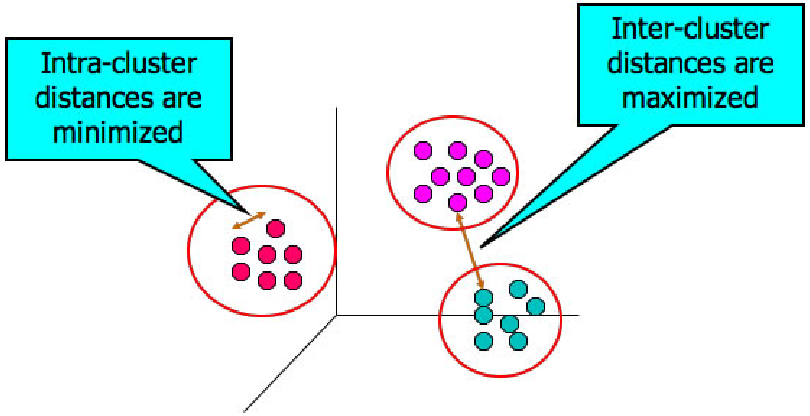
Outline

- Clustering *vs* Deep Clustering
- Unsupervised, Autoencoder-based Clustering
- Self-supervised, Contrastive-based Clustering

Clustering

- **Goal**: group objects into clusters *s.t.*
  - data objects within the same clusters are close to each other, while ensuring that different clusters are well-separated





- Most clustering algorithms rely on distance functions.
  - Distance measures (e.g., Euclidean) are meaningless in high-dimensional space
    - distance to nearest neighbors

    vs distance to farthest ones

Why Deep Clustering?

- **High-Dimensional Data**
  - Classical clustering (e.g., *k*-Means) struggles when data is very high-dimensional (images, text, etc.).
  - DNNs learn more compact, meaningful representations.
- **Nonlinear Feature Extraction**
  - Neural networks capture complex structures/patterns that are hard for traditional clustering algorithms to uncover.
- **Integrated Learning, Training, and Optimization**
  - The ability to train feature extraction and clustering objectives jointly often leads to better performance than a two-step "feature extraction → clustering" pipeline.

Deep Clustering

- Deep neural networks extract important features and capture the underlying structure of data → *encoded representations that make clustering easier and more meaningful*

- Typical Workflow
  1. **Representation learning** (embedding): map high-dimensional data into a lower-dimensional latent space
  2. **Clustering in latent space**: apply clustering algorithms (e.g., k-means) to the learned embedding
  3. ***Iterative refinement**: some methods alternate between feature learning and cluster assignment to iteratively improve both

(I) Unsupervised, Autoencoder-based Clustering

- **Core Idea**: Use an autoencoder to compress data into a latent space, then cluster those latent embeddings.
- **Advantages**: Keeps a reconstruction objective that preserves important data characteristics.

- **Representative Methods:**

  [1] DEC (Deep Embedded Clustering)

  [2] VaDE (Variational Autoencoder for Deep Embedding)

  [3] IDEC (Improved Deep Embedded Clustering)

[1] Xie, Junyuan, et al. "Unsupervised deep embedding for clustering analysis." ICML, 2016.
[2] Jiang, Zhuxi, et al. "Variational deep embedding: An unsupervised and generative approach to clustering." IJCAI, 2017.
[3] Guo, Xifeng, et al. "Improved deep embedded clustering with local structure preservation." IJCAI, 2017.

(II) Self-supervised, Contrastive-based Clustering

- **Core Idea**: Learn latent representations by pushing apart dissimilar data samples and pulling together similar ones — often via data augmentations.
- **Advantages**: Does not require labeled data; learned features tend to be robust and highlight semantic structure.

- **Representative Methods:**

  [1] DeepCluster: Iterative clustering + CNN training

  [2] SwAV: Online “clusters” (prototypes) plus contrastive ideas

  [3] SCAN: Contrastive learning via nearest-neighbor clustering

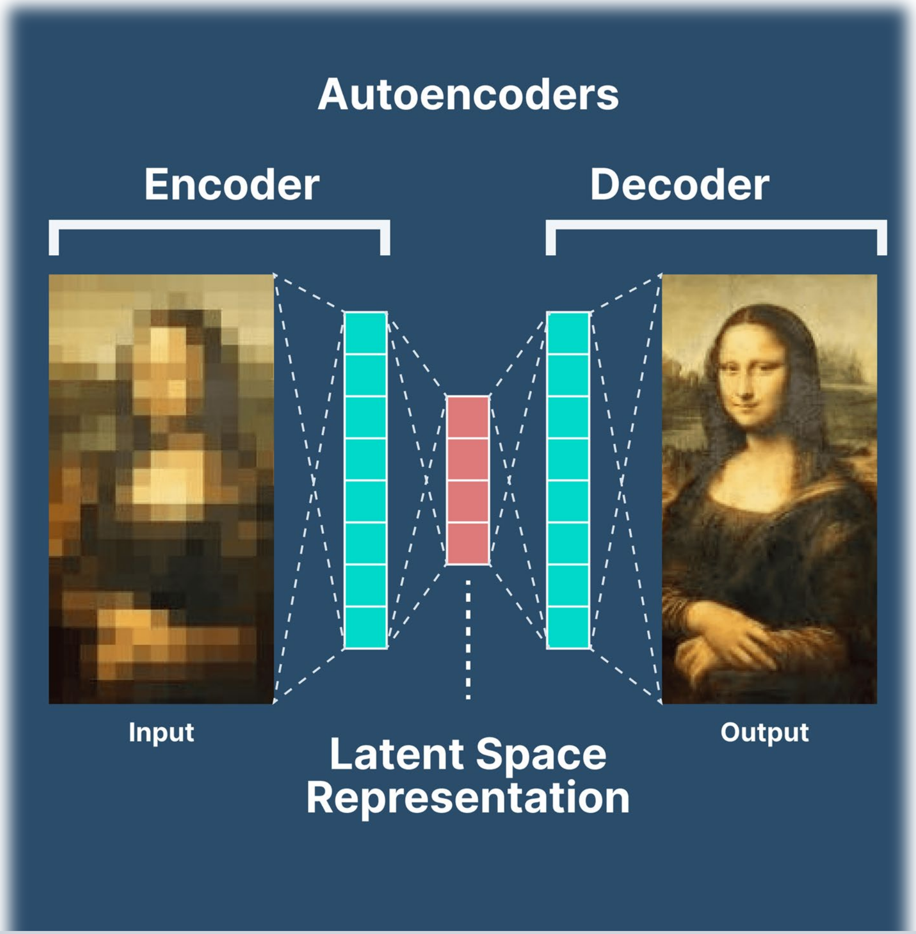
[1] Caron, Mathilde, et al. "Deep clustering for unsupervised learning of visual features." ECCV, 2018.
[2] Caron, Mathilde, et al. "Unsupervised learning of visual features by contrasting cluster assignments.‟ NEURIPS, 2020.
[3] Van Gansbeke, Wouter, et al. "Scan: Learning to classify images without labels." ECCV, 2020.

UNSUPERVISED, AUTOENCODER-BASED CLUSTERING

Autoencoder for Representation Learning

- A neural network used to learn compact (lower-dimensional), informative embedding in an unsupervised way
  - encoder-decoder architecture

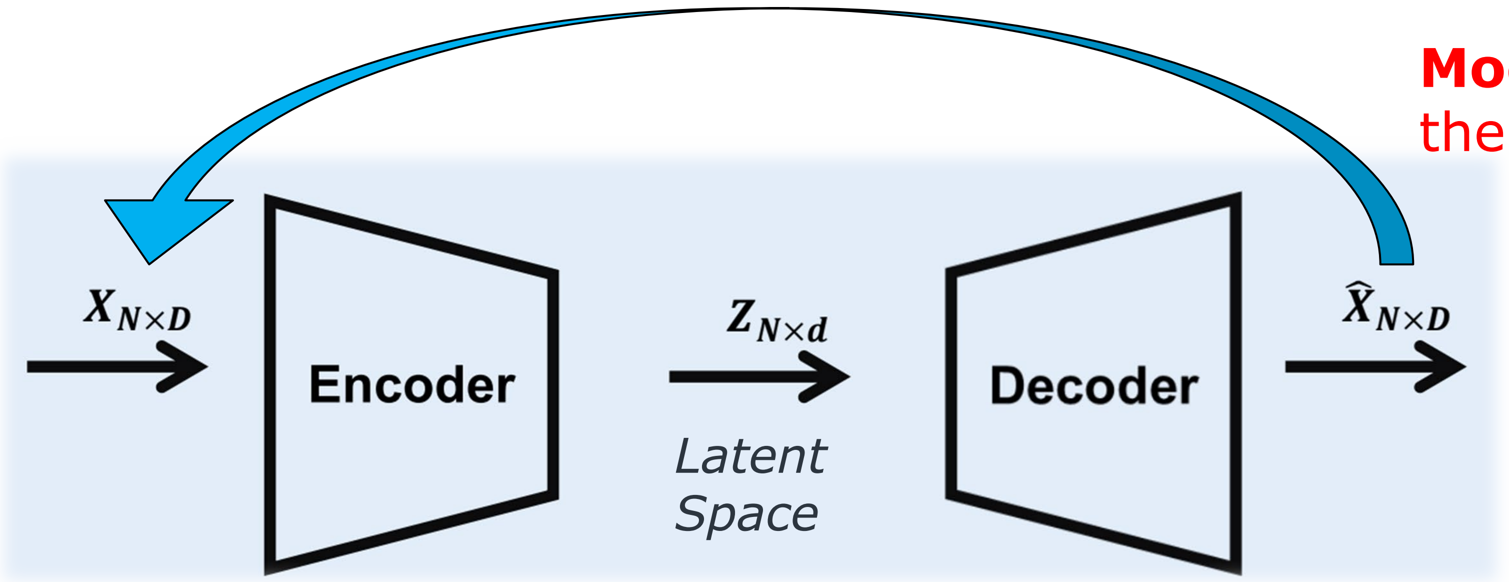
- Train the network to capture the most important patterns and reconstruct input data



The prediction (output) is a reconstruction of the input data.

Autoencoder: Architecture

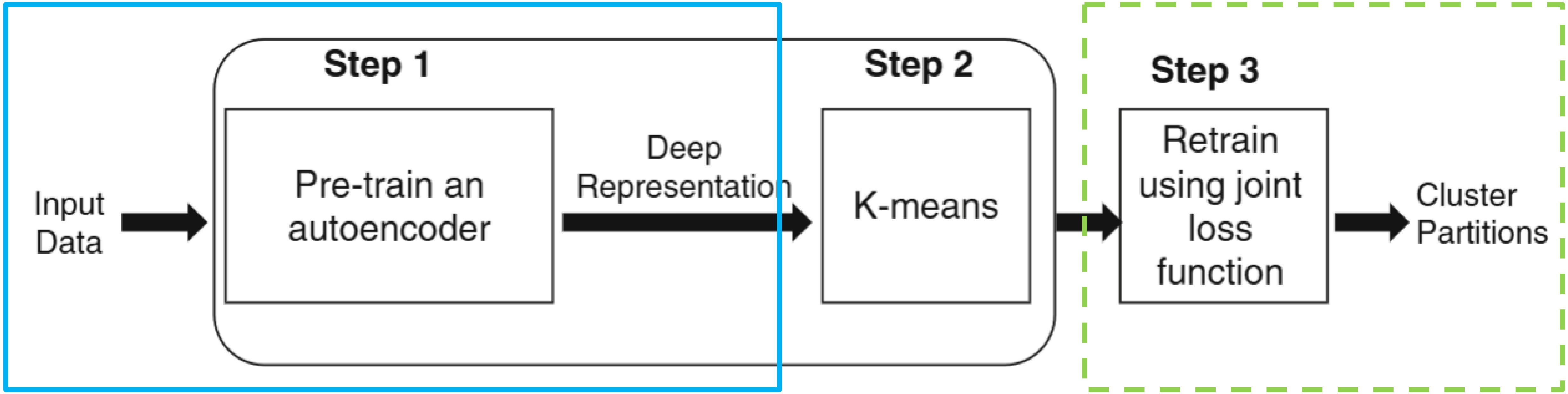
- **Encoder**: A module that "compress" input data into a latent space with lower dimensions
- **Decoder**: A module that "decompress" the embeddings and "reconstruct" the data back



Model Training: Minimize the reconstruction loss

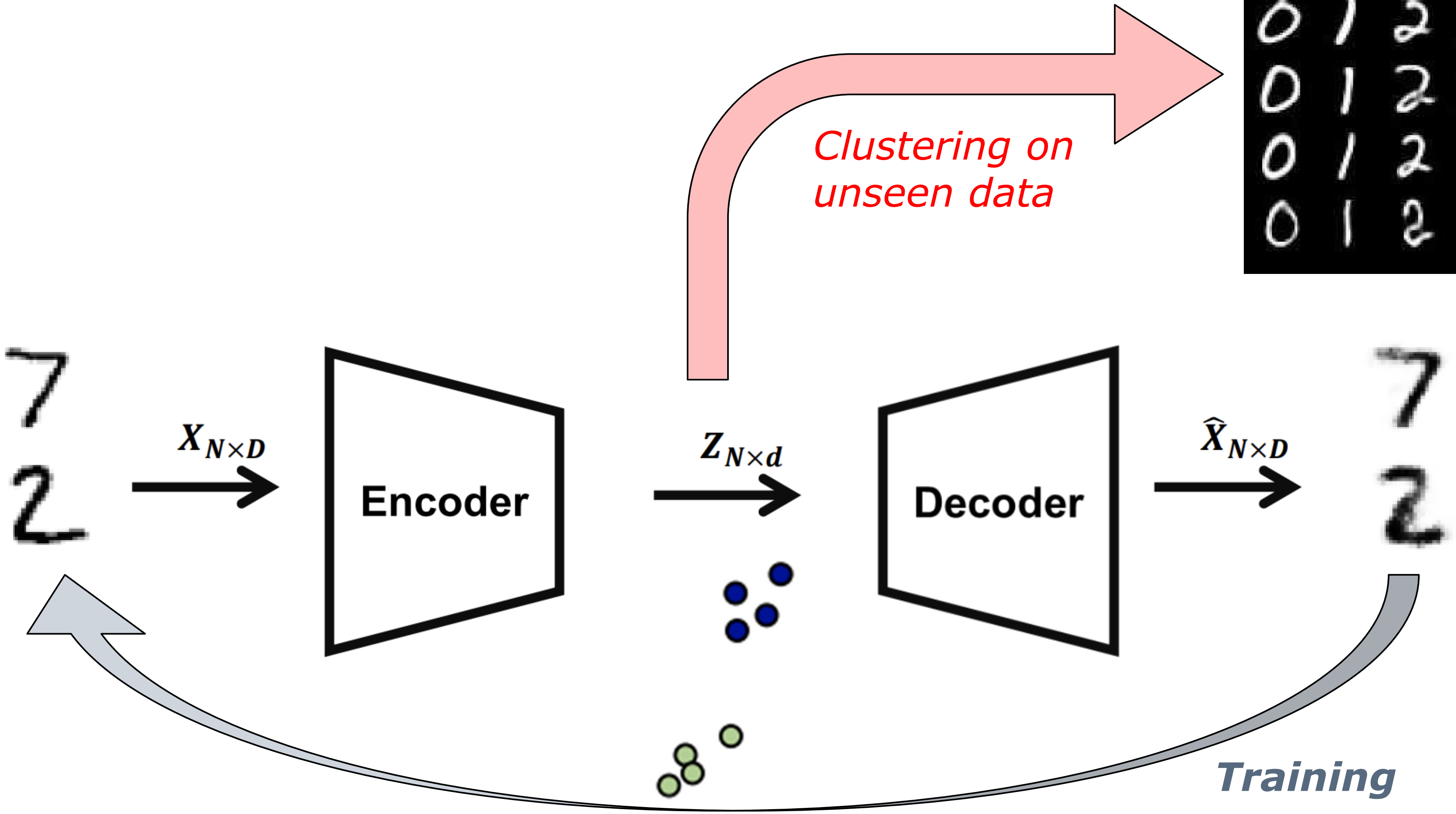
- **Input**: N data points in the D-dimensional space
- **Output**: cluster partitions

Autoencoder for Clustering



Joint optimization
- NN parameters
- cluster parameters

Autoencoder: Toy Example



Autoencoder-based Clustering

- **Learning Objectives:** minimize reconstruction error while optimizing cluster assignments.

- **Strengths:**
  - Preserves global structure and data semantics
  - Stable training dynamics and efficient for smaller datasets
  - Performs well on structured data (e.g., tabular, time series)
- **Limitations:**
  - Struggles with irrelevant details
  - Requires careful architecture design

SELF-SUPERVISED, CONTRASTIVE-BASED CLUSTERING

Overview of Learning Paradigms

- **Supervised Learning**: Requires labeled data to learn *mapping*
  - Example: image classification (Image → Class label), sentiment analysis (Text → Sentiment label)
- **Unsupervised Learning**: No labels are provided, focusing on discovering patterns or structures in the data – Clustering
- **Semi-Supervised Learning**: Mix of labeled & unlabeled data
  - To use a small set of labels and a large pool of unlabeled data
- **Self-Supervised Learning**: Generates its own “labels” directly from the data itself, with no need for external annotation
  - Example: predicting masked parts of an image or sentence

Contrastive Learning

- A self-supervised learning method that learns representations by contrasting similar (positive) and dissimilar (negative) data
  - Encourage similar data points to have closer representations
  - Pushes apart dissimilar data points in the feature space

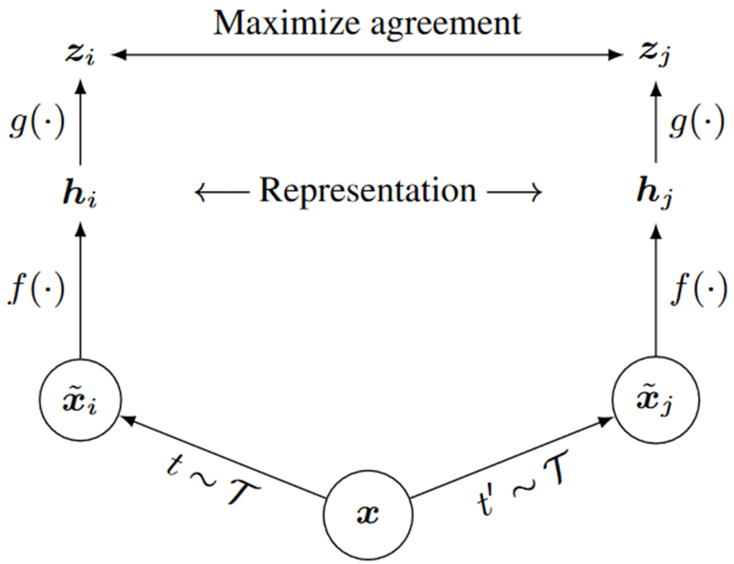
- **Representative methods:**

  [1] SimCLR: Maximizes agreement between augmented views of the same image.

  [2] MoCo: Builds a memory bank for negative samples to improve contrastive learning, especially for large-scale data

[1] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." ICML, 2020.
[2] He, Kaiming, et al. "Momentum contrast for unsupervised visual representation learning." CVPR, 2020.

Data Augmentation in SimCLR



A Simple Framework for Contrastive Learning of Visual Representations

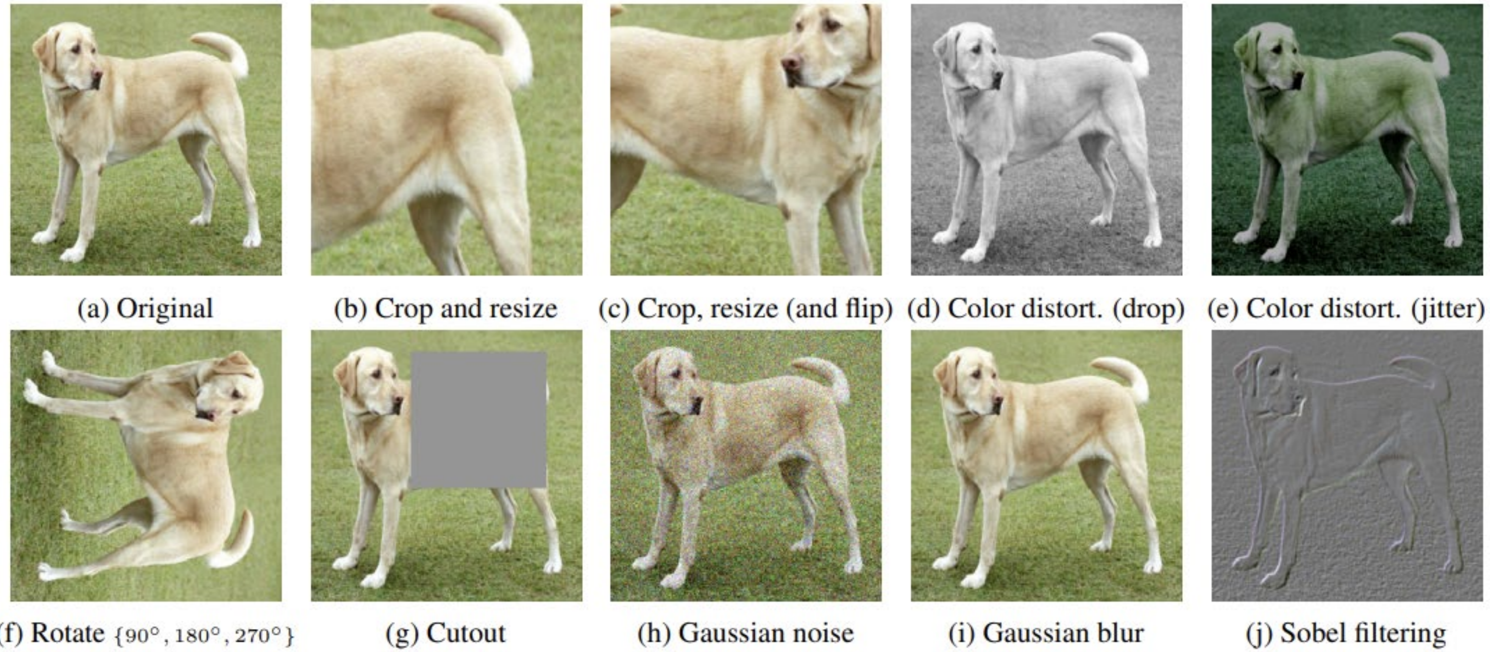


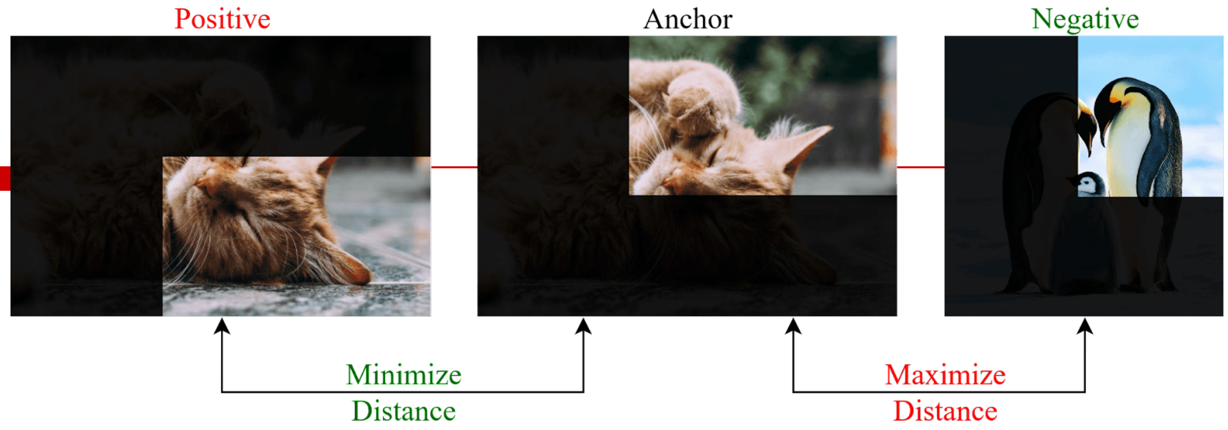
Figure 4. Illustrations of the studied data augmentation operators. Each augmentation can transform data stochastically with some internal parameters (e.g. rotation degree, noise level). Note that we *only* test these operators in ablation, the *augmentation policy used to train our models* only includes *random crop (with flip and resize)*, *color distortion*, and *Gaussian blur*. (Original image cc-by: Von.grzanka)

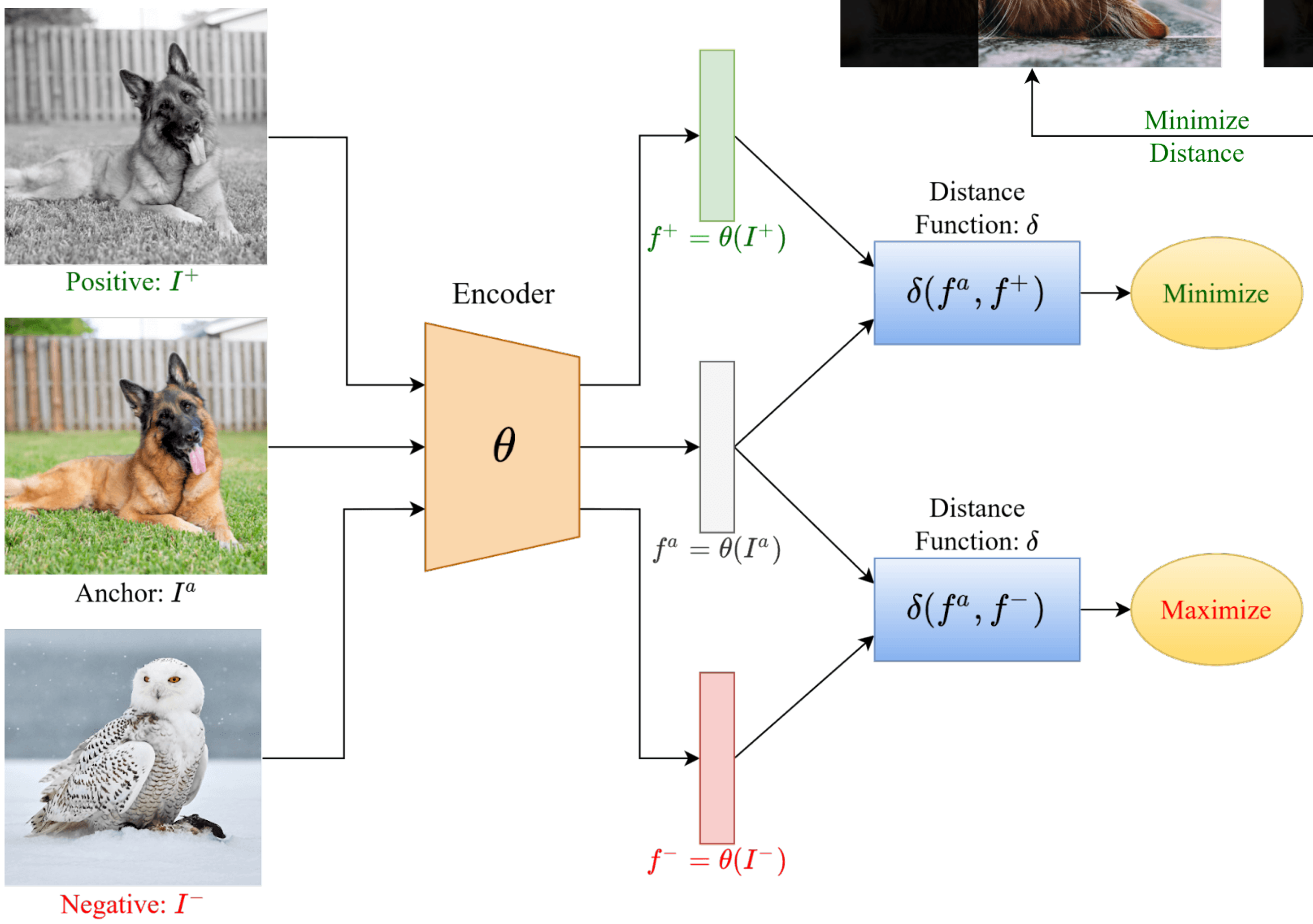
[1] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." ICML, 2020.

Contrastive Learning

Positive: diverse views of the same data points

Negative: views of different ones





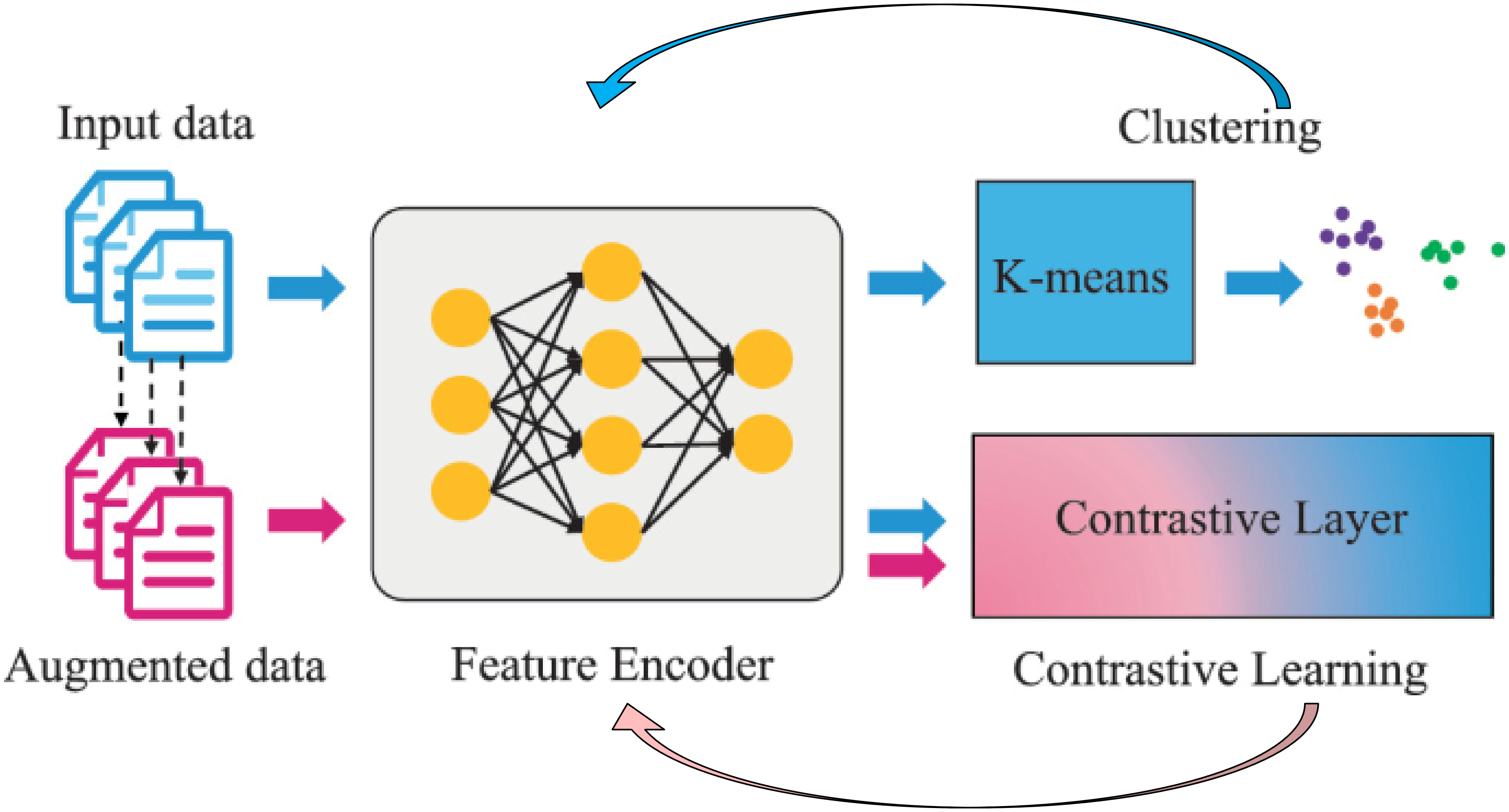
https://www.v7labs.com/blog/contrastive-learning-guide

From Contrastive Learning to Clustering

- Contrastive-Based Clustering
  - Combines the power of CL with clustering techniques
  - Learns robust representations that align well with clustering objectives: similar data points are close & well-separated clusters
  - Does not require explicit labels

- **Typical Workflow**
  - Pretrain a model using a CL framework (e.g., SimCLR)
  - Use the learned representations for clustering (e.g., K-means)
  - Optionally refine the clusters iteratively using joint objectives

Contrastive Clustering



Wu, Yang, et al. "Unsupervised Clustering with Contrastive Learning for Rumor Tracking on Social Media." CCF International Conference on Natural Language Processing and Chinese Computing. Cham: Springer Nature Switzerland, 2023.

Summary

- **Deep Clustering:** handles high-dimensional data through representation learning (without relying on distance measures)
- **Autoencoder-based Clustering**: learns low-dimensional embeddings by encoding and reconstructing data
  - Preserves global structure and is effective for structured data.
  - Jointly minimizes reconstruction and clustering objectives.
- **Contrastive-based Clustering:** learns robust representations by contrasting similar (positive) and dissimilar (negative) pairs.
  - Forms a feature space ideal for clustering.
  - No need for explicit labels and generalizes well across tasks.

Email: fengmei.jin@polyu.edu.hk

Office: PQ747

THANK YOU!