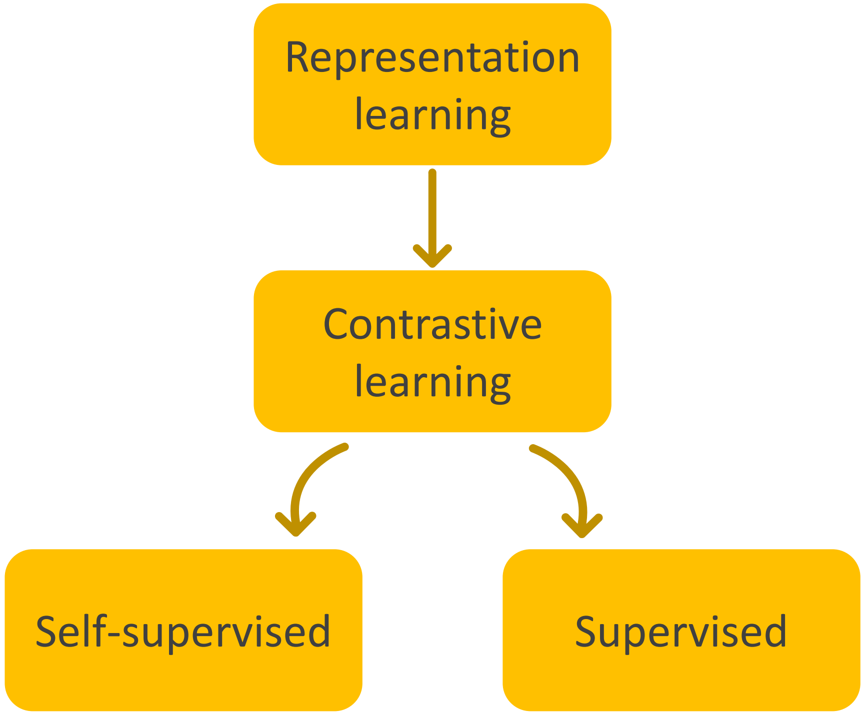
Cassandra Durr

Contrastive Learning: Fundamentals and Practical Applications

Introduction to Representation Learning

Representation Learning

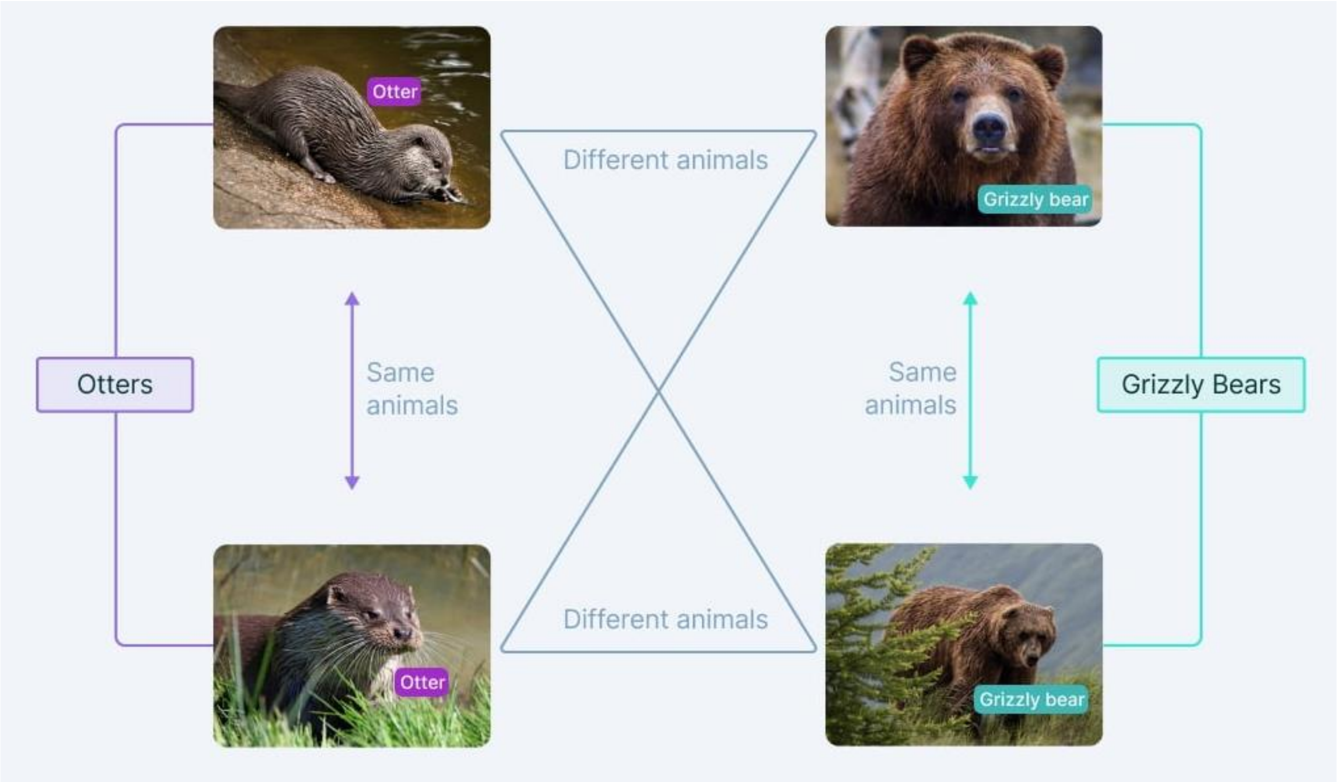
- Automatically discovers feature patterns in data
- Encodes features into numerical representations
- Performance is measured by performance on downstream tasks.



Introduction to Contrastive Learning

Contrastive Learning

- Encoding data to a lower-dimensional latent space (embedding space)
- Distinguishing positive instances (same class) and negative instances (different classes)
- Pushing together similar encodings and pulling apart dissimilar encodings



[1]

Self-supervised contrastive learning

Self-supervised contrasting learning

- Exploits unlabeled data by creating 'pseudo-labels' from input data properties.

Creating positive and negative instances in computer vision

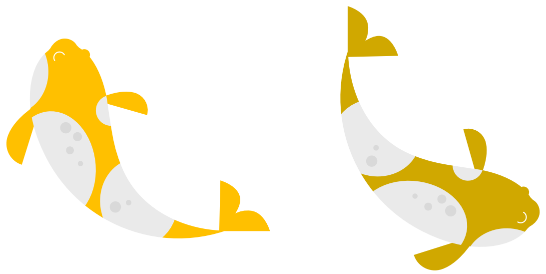
- Positive instances are generated by applying augmentations to anchor images or extracting patches from the same image.
- Negative instances are other images or patches from different images.
- We will discuss augmentation options in other fields later.

Pre-training

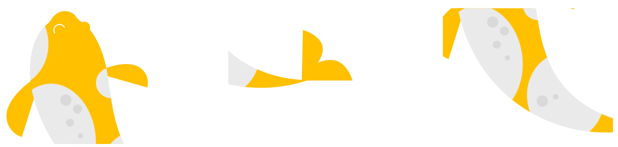
- Ideal for pre-training as it leverages abundant unlabeled data to learn object representations; can later be fine-tuned with smaller labeled datasets.

Creating positive instances in computer vision

- augmentation



- patches

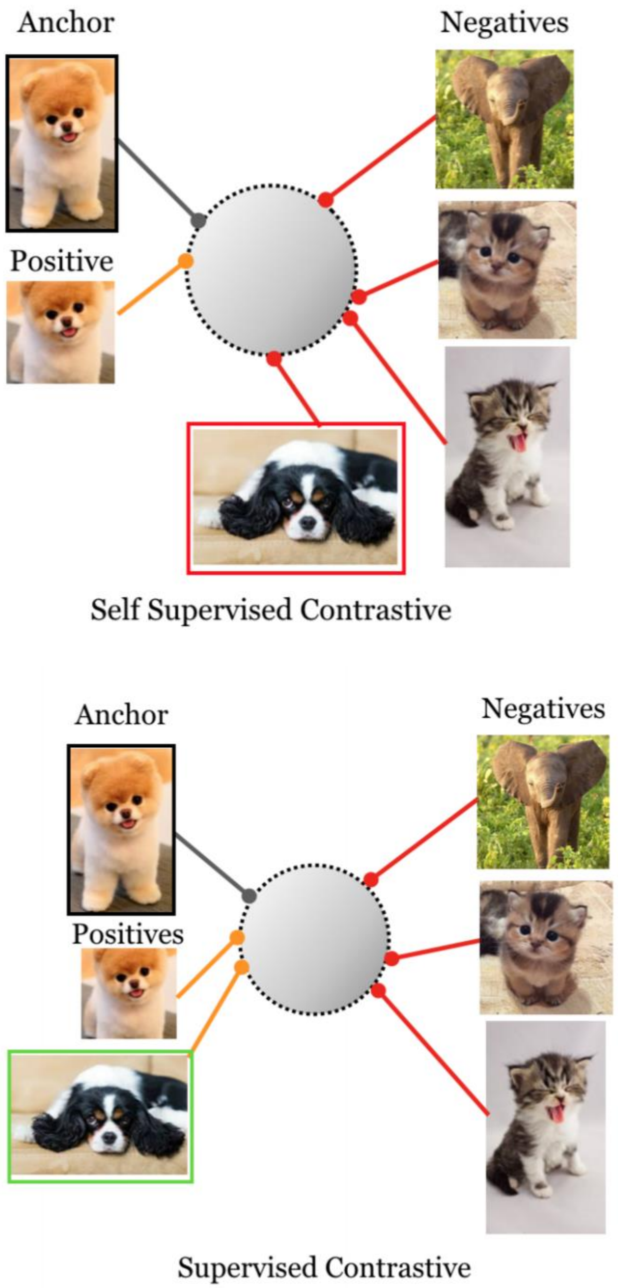


Supervised contrastive learning

- Addresses limitation of self-supervised learning where negatives might belong to the same class as the anchor and positive instances.
- Introduced by Google Research and MIT in 2021 as an extension to self-supervised contrastive learning.

Key Differences

- Uses class labels to create embeddings where objects from the same class are closely aligned.
- Proposes using multiple positive and negative instances per anchor image for more effective representations.



[3]

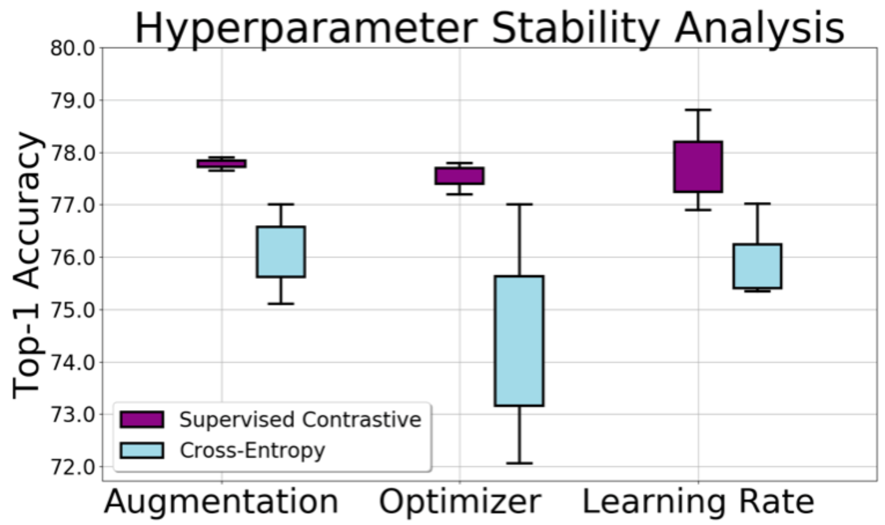
Supervised contrastive learning

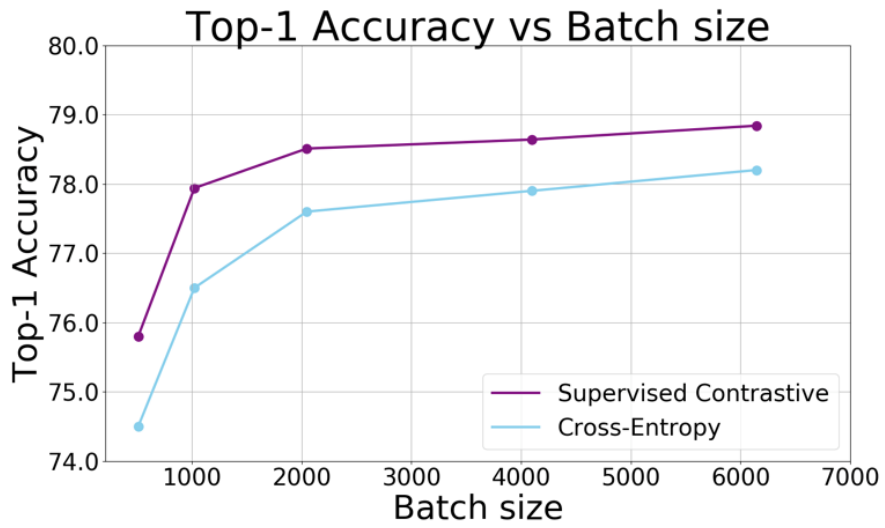
Benefits

- Achieves superior classification accuracy compared to models trained with cross-entropy loss.
- Less sensitive to hyperparameters compared to models trained with cross-entropy loss.

Drawbacks

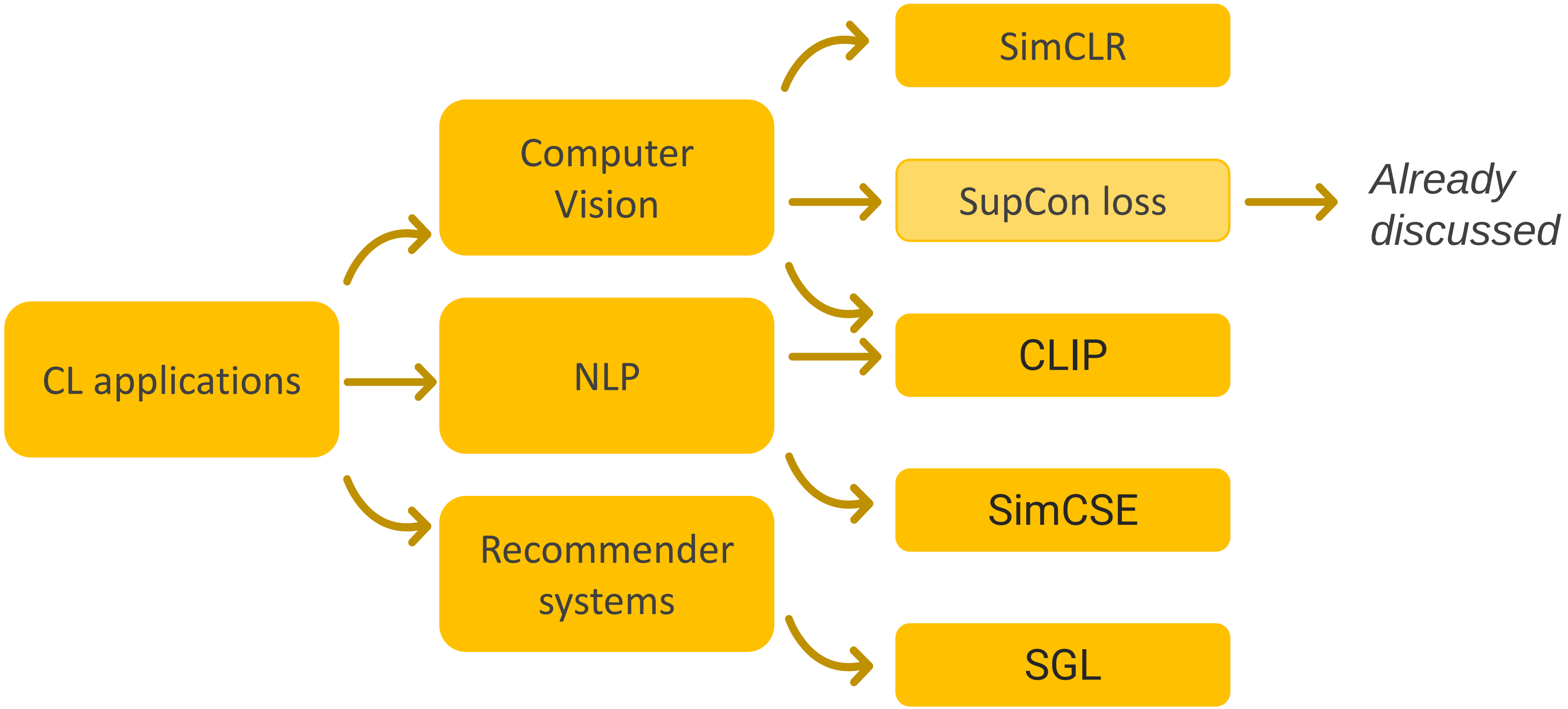
- Large batch sizes are required





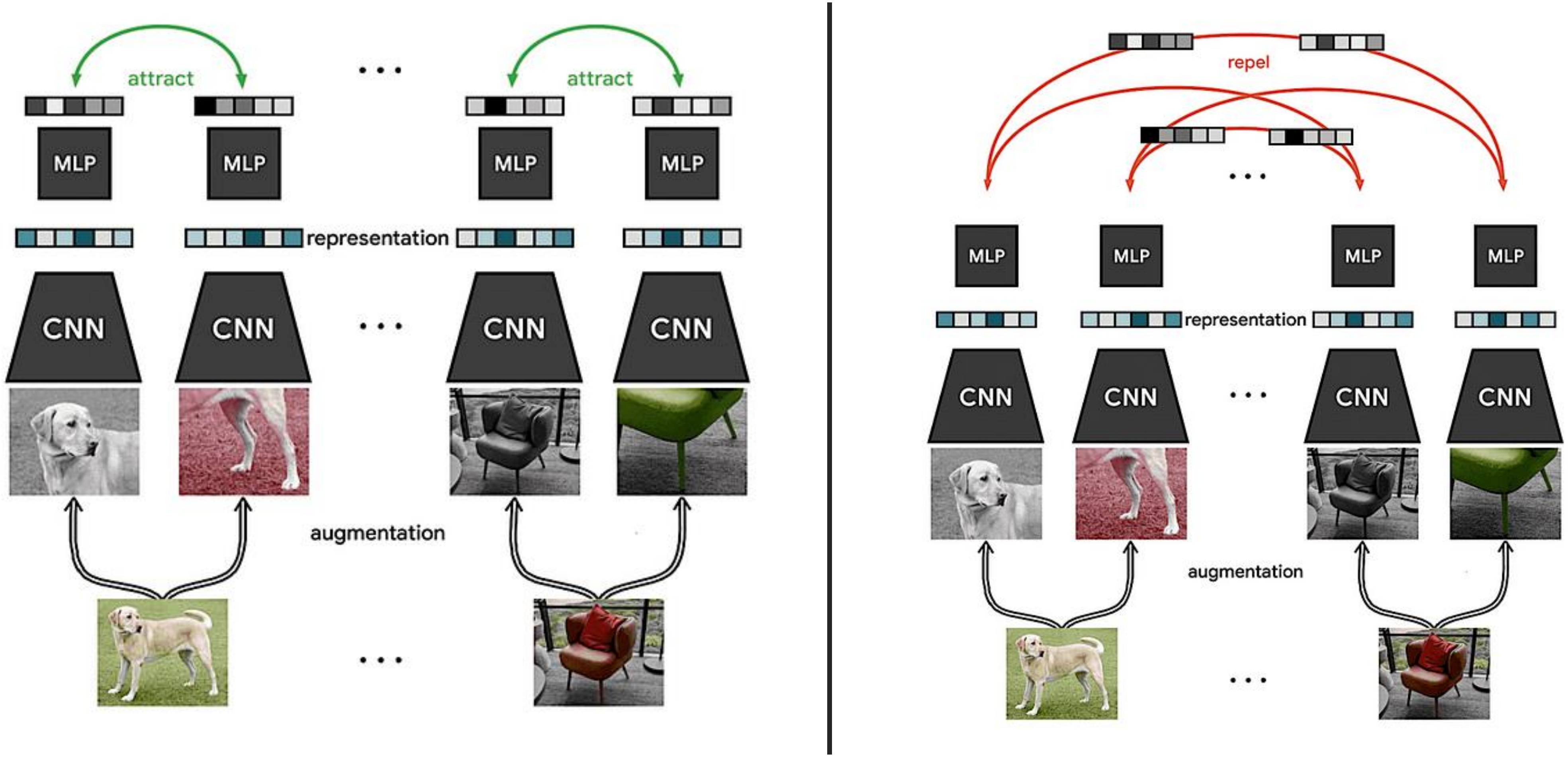
[3]

Practical applications of contrastive learning



Applications > Computer Vision > SimCLR

Illustration of SimCLR methodology



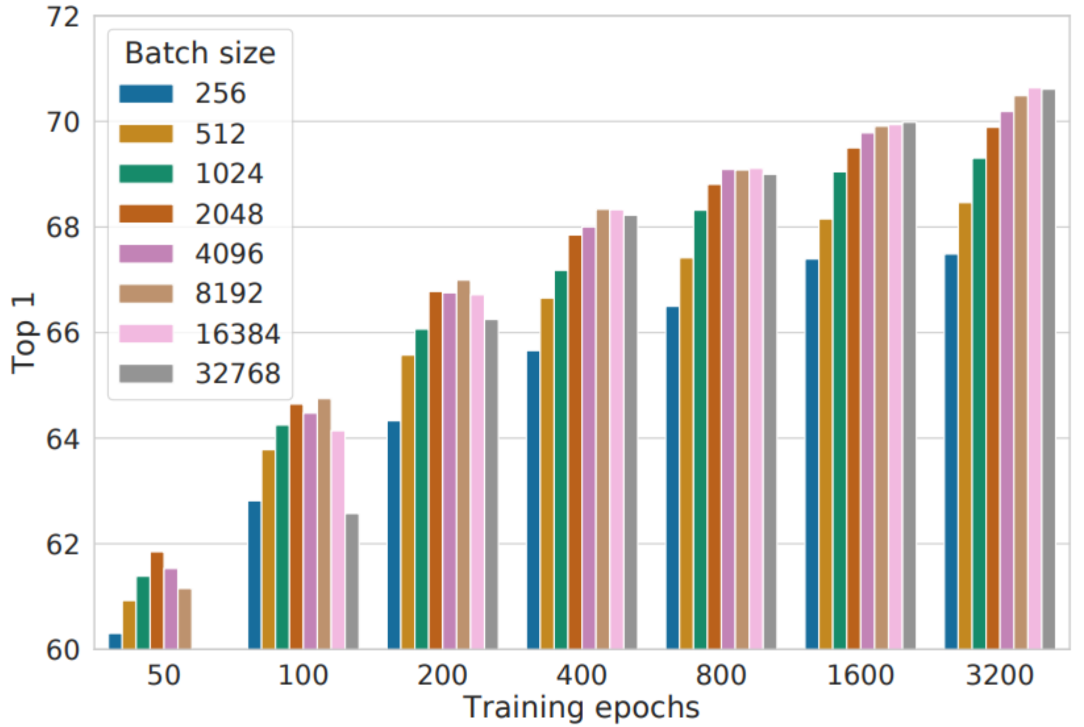
[4]

Applications > Computer Vision > SimCLR

Key Consideration

- SimCLR and Supervised Contrastive Learning require large batch sizes for more discriminative embeddings, leading to high computational resource demand.

Impact of batch size and training epochs on the performance of SimCLR



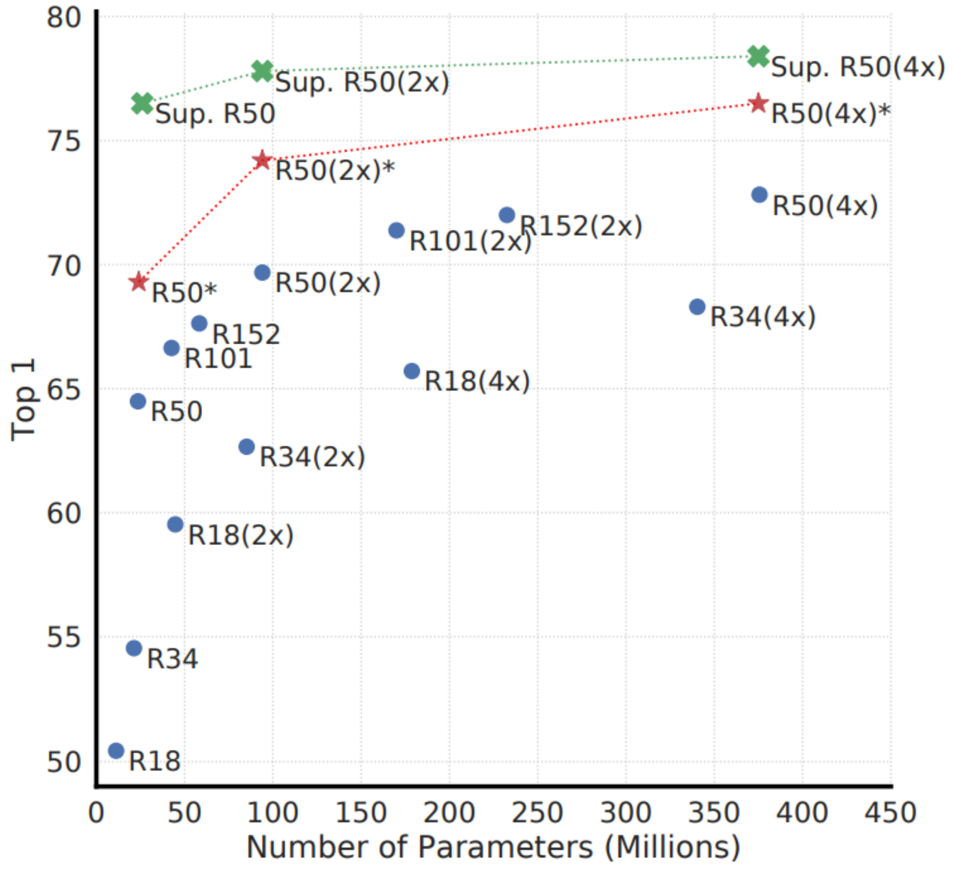
[2]

Applications > Computer Vision > SimCLR

Supervised vs self-supervised learning

- Self-supervised learning requires larger models and longer training periods but reduces the need for labeled data.
- Achieves comparable performance to supervised learning with enough computational resources.
- Decision between methods should consider available data and computational resources.

SimCLR: Accuracy comparison of supervised (90 epochs) and self-supervised CL (1000 epochs) models as the number of model parameters increases.



[2]

Applications > NLP & CV > CLIP

Contrastive Language-Image Pretraining

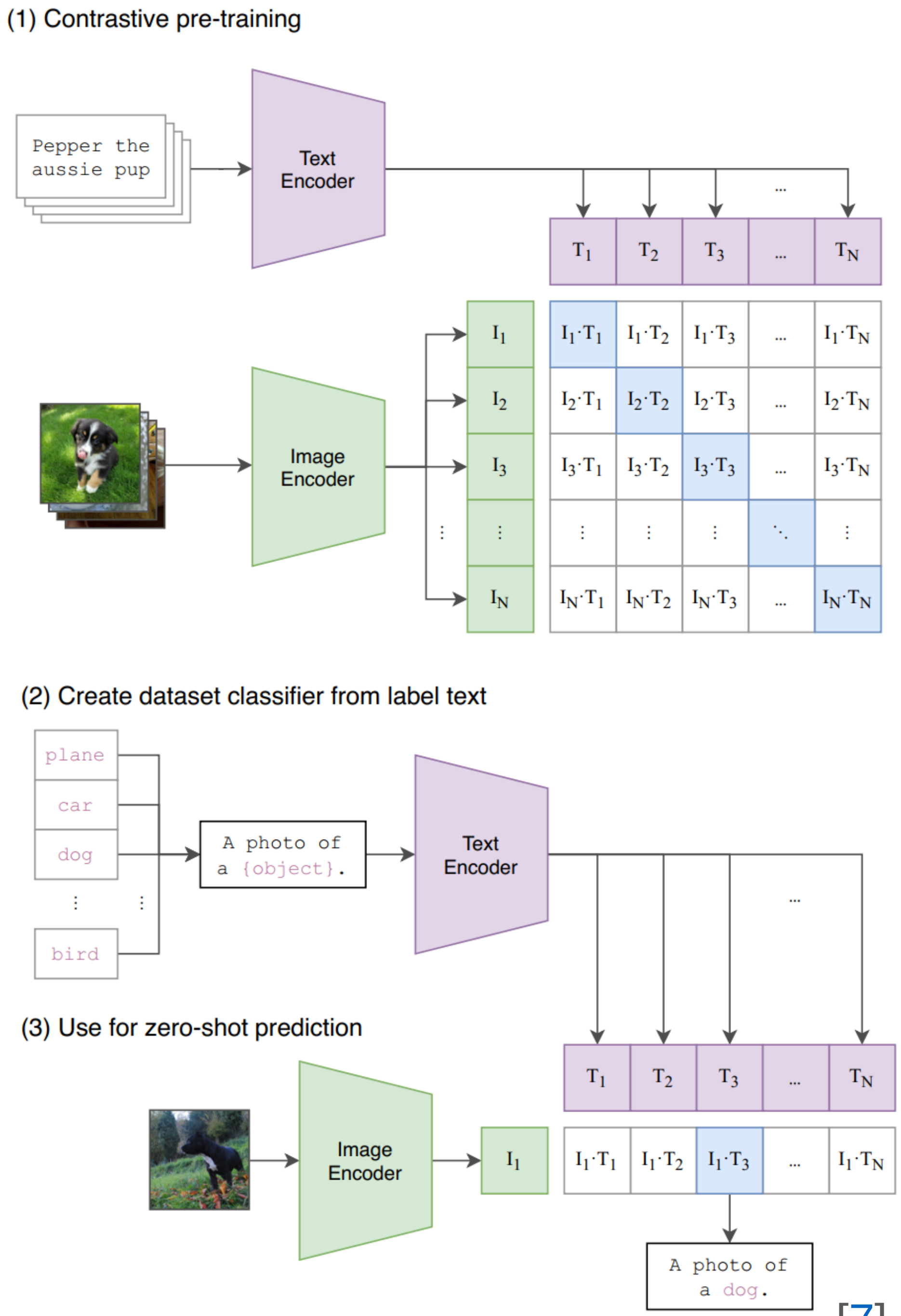
CLIP is a multi-modal model developed by OpenAI for understanding the relationship between images and text.

Pre-training

- It uses contrastive learning as a pretraining objective and enables zero-shot image classification.
- The model consists of text and image encoders trained on a large dataset of paired images and text captions.
- Contrastive learning pulls together the encodings of corresponding image-text pairs and pushes apart encodings from different pairs.

Zero-shot classification

- Given an image, its encoding is obtained using the trained image encoder, and text embeddings representing different classes are generated.
- Prompt engineering is performed to modify the input format for the text encoder.
- The final classification result is determined by calculating the similarity scores between the image encoding and text encodings.



[7]

Applications > NLP > SimCSE

SimCSE uses contrastive learning to produce high-quality sentence embeddings in both self-supervised and supervised contexts.

Self-Supervised Approach

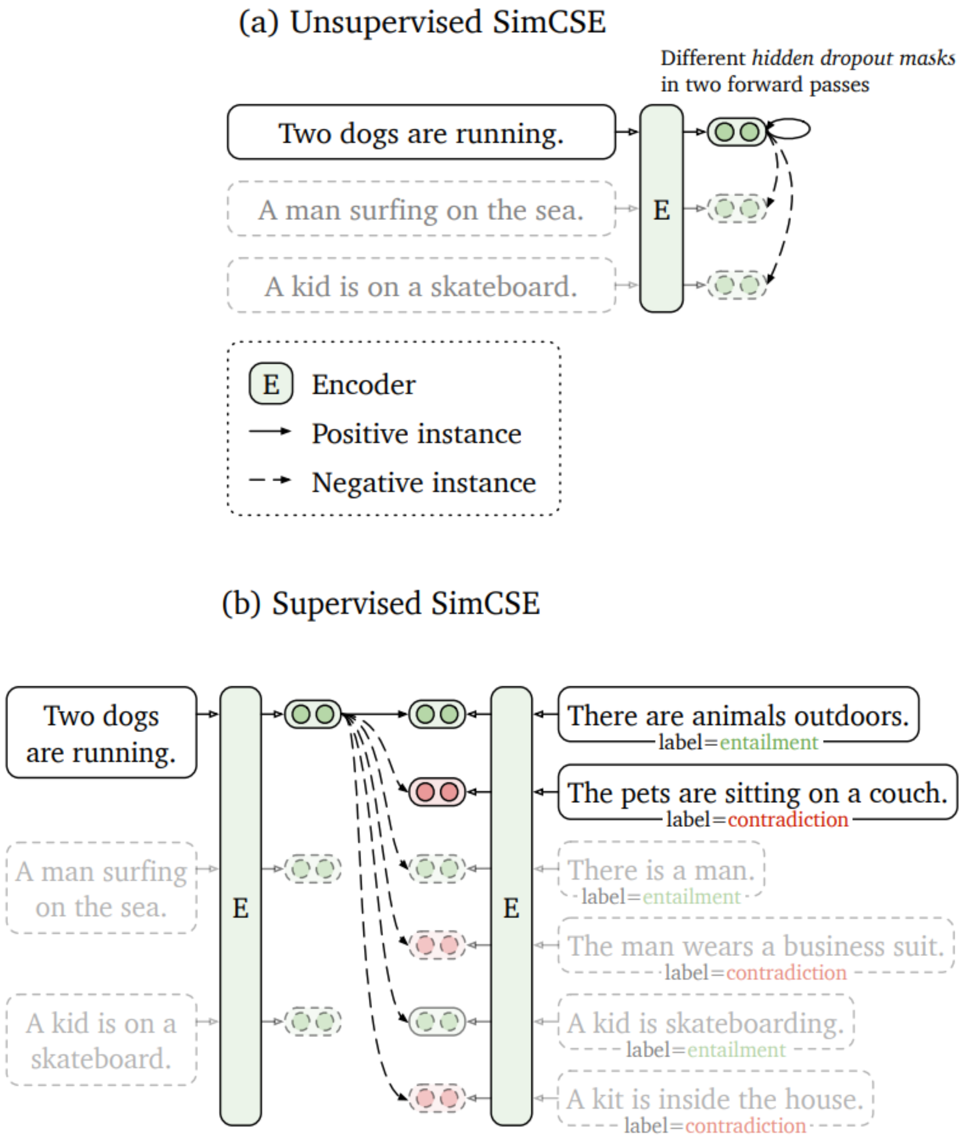
- Uses dropout as data augmentation: an anchor sentence is modified twice with different dropout masks.
- The modifications are pulled together and are pushed apart from the remaining in-batch sentences.

Supervised Approach

- Trains using a Natural Language Inference (NLI) dataset: each entry contains an anchor sentence, an entailment sentence, a contradiction sentence, and a neutral sentence.
- Entailment sentences are labeled as 'positive instances', contradictions as 'negative instances', and neutral sentences are discarded.
- Including 'hard negatives', i.e., specific contradictions of the anchor, improves the model performance.

Evaluation

- Uses Semantic Text Similarity tasks: computes cosine similarity between sentence embeddings and compares to human-annotated similarity.

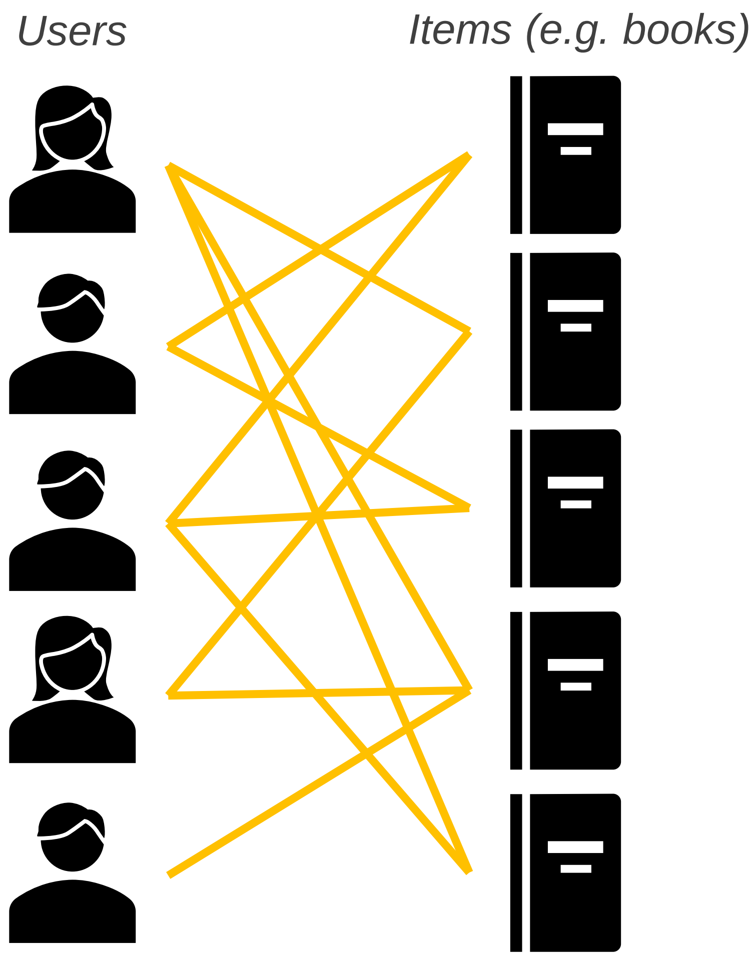


[6]

Applications > Recommender Systems

Recommender systems

- Model that seeks to predict the 'rating' or 'preference' that a user would give to an item.
- The model attempts to learn embeddings for each user and each item.
- Contrastive learning can be used to enhance the quality of the learned embeddings.

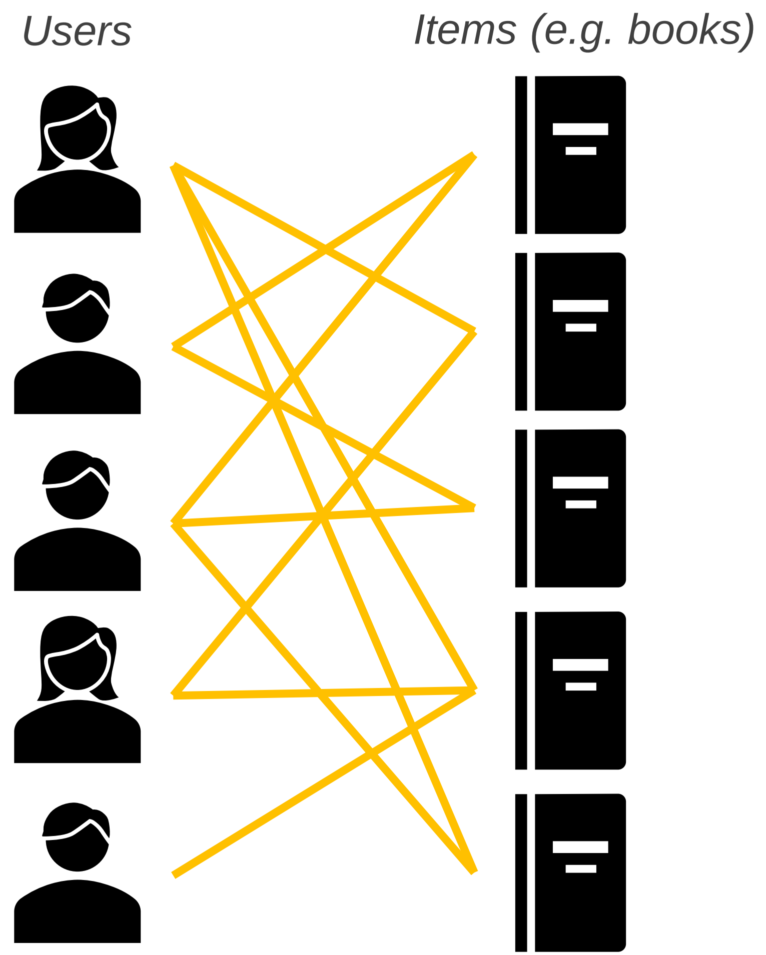


User-item interaction graph (bipartite graph)

Applications > Recommender Systems > SGL

Self-supervised Graph Learning (SGL)

- A novel learning paradigm known for improving the *accuracy* and *robustness* of Graph Convolutional Networks (GCNs) for recommendation systems.



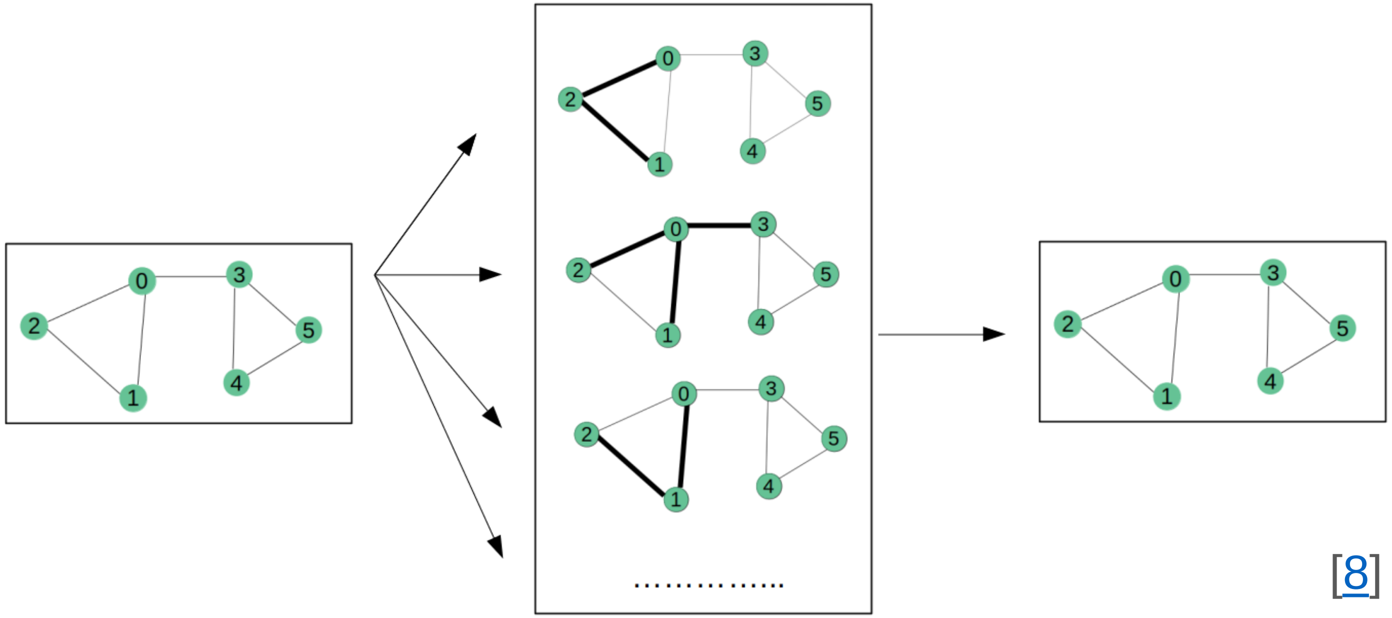
User-item interaction graph (bipartite graph)

Applications > Recommender Systems > SGL

Graph Convolutional Networks (GCNs)

- Type of neural network used to process data that is represented as graphs
- Key idea = aggregate information from a node's neighboring nodes and then use that aggregated information to update the node's own representation.

GCNs can be applied to user-item interaction graphs to learn user and item representations, based on edge connections.



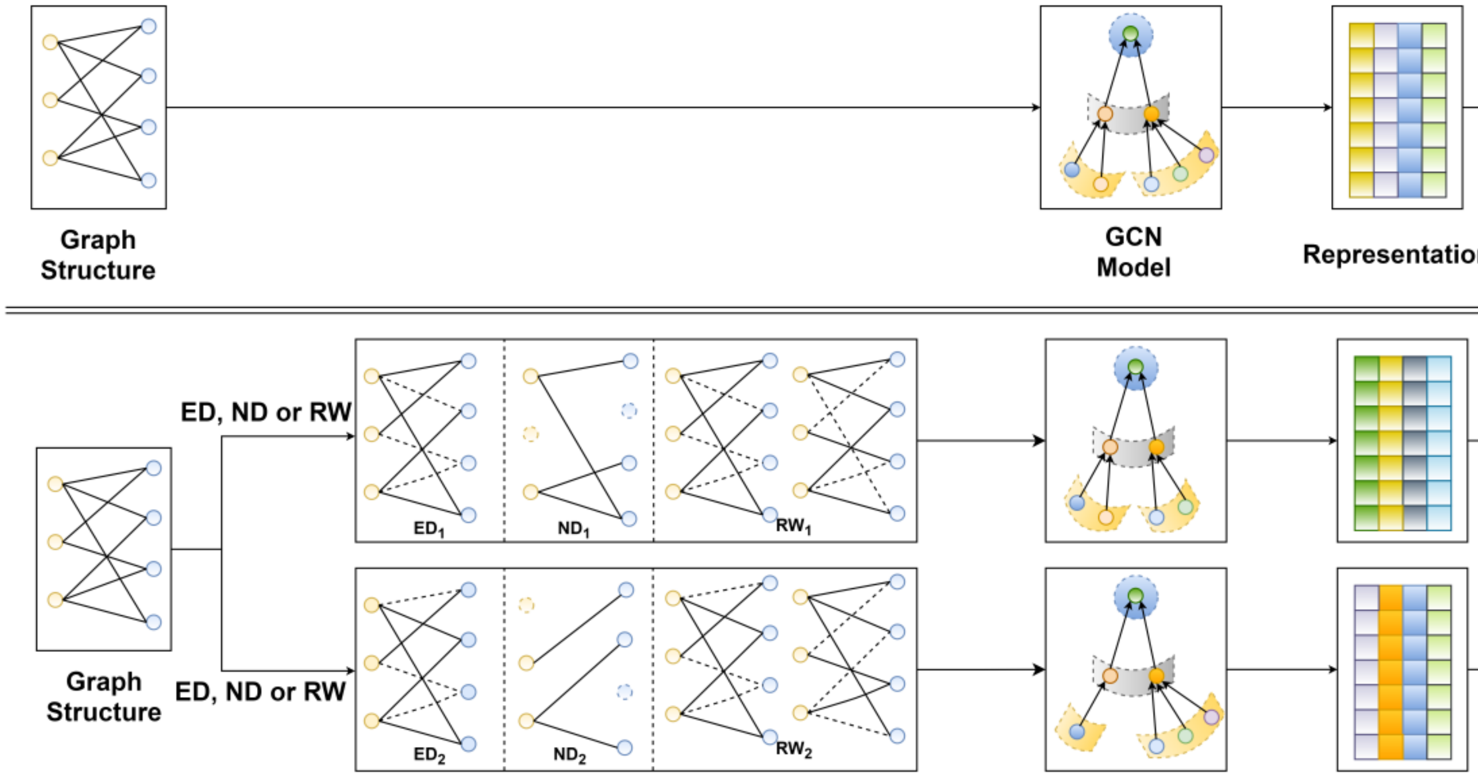
[8]

Applications > Recommender Systems > SGL

Classic recommendation task using the original graph

Multi-task learning

Self-supervised contrastive learning task using augmented graphs



Bayesian Personalized Ranking (BPR) loss

$$\mathcal{L}_{main} = \sum_{(u,i,j)\in O} -\log \sigma(\hat{y}_{ui} - \hat{y}_{uj})$$

$$\mathcal{L}_{ssl} = \mathcal{L}_{ssl}^{user} + \mathcal{L}_{ssl}^{item}$$

InfoNCE loss (a contrastive loss)

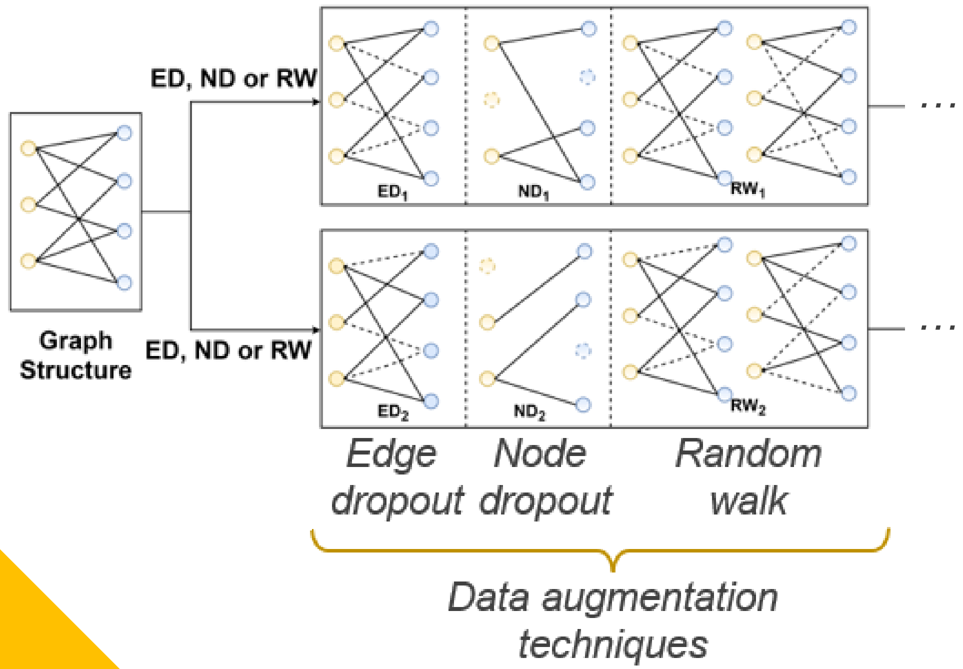
$$\mathcal{L}_{ssl}^{user} = \sum_{u\in\mathcal{U}} -\log \frac{\exp(s(\mathbf{z}'_u, \mathbf{z}''_u)/\tau)}{\sum_{v\in\mathcal{U}} \exp(s(\mathbf{z}'_u, \mathbf{z}''_v)/\tau)}$$

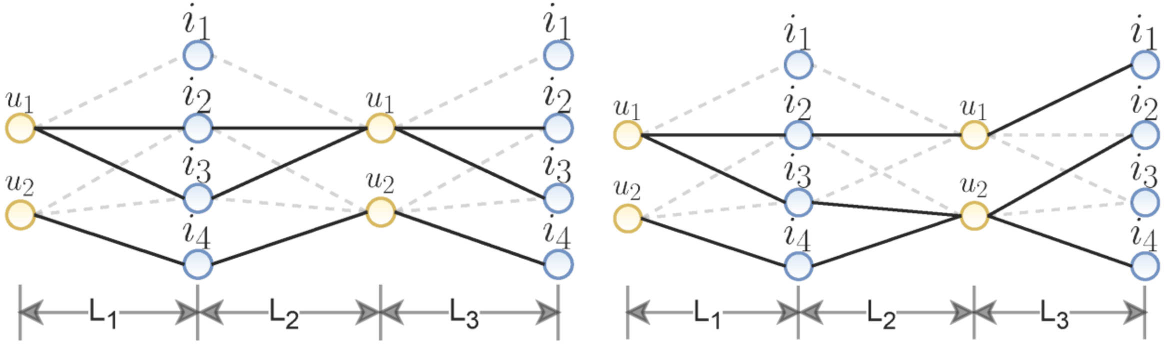
s = cosine similarity

[9]

Applications > Recommender Systems > SGL

Data augmentation process





(a) Edge Dropout **(b) Random Walk**

[9]

- ED: Randomly drop edges.
- ND: Randomly drop nodes.
- RW: The process constructs a new subgraph by performing a random walk starting from a given node.

Conclusion

- We’ve only scratched the tip of the iceberg – there are many other fields and examples to consider.
- Contrastive learning is a *representation learning* tool that aims to discover meaning representations by contrasting encodings from the same class, and from different classes.
- Contrastive learning can be applied in a *self-supervised* or a *supervised* context and there are pros and cons to each approach.
- *Data augmentation* is an important aspect of self-supervised contrastive learning.
- Contrastive learning can be applied across several domains.

References

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

[2] Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., 2020, November. A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PMLR.

[3] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C. and Krishnan, D., 2020. Supervised contrastive learning. *Advances in neural information processing systems*, *33*, pp.18661-18673.

[4] https://towardsdatascience.com/paper-explained-a-simple-framework-for-contrastive-learning-of-visual-representations-6a2a63bfa703

[5] He, K., Fan, H., Wu, Y., Xie, S. and Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729-9738).

[6] Gao, T., Yao, X. and Chen, D., 2021. SimCSE: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

[7] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. and Krueger, G., 2021, July. Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PMLR.

[8] https://towardsdatascience.com/understanding-graph-convolutional-networks-for-node-classification-a2bfdb7aba7b

[9] Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J. and Xie, X., 2021, July. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval* (pp. 726-735).


Questions?