SAFEAI- 2022

IS IT ALL A CLUSTER GAME? - EXPLORING OUT-OF-DISTRIBUTION DETECTION BASED ON CLUSTERING IN THE EMBEDDING SPACE

POULAMI SINHAMAHAPATRA, RAJAT KONER, KARSTEN ROSCHER, STEPHAN GÜNNEMANN

MOTIVATION

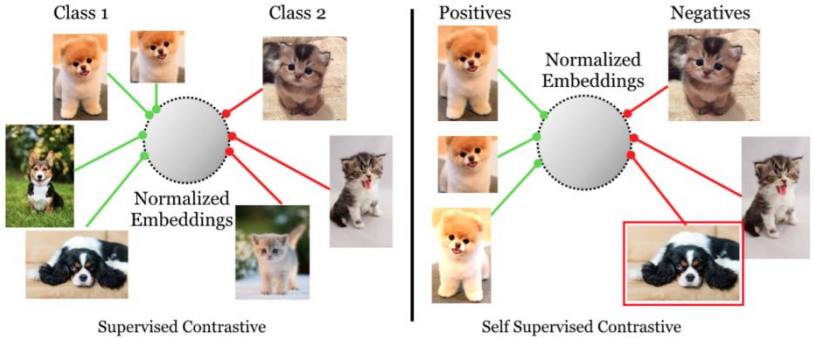
- Exploring Out-of-Distribution (OOD) detection problem
  - Embeddings of similar semantic data should cluster together in embedding space
  - Can we learn to distinguish In-Distribution (ID) from OOD samples based on clusters ?

KEY IDEAS

- Using Contrastive Learning (CL) methods
  - cluster similar instances/ classes together
  - push apart dissimilar instances/ classes

KEY IDEAS

- Using Contrastive Learning (CL) methods
  - cluster similar instances/ classes together
  - push apart dissimilar instances/ classes



Source: Khosla et. Al:"Supervised Contrastive Learning"

- CL methods used in this work:
  - Supervised Contrastive Learning (**SupCon**)
  - Self-supervised Contrastive Learning (**SimCLR**)

- Baseline:
  - Cross-Entropy (**CE**)

KEY IDEAS/ CONTRIBUTIONS

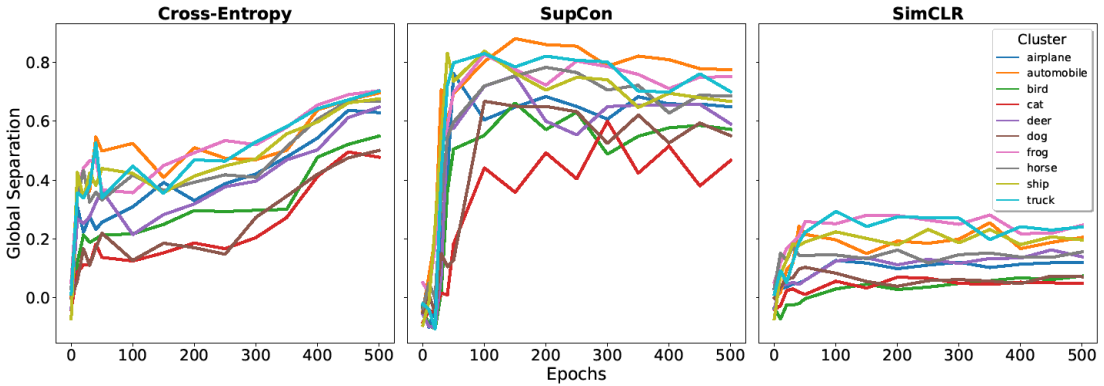
- Using Contrastive Learning (CL)
  - cluster similar instances/ classes
  - push apart dissimilar instances/ classes

- Analysing cluster formation in embedding space
  - Global Separation
  - Cluster Purity

KEY IDEAS/ CONTRIBUTIONS

- Using Contrastive Learning (CL)
  - cluster similar instances/ classes
  - push apart dissimilar instances/ classes

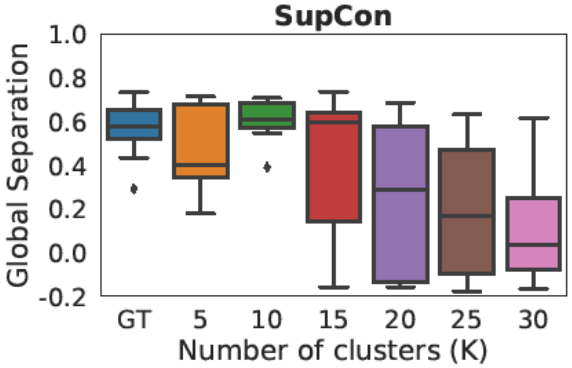
- Analysing cluster formation in embedding space
  - Global Separation
  - Cluster Purity

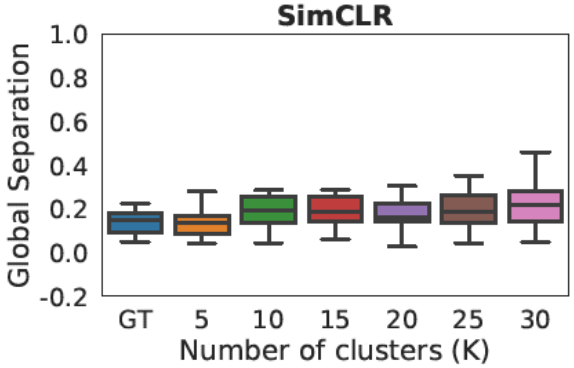


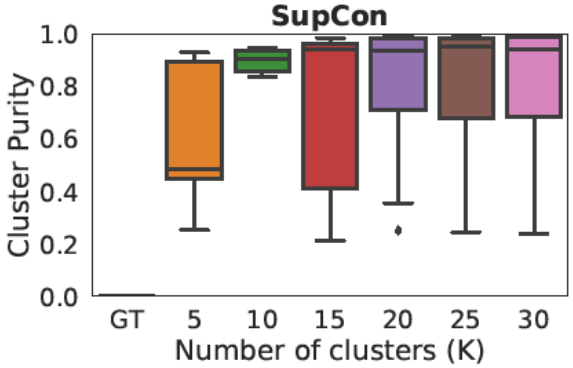
Evolution of cluster separation over time for supervised and self-supervised methods

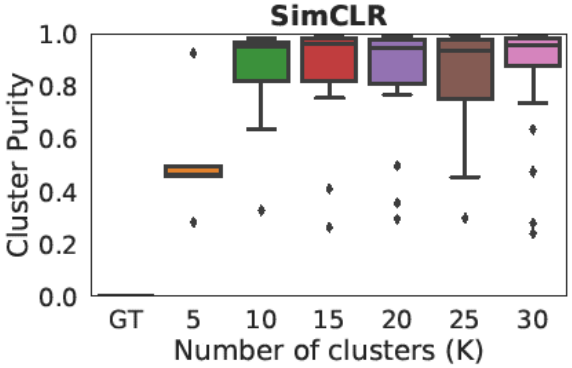
RESULTS: DETERMINING CLUSTER QUALITY

— Comparison of cluster quality for SupCon and SimCLR based on Global Separation and Cluster Purity for GT classes as well as K-means based cluster









KEY IDEAS/ CONTRIBUTIONS

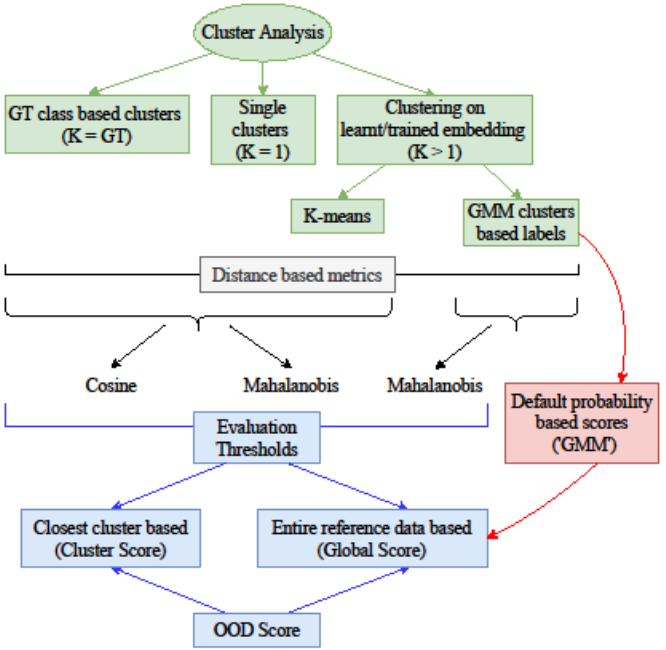
- Using Contrastive Learning (CL)
  - cluster similar instances/ classes
  - push apart dissimilar instances/ classes

- Analysing cluster formation in embedding space
  - Global Separation
  - Cluster Purity

- Using distance-based metrics to perform OOD detection
  - Metrics: Mahalanobis and Cosine Similarity
  - AUROC as OOD evaluation scores
  - Cluster-based (K-means/GMM) and Global scores

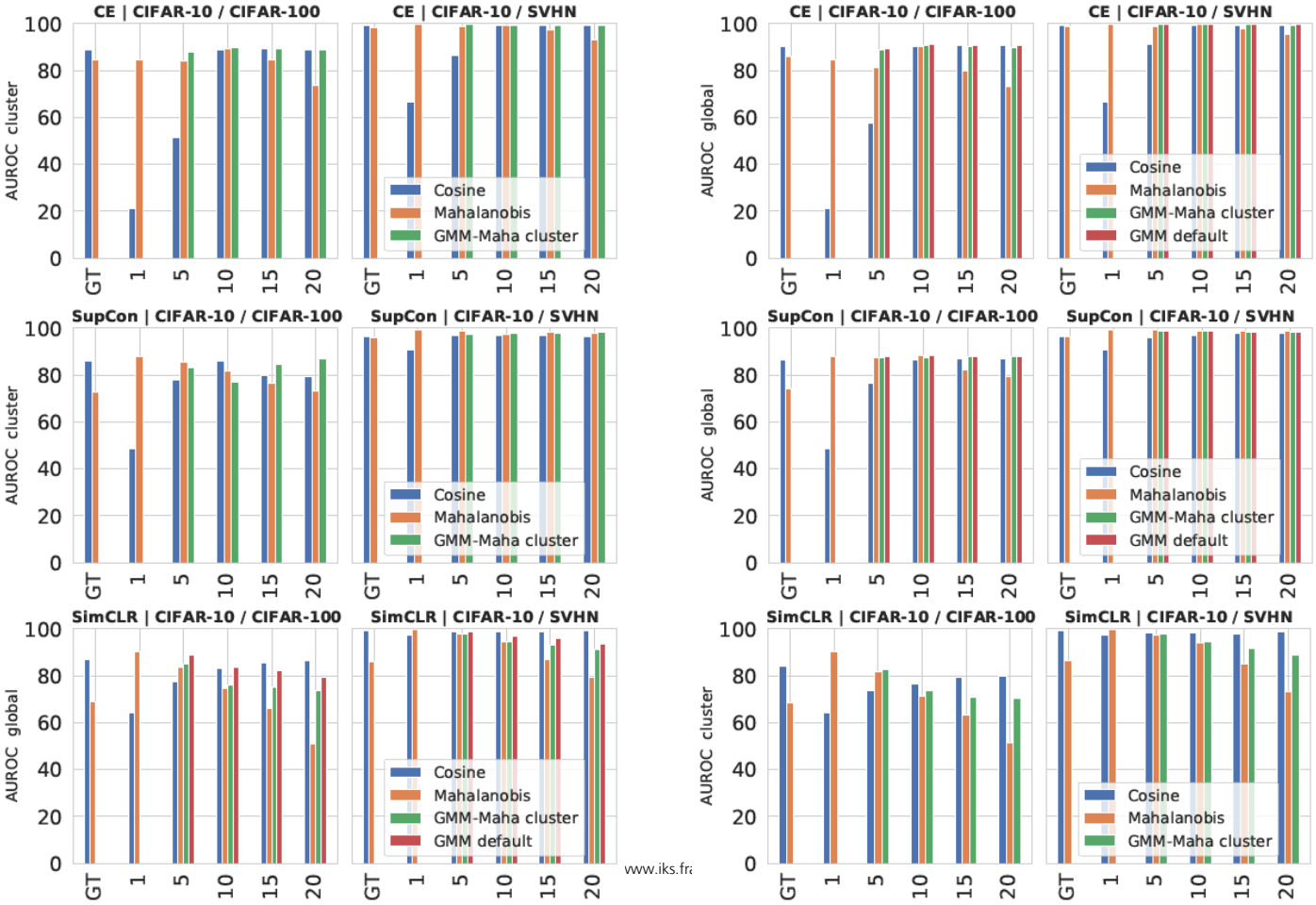
KEY IDEAS/ CONTRIBUTIONS

- Using Contrastive Learning (CL)
  - Cluster similar instances/ classes
  - Push apart dissimilar instances/ classes
- Analysing cluster formation in embedding space
  - Global Separation
  - Cluster Purity
- Using distance-based metrics to perform OOD detection
  - Metrics: Mahalanobis and Cosine Similarity
  - AUROC as OOD evaluation scores
  - Cluster-based (K-means/GMM) and Global scores



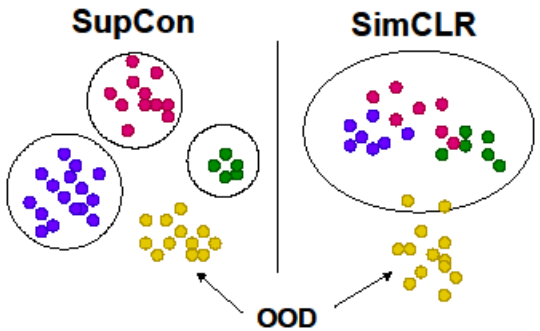
Schematic of our OOD evaluation pipeline

RESULTS: OOD DETECTION



CONCLUSION AND FUTURE WORK

— Depending on supervised or self-supervised training distinct or overlapping clusters of embeddings exist

— Simple less expensive methods like CE provide competitive results

— Across all the different trends investigated, there's no clear winner yet and needs to be further explored