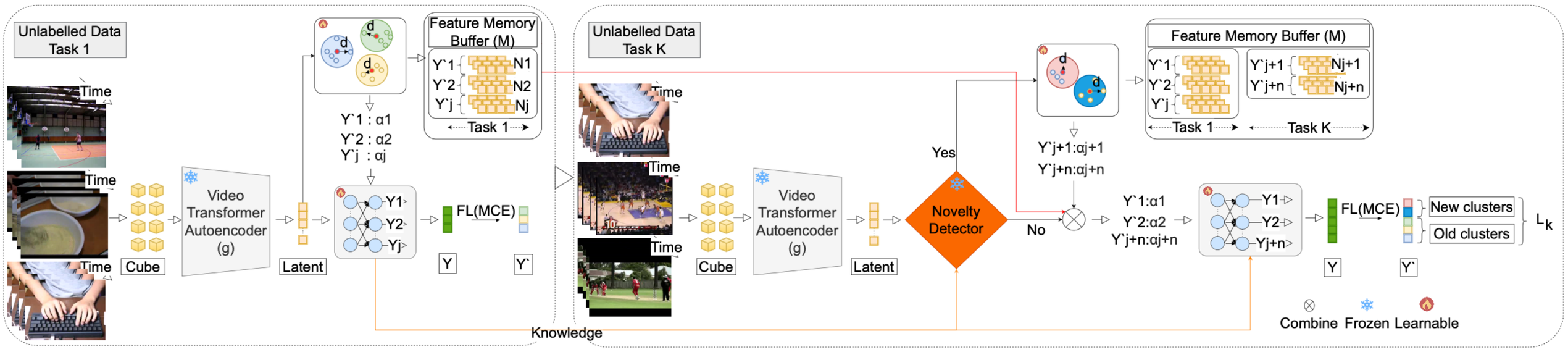
Unsupervised Video Continual Learning via Non-Parametric Deep Embedded Clustering

Nattapong Kurpukdee and Adrian G. Bors,
Department of Computer Science, University of York, York YO10 5GH, UK
Email : nattapong.kurpukdee@york.ac.uk, adrian.bors@york.ac.uk

BMVC 2025



Overview of the proposed unsupervised video continual learning based on the Kernel Density Estimation (uVCL-KDE).

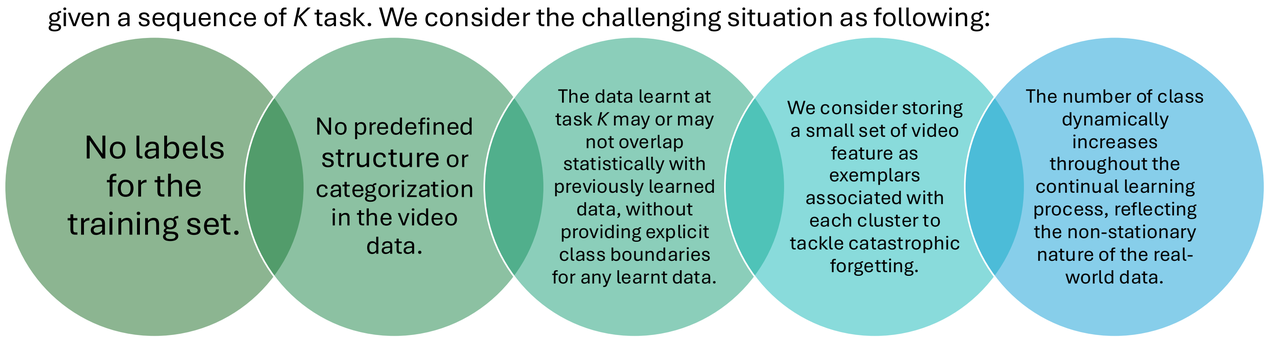
Abstract

- ✓ Prior studies have focused on supervised continual learning, relying on the knowledge of labels and task boundaries [1, 2], while having labelled data is costly and not practical.
- ✓ To address this gap, we propose a realistic scenario for unsupervised video continual learning (uVCL), where neither task boundaries nor labels are provided when learning a succession of tasks.
- ✓ We proposed an approach to use Kernel Density Estimation (KDE) of deep embedded video features extract by unsupervised video transformer networks as a non-parametric probabilistic representation of the data.
- ✓ We introduce a novelty detection criterion for the incoming new task data, dynamically enabling the expansion of memory clusters, aiming to capture new knowledge when learning a succession of tasks.
- ✓ To prevent catastrophic forgetting problems, we leverage the use of transfer learning from the previous task as initial state for the knowledge transfer to the current learning task.
- ✓ Our proposed model is evaluated on three standard video action recognition datasets including UCF101, HMDB51, and Something-to-Something V2 without using any labels or class boundaries.

Methodology

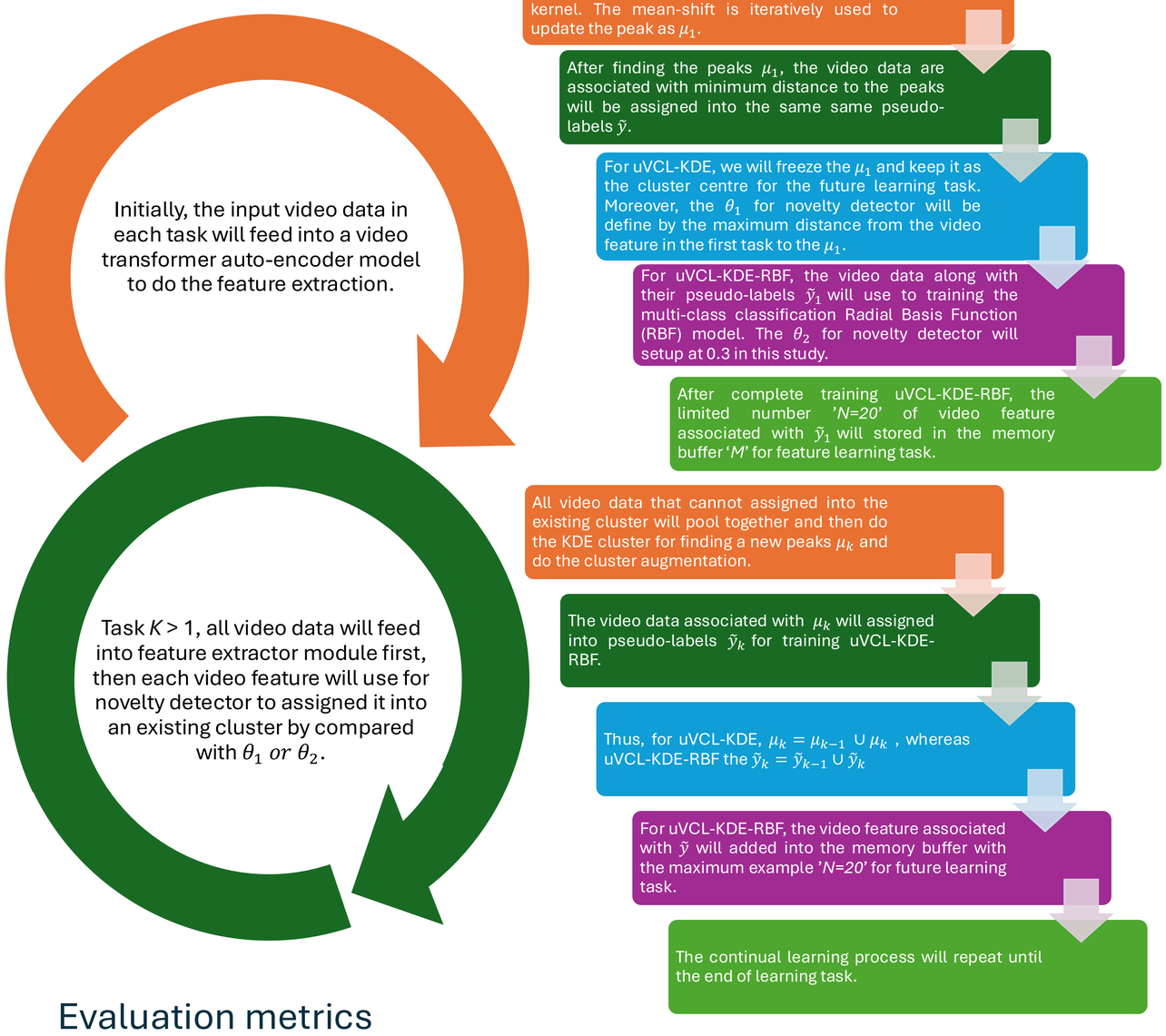
Problem setup

We study unsupervised video continual learning, aiming to learn and structure a data space *H*, given a sequence of *K* task. We consider the challenging situation as following:



Architecture

The overview of our proposed pipeline is illustrated in the figure above.



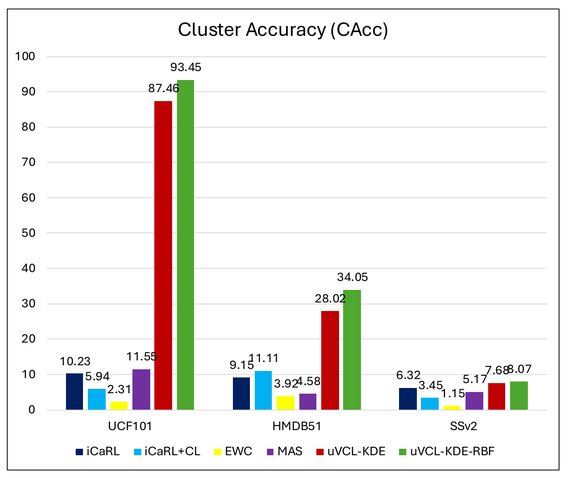
Evaluation metrics

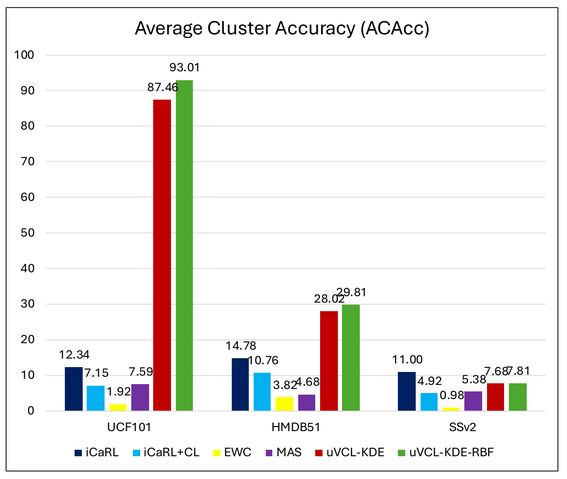
We evaluate the cluster accuracy (CAcc), first we employ the Hungarian matching algorithm [3] to associate each pseudo-label with the ground truth label, where the label use for testing only.

- ✓ Then we evaluate the average unsupervised continual learning accuracy (ACAcc) [1, 4].
- ✓ After that, we evaluate the Forward forgetting (FWF) [1, 4]. A large positive Forward Forgetting is also known as catastrophic forgetting.
- ✓ Lastly, we evaluate the Backward Forgetting (BWF) [1, 4]. A large negative backward forgetting is also known as catastrophic forgetting.

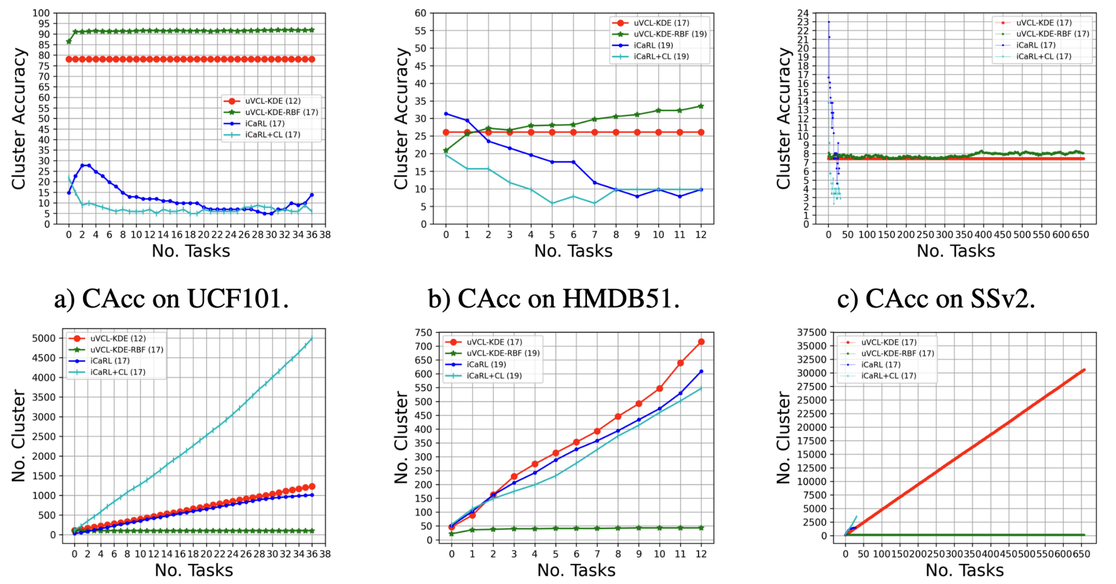
Results

The performance on UCF101, HMDB51, and SSv2.





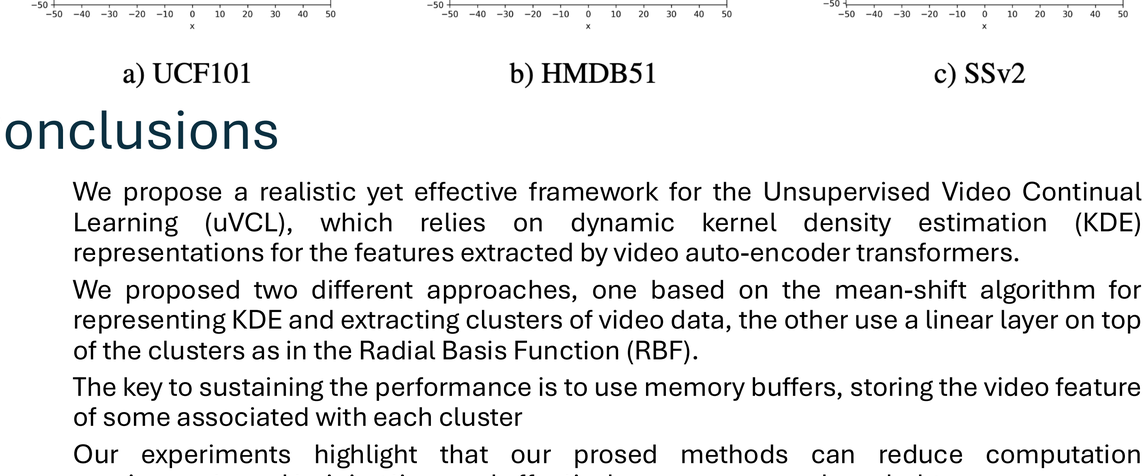
The progressive performance UCF101, HMDB51, and SSv2 considering the first fold data. Inside the brackets for each method is the bandwidth *h* for the mean-shift clustering.



a) CAcc on UCF101. b) CAcc on HMDB51. c) CAcc on SSv2.

d) No. of clusters on UCF101. e) No. of clusters on HMDB51. f) No. of clusters on SSv2.

We visualize the laten space in the memory for each cluster after learning all tasks by using t-SNE for feature reduction to 2-Dimensions with perplexity of 40. + represents the cluster centre, and a number represents the cluster ID.



a) UCF101 b) HMDB51 c) SSv2

Conclusions

- ✓ We propose a realistic yet effective framework for the Unsupervised Video Continual Learning (uVCL), which relies on dynamic kernel density estimation (KDE) representations for the features extracted by video auto-encoder transformers.
- ✓ We proposed two different approaches, one based on the mean-shift algorithm for representing KDE and extracting clusters of video data, the other use a linear layer on top of the clusters as in the Radial Basis Function (RBF).
- ✓ The key to sustaining the performance is to use memory buffers, storing the video feature of some associated with each cluster
- ✓ Our experiments highlight that our prosed methods can reduce computation requirements and training time and effectively preserves pass knowledge.
- ✓ Future work, we will employ a dynamic novelty detector criterion for deciding when learn new information and define new clusters.

References

[1] Andrés Villa, Kumail Alhamoud, Victor Escorcia, Fabian Caba Heilbron, Juan León Alcázar, and Bernard Ghanem. vCLIMB: A novel video class incremental learning benchmark. In Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recog. (CVPR), pages 19013–19022, 2022.
[2] Nattapong Kurpukdee and Adrian G. Bors. Temporal transformer encoder for video class incremental learning. In Proc. of the IEEE International Conference on Image Processing (ICIP), pages 1295–1301, 2024.
[3] Harold W Kuhn. The hungarian method for the assignment problem. Naval research logistics quarterly, 2(1-2):83–97, 1955.
[4] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. Advances in neural information processing systems, 30, 2017.