Else-Net: Elastic Semantic Network for Continual Action Recognition from Skeleton Data (Supplementary Material)

Tianjiao Li [1] Qiuhong Ke [2] Hossein Rahmani [3] Rui En Ho [1] Henghui Ding [4] Jun Liu [1] *

[1] Singapore University of Technology and Design [2] University of Melbourne

[3] Lancaster University [4] Nanyang Technological University

{tianjiao_li, ruien_ho}@mymail.sutd.edu.sg, jun_liu@sutd.edu.sg

h.rahmani@lancaster.ac.uk, qiuhong.ke@unimelb.edu.au, ding0093@e.ntu.edu.sg

1. Ablation Study on Exploration of New Memory Blocks

We evaluate the impacts on exploration of brand new memory blocks and conduct experiments on the NTU RGB+D dataset (cross subject protocol). The results are shown in Tab. 1. The "Else-Net" is able to search for the most relevant blocks among the existing memory blocks, as well as explore candidate new memory blocks. However, the "Else-Net w/o Exploration" does not explore new memory blocks during continual learning process, i.e., it can only select relevant blocks among exiting memory blocks. As we can see in Tab. 1, the performance drops when the model does not explore new memory blocks. This shows that our Else-Net achieves better continual learning performance.

Table 1. Performance comparison (%) on exploration of new memory blocks. Exploration of brand new memory blocks can grand the model with abilities in effectively learning new human actions and preserving the past knowledge to mitigate catastrophic forgetting problems.

Methods	ACC	FM	LA
Else-Net w/o Exploration	80.2	8.7	85.5
Else-Net	84.4	5.1	87.6

2. Visualization of Block Searching

To demonstrate how the proposed Else-Net searches for the most relevant memory blocks w.r.t. the current input human actions, we visualize the process of Block Searching in Fig. 1.

As we can see in Fig. 1, our Else-Net first learns to recognize "wipe face". And then, when learning the action "cross hands in front", the model tends to share some memory blocks in different body part branches to take better use of past knowledge, e.g., the Else-Net selects the same Memory Blocks in the torso branch. However, for the right arm parts between these two actions, large discrepancies can be observed. Therefore, the Else-Net explores brand new Memory Blocks to capture the new knowledge, to achieve effective learning. It is also worth noting that although no new Memory Blocks are explored for left leg branch, the discrepancies are also captured by our Else-Net via employment of different existing blocks.

Therefore, this shows that the proposed Else-Net has the abilities in capturing the homogeneous features between the newly-incoming human actions and past knowledge from previous actions via sharing common Memory Blocks, and thus the new actions can be effectively learned. In the meantime, the Else-Net only updates the selected relevant memory blocks and freezes the non-selected irrelevant ones. In such a case, our model is also capable of consolidating the past memories to mitigate the catastrophic forgetting problems.

3. Summary of Techniques

Continual Learning aims to learn new tasks sequentially without forgetting previously learned knowledge. Hayes *et al.* [4] proposed to mimic the consolidation processing of human brains by encoding the raw images into latent embedding features and storing the features into the episodic memory, using product quantization technique. Also, the authors randomly selected some classes to train the model offline with the purpose of giving prior knowledge to the recognition model, while the other classes are continuously learned. Lopez-Paz *et al.* [20] proposed to use the task descriptors to provide additional contextual information when learning continuously-input tasks. By constraining the angles between the loss gradient vectors of cur-

*Corresponding Author.

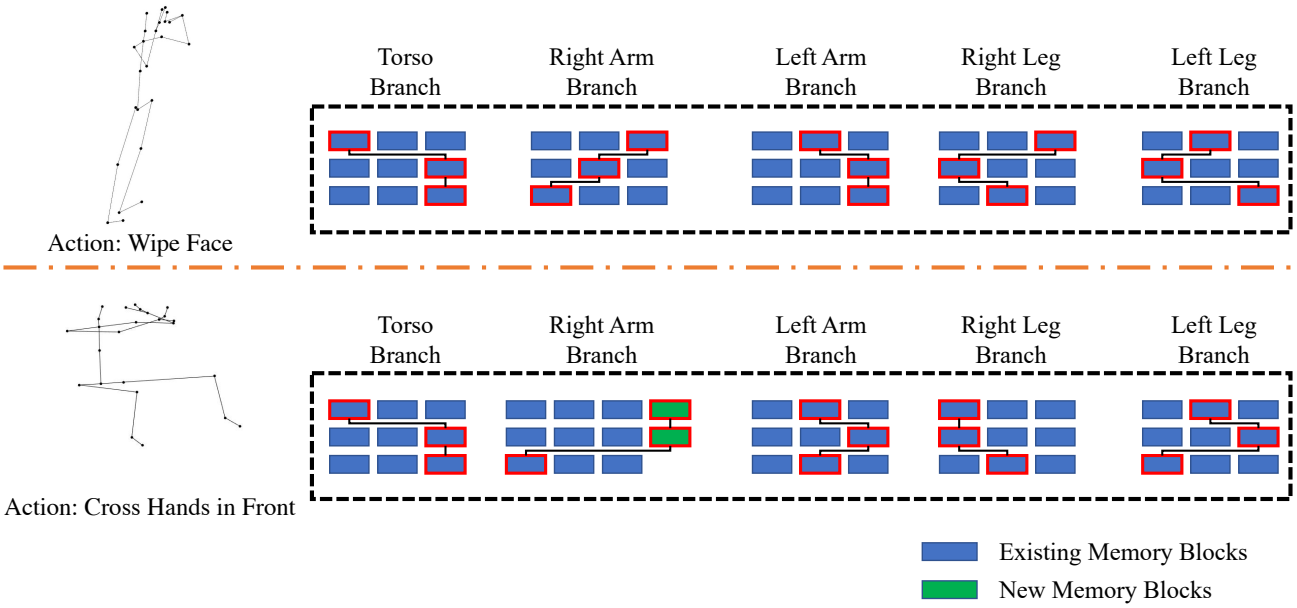

Figure 1. Visualization of the proposed block searching scheme. The Memory Blocks in red boundaries are the selected most relevant blocks. The proposed Else-Net learns to recognize "wipe face" and "cross hand in front" sequentially. For these two human actions "wipe face" and "cross hands in front", different local body parts may share homogeneity, e.g., the torso parts for both actions share similar features, and thus they share the same memory blocks. However, when learning the right arms part of "cross hands in front", the past knowledge learned from "wipe face" becomes less relevant. Consequently, to preserve the past knowledge and to effectively learn new knowledge, the proposed Else-Net dynamically explores brand new memory blocks to memorize the new knowledge from the right arm part.

rent and previous tasks, the authors can guarantee that when new unseen tasks coming in, the gradients of the newly input tasks do not violate the previous tasks, which thus mitigates the catastrophic forgetting problems of the continual learning model.

Episodic Memory introduced in [20] stored the previously learned task samples and the corresponding task descriptors. The authors used the task descriptors in the episodic memory to provide contextual information for all tasks, to constrain the gradient vector angles between the previous and current tasks in order to mitigate catastrophic forgetting problems.

MS-G3D proposed by Liu *et al.* [19] contains **(1)** a disentangled multi-scale aggregation scheme to eliminate redundant dependencies between node features from different neighbourhood, and capture long-range joint relationships on human skeletons, and **(2)** a unified spatial-temporal modelling scheme to extract spatial and temporal features simultaneously by modelling with cross-spacetime edges in a spatial temporal graph.