Viability of Convolutional Variational Autoencoders for Lifelong Class Incremental Similarity Learning

Jiahao Huo[1] and Terence van Zyl[1]

University of Johannesburg, Johannesburg GT, 2092, South Africa
216045414@student.uj.ac.za, tvanzyl@uj.ac.za

Abstract. Incremental similarity learning in neural networks poses a challenge due to catastrophic forgetting. To address this, previous research suggests that retaining "image exemplars" can proxy for past learned features. Additionally, it is widely accepted that the output layers acquire task-specific features during later training stages, while the input layers develop general features earlier on. We lock the input layers of a neural network and then explore the feasibility of producing "embedding" models from a VAE that can safeguard the essential knowledge in the intermediate to output layers of the neural network. The VAEs eliminate the necessity of preserving "exemplars". In an incremental similarity learning setup, we tested three metric learning loss functions on CUB-200 (Caltech-UCSD Birds-200-2011) and CARS-196 datasets. Our approach involved training VAEs to produce exemplars from intermediate convolutional and linear output layers to represent the base knowledge. Our study compared our method with a previous technique and evaluated the baseline knowledge (Ω_{base}), new knowledge (Ω_{new}), and average knowledge (Ω_{all}) preservation metrics. The results show that generating exemplars from the linear and convolutional layers is the most effective way to retain base knowledge. It should be noted that embeddings from the linear layers result in better performance when it comes to new knowledge compared to convolutional embeddings. Overall, our methods have shown better average knowledge performance ($\Omega_{all} = [0.7879, 0.7805]$) compared to iCaRL ($\Omega_{all} = [0.7476, 0.7683]$) in the CUB-200 and CARS-196 experiments, respectively.Based on the results, it appears that it is important to focus on embedding exemplars for the intermediate to output layers to prevent catastrophic forgetting during incremental similarity learning in classes. Additionally, our findings suggest that the later linear layers play a greater role in incremental similarity learning for new knowledge than convolutions. Further research is needed to explore the connection between transfer learning and similarity learning and investigate ways to protect the intermediate layer embedding space from catastrophic forgetting.