MAXIM BAZHENOV, UC San Diego
Sleep: from biological to artificial systems

Artificial neural networks are known to exhibit a phenomenon called catastrophic forgetting, where their performance on previously learned tasks deteriorates when learning new tasks sequentially. In contrast, human and animal brains possess the remarkable ability of continual learning, enabling them to incorporate new information while preserving past memories. Empirical evidence indicates that sleep plays a crucial role in the consolidation of recent memories and safeguarding against catastrophic forgetting of previously acquired knowledge. Here we tested the hypothesis that implementing a sleep-like phase in artificial neural networks can protect old memories during new training and alleviate catastrophic forgetting. Sleep was implemented as off-line training with local unsupervised Hebbian plasticity rules and noisy input. In an incremental learning framework, sleep was able to recover old tasks that were otherwise forgotten. Previously learned memories were replayed spontaneously during sleep, forming unique representations for each class of inputs. Representational sparseness and neuronal activity corresponding to the old tasks increased while new task related activity decreased. In the weight space, sleep moved the system towards the region representing the intersection of the loss function minima for individual tasks. Our study sheds light on a potential synaptic weight dynamics strategy employed by the brain during sleep to enhance memory performance for continual learning.