Do Your Best and Get Enough Rest for Continual Learning

Hankyul Kang,[1] Gregor Seifer,[1] Donghyun Lee,[2] Jongbin Ryu[1]

[1]Ajou University, [2]KAIST

Motivation: Forgetting Curve Theory

- Ebbinghaus's theory says, "human memory tends to fade as time goes on; however, memory retention can be improved by repeated learning in an ***optimal recall interval***."
- We apply his theory to neural networks in continual learning scenarios.

	Within February	After 6 months
Person A	Memorize 10 words every **2** days.	Remembers only **3** words.
Person B	Memorize 10 words every **4** days.	Remembers **10** words.

Figure 1. **Hermann Ebbinghaus** (1850-1909, German psychologist)

https://en.wikipedia.org/wiki/Hermann_Ebbinghaus

Motivation: Forgetting Curve Theory

- Expanding the ***recall interval*** improves long-term memory retention of neural networks by repeatedly recalling memory with moderate difficulty, whereas an excessive recall interval decreases it.
- Long-term ***recall intervals*** lead to a high degree of forgetting.

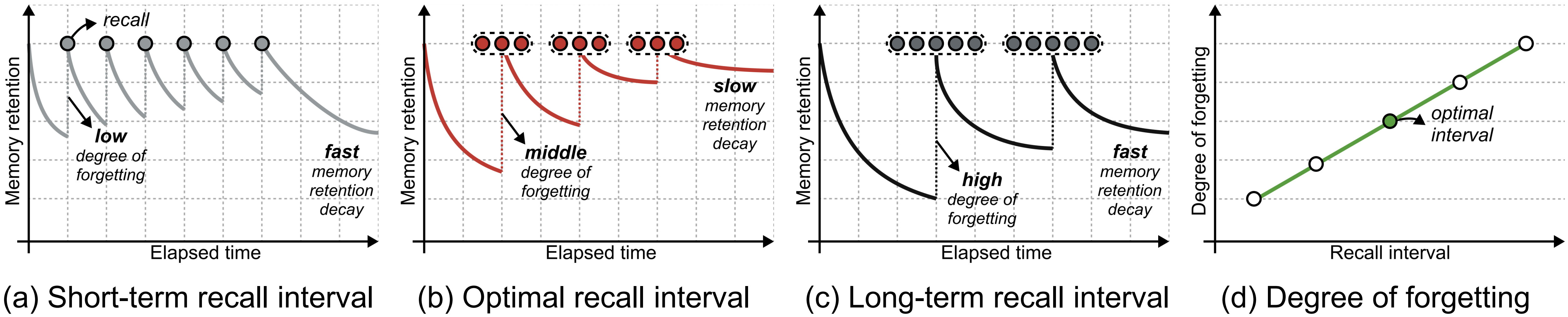


Figure 2. **Conceptual Graph of Forgetting Curve.**

Method: View Batch Model

- We structure a view-batch to have multiple views of a single sample.
- The ***recall interval*** of a single sample increases with the view-batch.

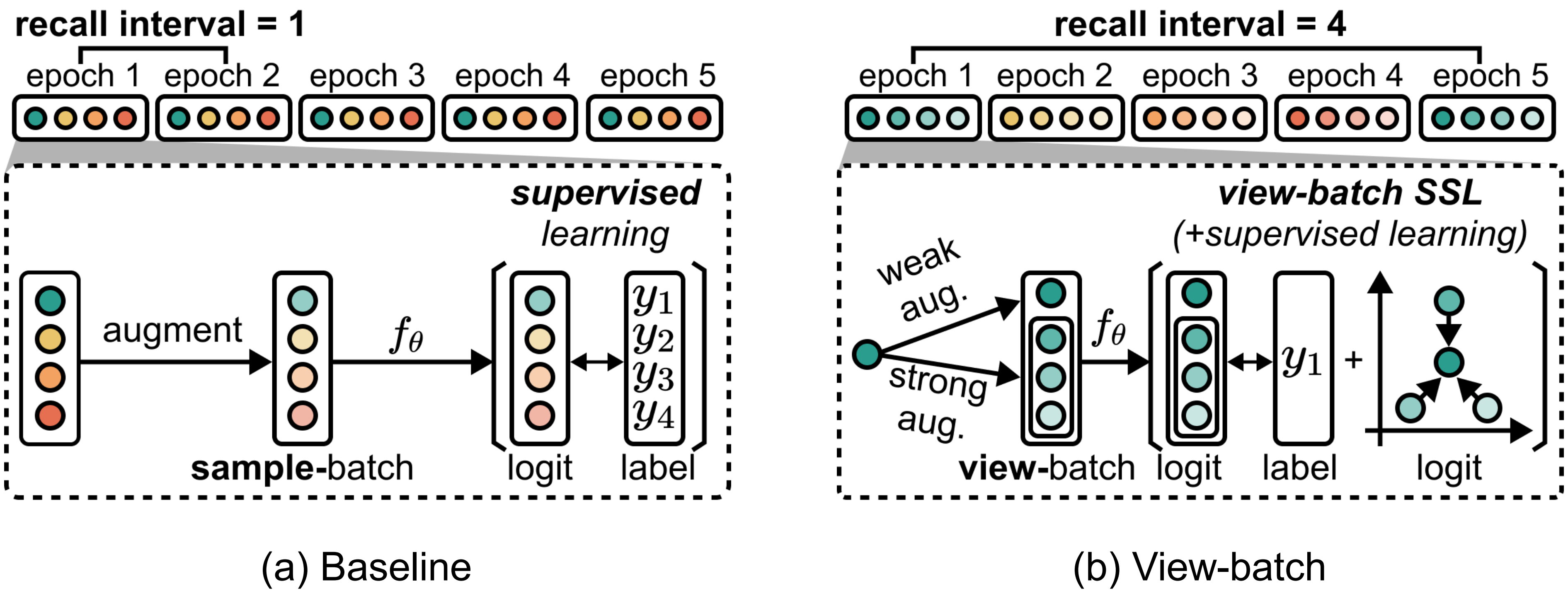
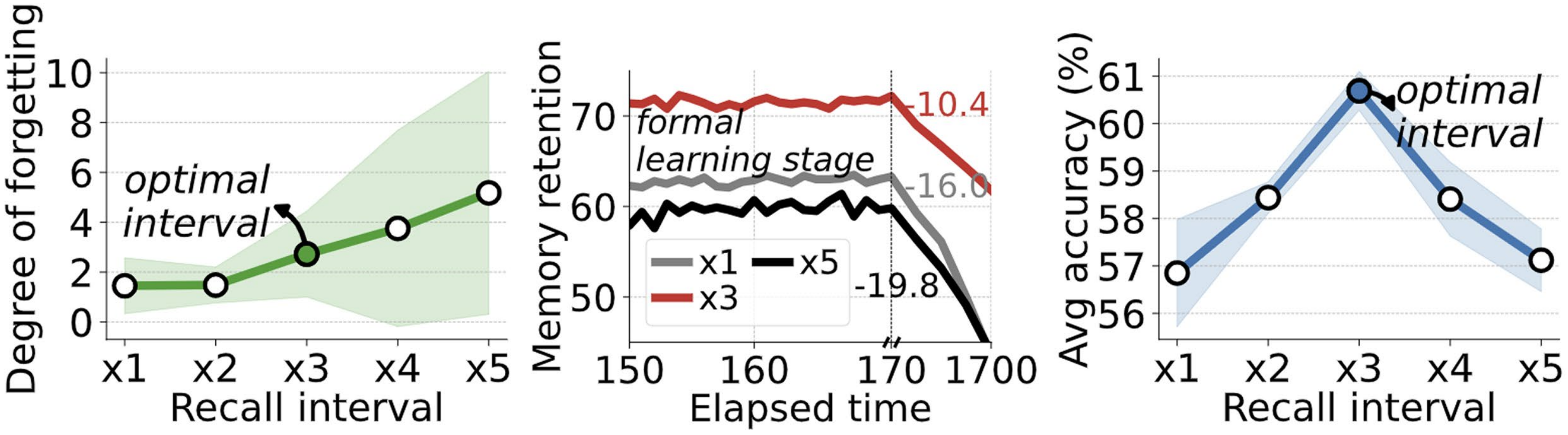


Figure 3. **Schematic illustration of the proposed view-batch model.** (a) trains whole samples (denoted as different colors) per epoch. In contrast to (a), (b) learns multiple views of the same sample (marked as different shades) to ensure enough time-space between recall intervals.

Empirical Findings

- We show that (a) the degree of forgetting increases with ***recall intervals***.
- We demonstrate that x3 achieves (c) the best performance, thanks to (b) the slow memory retention decay.



(a) Degree of forgetting (b) Forgetting curve (c) Accuracy

Figure 4. **Empirical findings.**

Comprehensive benchmark

- We provide comprehensive comparisons of various factors for continual learning. TCIL is used as a baseline method on the CIFAR100 dataset.

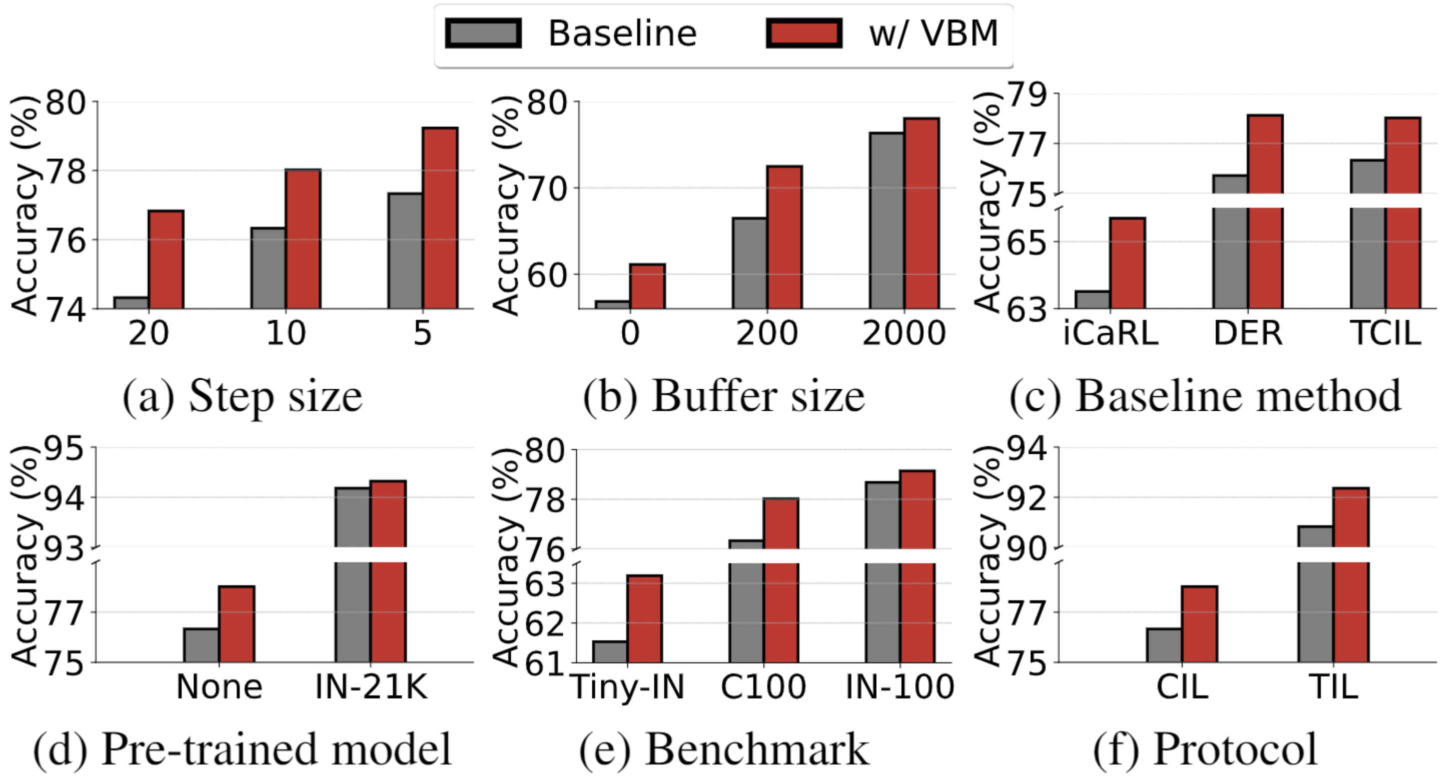


Figure 4. **Comprehensive benchmark.**

Conclusion

- Contribution**:** We propose the view-batch model that ***optimizes the recall interval between retraining samples in order to enhance memory decay*** in the continual learning task.

- Benefits**:** We showcase ***the effectiveness of our approach on various state-of-the-art continual learning methods***, where ours consistently improves the performance.

- Drawbacks**:** We limit the experimental scope to the continual learning tasks.