# CHAPTER V
# CONCLUSION

## 5.1. Conclusion

Based on the research results that have been conducted, the application of the SimAM-EMA hybrid attention module on the ResNet-50 architecture for image classification at 64 × 64 pixel resolution, the following conclusions can be drawn.

1. The integration of SimAM and EMA modules into ResNet-50 bottleneck blocks was successfully implemented with a stable training process without structural conflicts. SimAM is placed after Conv2 and before BN2 to perform feature weight modulation before normalization, while EMA is placed after Conv3 before the residual summation operation to capture cross-channel dependencies in the normalized representation. In the Hybrid architecture, SimAM and EMA attention modules are selectively placed on stage 3 and stage 4 that process high-level semantic representations. The characteristic difference between SimAM which is parameter-free and EMA which is randomly initialized became the basis for applying a differential learning rate and an EMA freeze mechanism during the first three epochs. This strategy was applied to maintain optimization stability between the pretrained backbone and the EMA module, so that the training process can proceed more effectively and stably without representational degradation in the early training phase.
2. The Hybrid SimAM-EMA model produces the best performance with Top-1 accuracy of 77.84% and Top-5 of 93.44% on the Tiny ImageNet-200 dataset. This model surpasses all ablation reference variants, namely increasing +1.23 percentage points compared to Baseline (76.61%), +1.61 percentage points compared to SimAM-Only (76.23%), and +3.66 percentage points compared to EMA-Only (74.18%). The difference between Top-1 and Top-5 accuracy across all models ranges between 15 and 18 percentage points, with the largest gap in Baseline (16.18 pp) and EMA-Only (17.70 pp), and the narrowest in Hybrid (15.60 pp). This substantial gap is a direct consequence of the characteristics of Tiny ImageNet-200 which contains 200 classes with high visual similarity and image resolution limited to 64 × 64 pixels. This condition causes models to often be able

to place the correct class in the top five candidates, but not always able to distinguish it at first rank consistently. The fact that the Top-1 to Top-5 gap in Hybrid is narrower than other models indicates that the combination of SimAM and EMA improves feature discriminability for the single most confident prediction, consistent with the theoretical argument in attention mechanisms that the superiority of channel attention mechanisms is more pronounced in Top-1 compared to Top-5. Further ablation analysis shows that pretrained ResNet-50 already has strong representational capacity so that performance improvement becomes more challenging. SimAM-Only shows that parameter-free modules have not been able to produce feature selection that is sufficiently discriminative for the 200-class distribution, while EMA-Only demonstrates that parametric modules with random initialization require appropriate training configuration so that the optimization process can proceed stably. These findings simultaneously confirm that the special training configuration in the Hybrid model is an integral part of the architectural design and not merely an external factor that provides additional advantages.

3. From the computational efficiency standpoint, the Hybrid SimAM-EMA model shows different characteristics across four main aspects, namely parameter count, GFLOPs, training time, and inference latency. In the parameter aspect, the parameter count only increased by +0.84% from 23.92M to 24.12M, where all additions come from the EMA module because SimAM is parameter-free. In the GFLOPs aspect, the Hybrid model experienced an increase of +23.3% from 8.18G to 10.08G, almost identical to EMA-Only (+23.2%), indicating that the theoretical computational complexity in the Hybrid architecture is dominated by EMA while SimAM's contribution to GFLOPs is relatively very small (+0.09%). In the training time aspect, the Hybrid model requires 1.04 hours for 20 epochs, higher than Baseline (0.71 hours) and SimAM-Only (0.73 hours), but significantly lower than EMA-Only (2.06 hours). This training time saving of 1.02 hours compared to EMA-Only was obtained through the strategy of selectively placing EMA on stage 3 and stage 4 which have smaller spatial resolution, so that the tensor manipulation overhead that is EMA's main bottleneck source can be substantially suppressed. In the inference latency aspect, Hybrid records 20.39 ms per image, lower than EMA-Only (22.70 ms) but far above Baseline (7.97 ms) and SimAM-Only (12.50 ms).

A significant discrepancy was also found between the theoretical GFLOPs increase and actual inference latency in SimAM, where the GFLOPs increase was only +0.09% but latency increased by up to +56.9%. This phenomenon indicates that elementwise operations can produce memory access and kernel launch overhead that is not reflected in GFLOPs calculations, so model efficiency evaluation needs to consider actual latency measurements in addition to GFLOPs alone. Viewed from the perspective of computational cost feasibility, the Top-1 accuracy improvement of 1.23 percentage points obtained by Hybrid can be considered proportional in certain contexts. It should be noted that EMA-Only with identical parameter and GFLOPs loads actually experienced an accuracy drop of 2.43%, so the performance difference between Hybrid and EMA-Only reaches 3.66 percentage points with exactly the same number of parameters. This shows that Hybrid's accuracy improvement is not merely the result of capacity addition, but from the quality of computational distribution. Nevertheless, the additional training time of +46.5% and inference latency of +155.8% compared to Baseline are real consequences that limit Hybrid's feasibility in applications with strict resource constraints or response time requirements. With these characteristics, the Hybrid model is more suitable for use in scenarios that prioritize accuracy with tolerance for higher training time and inference latency, such as medical image analysis or offline industrial quality inspection, while Baseline remains the more practical choice for real-time applications.

### 5.2. Recommendations

Based on the results of the application of the SimAM-EMA hybrid attention module on the ResNet-50 architecture for image classification at 64 × 64 pixel resolution found in this study, there are several recommendations that can serve as references for further research.

1. Exploration of Module Integration Positions. This study used one module placement configuration selected based on initial architectural analysis. Further research is recommended to explore variations in SimAM and EMA integration positions within bottleneck blocks, for example placing EMA before the residual operation at different positions or applying SimAM at earlier stages to identify

whether there are placement configurations that produce more optimal representational interactions.

2. Testing on More Diverse Architectures and Datasets. The findings of this study were obtained in the context of ResNet-50 on Tiny ImageNet-200 at 64 × 64 pixel resolution. To test the generalizability of the SimAM-EMA hybrid module characteristics, further research is recommended to apply the same architecture on different backbones such as ResNet-101 or EfficientNet, as well as on datasets with different resolutions and class distributions, to verify whether the complementary interaction pattern between SimAM and EMA is consistent across contexts.
3. Inference Efficiency Optimization. The Hybrid model's latency profile reaching 20.39 ms per sample limits its application in real-time scenarios. Further research can explore other compression techniques on the trained Hybrid model, with the aim of maintaining the accuracy advantages already achieved while simultaneously reducing latency overhead so that this architecture is feasible to apply on devices with limited computational capacity.