# 효율적인 비전 맘바 동작에 관한 연구

이현진, 이종석*

연세대학교

hyeonjin.lee@yonsei.ac.kr, *jong-seok.lee@yonsei.ac.kr

# A Study on the Efficient Operation of Vision Mamba

Hyeonjin Lee, Jong-Seok Lee*

Yonsei University

요 약

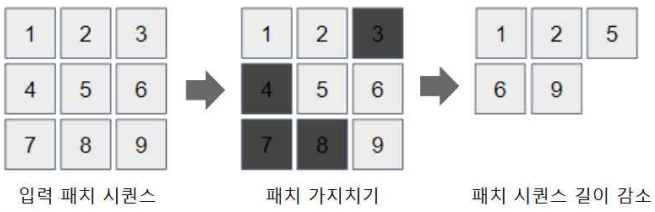
영상 처리 작업을 위한 인공신경망 모델로 새롭게 등장한 비전 맘바 모델은 그 효과적인 학습 방식 때문에 크게 주목을 받고 있다. 학습된 비전 맘바 모델을 더욱 효과적으로 동작시키기 위한 방법으로, 본 연구에서는 패치 가지치기 기술을 제안한다. 기존의 비전 트랜스포머 모델을 대상으로 연구되었던 패치 가지치기 기술을 비전 맘바 모델에 적용하기 위하여, 비전 맘바 모델이 여러 이미지 샘플들에서 일관되게 중요하게 여기는 패치를 판단하는 방법을 설계하고 이를 기반으로 모델이 처리해야 하는 패치 시퀀스의 길이를 줄여 비전 맘바 모델의 효율적인 동작을 구현한다.

## Ⅰ. 서 론

딥러닝 (deep learning) 기술이 발전함에 따라, 다양한 영상 처리 작업을 수행하는 인공신경망 모델 역시 계속해서 발전하였다. 다층 퍼셉트론, 합성곱 신경망, 비전 트랜스포머 등의 모델들이 차례로 영상 처리 인공신경망 모델의 뼈대로서 연구되었는데, 최근에는 비전 맘바 모델 [1]이 새롭게 등장하면서 이에 대한 연구가 활발히 진행되고 있다.

비전 맘바 모델은 이미지를 패치 시퀀스 (patch sequence)로 변형하여 처리하는 모델로서, 기존의 비전 트랜스포머 모델보다 더 효율적인 연산으로 높은 성능의 모델을 학습시킬 수 있다는 강점이 있어 크게 주목받고 있다. 하지만 여전히 많은 계산 비용을 수반하는 영상 처리 작업의 특성상, 학습된 모델의 더욱 효율적인 동작을 위해 추가적인 연구가 필요하다. 이를 위하여 비전 맘바 모델의 경량화 연구 [2]도 등장을 했으나, 이는 모델의 구성 파라미터의 개수 자체에 변형을 주는 방법이기 때문에 떨어진 성능의 모델을 다시 기존의 좋은 성능의 모델로 복원시키는 것이 불가능하다는 단점이 존재한다.

본 논문에서는, 패치 가지치기 기술을 적용해 비전 맘바 모델은 변형시키지 않으면서 모델의 효율적인 동작은 가능하게 만드는 방법을 제안한다. 패치 가지치기 기술은 중요하지 않은 패치들은 모델의 연산에 사용하지 않음으로써 모델은 변형시키지 않고 모델의 연산량만 줄이는 방법이다. (그림 1) 하지만 패치 가지치기 기술은 비전 트랜스포머 모델의 효율적인 동작을 위해 연구되었던 기술이기 때문에, 패치의 중요도를 판별하는 기준이 비전 트랜스포머의 연산 방법에 맞추어져 있다. [3] 이 때문에 기존 비전 트랜스포머 모델의 패치 가지치기 기술을 비전 맘바 모델에 바로 적용하기에는 그 한계가 존재한다. 따라서, 비전 맘바 모델에 적용할 수 있는 패치 가지치기의 새로운 기준을 수립하고, 이 기준에 따라 중요하지 않은 패치들은 가지치기하는 기술을 구현하여, 모델의 변형 없이도 비전 맘바 모델의 영상 처리 작업 연산량을 줄이는 방법을 설계한다.




〈그림 1〉 패치 가지치기를 통해 줄어드는 패치 시퀀스

## Ⅱ. 본론

앞서 서론에서 서술했듯, 기존의 패치 가지치기 기술은 비전 트랜스포머 모델을 대상으로 연구된 기술이기 때문에 이를 비전 맘바 모델에 적용하기 위해서는 패치의 중요도를 판단하는 기준이 달라져야 한다. 비전 트랜스포머 모델에서만 얻을 수 있는 어텐션 점수 (attention score)를 기반으로 패치 가지치기를 진행하는 대신, 본 논문에서는 모델 출력값에 대한 패치의 그래디언트 (gradient)를 기준으로 패치 가지치기를 진행한다. 모델의 출력값이 입력 패치에 대해 얼마나 변화하는지를 계산한 그래디언트 값의 절댓값이 클수록 해당 패치가 예측 결과에 미치는 영향력이 더 큰 중요한 패치라고 가정한다.

본 논문에서는 매 입력 이미지 샘플마다 패치의 중요도를 판단하지 않고, 여러 이미지 샘플들에 공통으로 적용할 수 있는 패치의 중요도를 구한 뒤 비전 맘바 모델에 한 번의 패치 가지치기 작업을 수행한다. 이를 위하여 학습된 비전 맘바 모델이 일관되게 중요한 패치로 여기는 패치의 위치를 미리 파악해야 하므로, 모집단 이미지 샘플들이 비전 맘바 모델을 통과할 때의 패치의 중요도를 우선 평가한다. 이를 기준으로 학습된 비전 맘바 모델을 구성하는 각 층에서 일관되게 중요한 위치의 패치만 남기고 패치 가지치기를 실행한다.

제안하는 패치 가지치기 방법의 효과를 확인하기 위해, 본 논문에서는 대표적인 비전 맘바 모델 ViM [4]의 Tiny 모델에 이 패치 가지치기 방법을 적용해 실험한다. 실험한 영상 처리 작업은 CIFAR100 이미지 데이터셋 [5]에 대한 이미지 분류 작업이다. 그 결과, 제안한 패치 가지치기 기술을 통해 비전 맘바 모델이 처리하는 패치 시퀀스의 길이가 단계적으로 줄어들면서, 모델의 이미지 분류 정확도는 76.07 %에서 71.08 %로 약간 줄어드는 것에 비해 모델의 총연산량 수치는 4.04 GFLOPs에서 3.25 GFLOPs로 크게 감소하는 것을 확인할 수 있다.

## Ⅲ. 결론

딥러닝 기술의 발전에 따라 새롭게 등장한 비전 맘바 모델은, 기존의 비전 트랜스포머 모델보다 더 효율적으로 영상 처리 작업을 학습시킬 수 있는 모델이다. 본 논문에서는 이 학습된 비전 맘바 모델을 더욱 효율적으로 동작시킬 수 있는 방법으로서, 비전 트랜스포머 모델에서 연구되었던 패치 가지치기 기술을 비전 맘바에 적용할 수 있는 방식으로 설계하였다. 제안한 패치 가지치기 기술을 실제 비전 맘바 모델에 적용한 뒤 이미지 분류 작업을 수행시킨 결과, 모델이 처리해야 하는 패치 시퀀스 길이를 단계적으로 줄임으로써 모델의 총연산량을 크게 줄여 효율적인 영상 처리 작업 수행이 가능하였다.

## ACKNOWLEDGMENT

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00453301, RS-2024-00453883)

## 참 고 문 헌

[1] Zhang, H., Zhu, Y., Wang, D., Zhang, L., Chen, T., and Ye, Z. “A survey on visual mamba.” Applied Sciences, 14(13), 5683. (2024)

[2] Lei, X., Zhang, W., and Cao, W. “DVMSR: Distillated Vision Mamba for Efficient Super-Resolution.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 6536 - 6546 (2024)

[3] Tang, Y., Han, K., Wang, Y., Xu, C., Guo, J., Xu, C., and Tao, D. “Patch slimming for efficient vision transformers.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12165 - 12174 (2022)

[4] Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., and Wang, X. “Vision Mamba: Efficient visual representation learning with bidirectional state space model.” In Proceedings of the 41st International Conference on Machine Learning (ICML), (2024)

[5] Krizhevsky, A., and Hinton, G. “Learning multiple layers of features from tiny images” Master's thesis, Department of Computer Science, University of Toronto. (2009)