ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 23-03-07
j.inst.Korean.electr.electron.eng.Vol.27,No.3,251~257,September 2023
https : //doi.org/10.7471/ikeee.2023.27.3.251
39

비전 트랜스포머 성능향상을 위한 이중 구조 셀프 어텐션

A Dual-Structured Self-Attention for improving the Performance of Vision Transformers

이 광 엽*, 문 환 희*, 박 태 룡*★

Kwang-Yeob Lee*, Hwang-Hee Moon*, Tae-Ryong Park*★

Abstract

In this paper, we propose a dual-structured self-attention method that improves the lack of regional features of the vision transformer's self-attention. Vision Transformers, which are more computationally efficient than convolutional neural networks in object classification, object segmentation, and video image recognition, lack the ability to extract regional features relatively. To solve this problem, many studies are conducted based on Windows or Shift Windows, but these methods weaken the advantages of self-attention-based transformers by increasing computational complexity using multiple levels of encoders. This paper proposes a dual-structure self-attention using self-attention and neighborhood network to improve locality inductive bias compared to the existing method. The neighborhood network for extracting local context information provides a much simpler computational complexity than the window structure. CIFAR-10 and CIFAR-100 were used to compare the performance of the proposed dual-structure self-attention transformer and the existing transformer, and the experiment showed improvements of 0.63% and 1.57% in Top-1 accuracy, respectively.

요 약

본 논문에서는 비전 트랜스포머의 셀프 어텐션이 갖는 지역적 특징 부족을 개선하는 이중 구조 셀프 어텐션 방법을 제안한다. 객체 분류, 객체 분할, 비디오 영상 인식에서 합성곱 신경망보다 연산 효율성이 높은 비전 트랜스포머는 상대적으로 지역적 특징 추출 능력이 부족하다. 이 문제를 해결하기 위해 윈도우 또는 쉬프트 윈도우를 기반으로 하는 연구가 많이 이루어지고 있으나 이러한 방법은 여러 단계의 인코더를 사용하여 연산 복잡도의 증가로 셀프 어텐션 기반 트랜스포머의 장점이 약화 된다. 본 논문에서는 기존의 방법보다 locality inductive bias 향상을 위해 self-attention과 neighborhood network를 이용하여 이중 구조 셀프 어텐션을 제안한다. 지역적 컨텍스트 정보 추출을 위한 neighborhood network은 윈도우 구조보다 훨씬 단순한 연산 복잡도를 제공한다. 제안된 이중 구조 셀프 어텐션 트랜스포머와 기존의 트랜스포머의 성능 비교를 위해 CIFAR-10과 CIFAR-100을 학습 데이터를 사용하였으며 실험결과 Top-1 정확도에서 각각 0.63%과 1.57% 성능이 개선되었다.

Key words : Vistion Transformer, locality inductive, self-attention, neighborhood network, context information

*Dept. of Computer Eng., Seokyeong University
★Corresponding author
E-mail: kylee@skuniv.ac.kr, Tel: +82-2-940-7745
※Acknowledgment:
This work was supported by Seokyeong University in 2022 and by Seokyeong University in 2023.
Manuscript received Sep. 04, 2023; revised Sep. 14, 2023; accepted Sep. 15, 2023.

Ⅰ. 서론

자연스러운 단어 시퀀스 모델을 찾아내는 자연어처리 신경망은 연속형 데이터를 잘 처리하는 순환신경망(RNN)으로 설계된다. 순환신경망은 시계열 특징을 갖고 있어서 이전 단계의 출력이 다음 단계의 출력을 결정하는 데 사용되기 때문에 과거 출력값이 잘 기억되어 있으면 다음 단계를 잘 결정할 수 있다. 그러나, 많은 단계가 진행되면서 점차 과거 출력값이 사라지는 기울기 소멸 문제(vanishing gradient problem)가 발생한다. 이를

해결하기 위해 RNN 에서는 LSTM과 같은 복잡한 은닉층을 사용하게 되는데 이러한 은닉층은 계산의 복잡도가 높아 순환신경망 처리 시간을 높이고 많은 파라메터를 사용하여 메모리 사용량이 크게 증가하는 문제가 있다.

이러한 문제를 해결하는 방법으로 입력되는 모든 단어를 동시에 고려하는 어텐션을 기반으로 하는 트랜스포머가 시계열 신경망의 새로운 기술로 등장하였다[1]. 트랜스포머는 긴 길이의 시퀀스 에서도 RNN에 비하여 장점이 있다. 트랜스포머는 내부 모듈에서 입력을 출력에 더하는 잔차 연결(Residual connection)이 사용되었다. 잔차 연결이 없는 LSTM은 적은 층의 신경망 구조만 가능해도 트랜스포머는 수십 층의 구조 설계도 가능한 장점이 있다.

트랜스포머의 단순한 구조는 지금까지 이미지 객체 분류, 인식 등에서 대표적인 솔루션으로 자리잡고 있는 합성곱 신경망(CNN)을 대체하기 시작했다. 특히, 해상도가 높은 이미지를 합성곱 신경망에서는 한 번에 입력하여 처리하기에 연산 복잡도와 메모리 사용량이 매우 커서 효과적인 기술이라 할 수 없다. 그러나 패치 단위로 분할하여 입력한 후 동시에 인코딩이 가능한 트랜스포머 구조를 적용하면 다양한 스케일의 입력 이미지를 효과적으로 처리할 수 있어 비전 신경망으로 많은 연구와 활용이 되기 시작하였으며 이를 비전 트랜스포머라고 부르고 있다.

어텐션 기반의 비전 트랜스포머는 입력값을 임베딩하여 동시에 인코딩한 후 디코딩 과정을 거쳐 어텐션 스코어 값이 출력된다. 이 과정에서 디코딩 출력이 인코딩의 어텐션 쿼리(Q)값이 되는 것이 일반적인 구조다. 그러나, 인코딩 자신의 값으로 쿼리(Q), 키(K), 밸류(V)를 만드는 셀프 어텐션이 연산량이 적고 처리 시간이 짧아 트랜스포머에서 널리 사용하고 있다. 그러나 512길이의 쿼리를 사용하는 비전 트랜스포머에서 셀프 어텐션의 경우에도 식(1)과 같은 연산이 요구되기 때문에 파라메터의 개수가 크게 늘어나 파라메터를 저장하는 메모리 사용량이 크게 요구된다. 이러한 단점을 보완하기 위해 non-local network를 사용하면 파라메터를 절반으로 줄일 수가 있다.

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

또한, 어텐션 기반의 트랜스포머는 합성곱 신경망에 비하여 지역적 특징추출에 큰 단점이 있다. 어텐션은 쿼리와 모든 위치의 임베딩 정보의 관계를 계산하기 때문에 전역적 특징추출에는 매우 우수한 성능을 보이지만 커널을 사용하는 합성곱 신경망처럼 지역적 특징추출에 낮은 성능을 보인다. 셀프 어텐션의 대체 방법으로 사용되는 non-local network의 경우에서도 지역적 특징추출에 높은 성능을 보일 수 없는 구조이다. 따라서, 트랜스포머에 높은 지역적 특징추출 성능을 보완하는 방법이 요구된다.

본 논문에서는 셀프 어텐션 기반의 트랜스포머의 복잡한 연산량 및 메모리 사용량을 줄이고 지역적 특징추출 성능을 높이는 방법으로 neighbor hood network를 혼용하는 이중 구조 트랜스포머를 제안한다. 기존의 트랜스포머와 성능을 비교하는데 CIFA-10 데이터를 학습 및 테스트에 이용한다.

II. 비전 트랜스포머 선행연구

이미지에서 객체 분류, 인식을 위해서는 이미지의 전역 context와 지역 context 정보를 고루 추출하는 방법이 요구된다. 이와 관련된 연구가 최근에도 많이 진행되고 있으며 그 가운데 이미지를 픽셀별로 클래스를 분류하는 semantic segmentation 기술이 기초가 된다. semantic segmentation은 대부분 완전결합 합성곱 신경망격자 4*44*4(FCN)을 기반으로 하여 픽셀 단위의 end-to-end 신경망을 사용하고 SegNet, U-Net과 같은 인코더-디코더 신경망 구조가 우수한 성능을 보인다.

그러나, 인코더는 강한 특징을 추출하기 위해 특징맵의 크기를 점차 줄여나가기 때문에 특징 이미지의 해상도가 낮아져 최종적으로 segmentation의 성능이 높아지지 않는다. 이를 해결하기 위해 [2] 연구에서 전역 context 정보와 지역 context 정보를 분리한 후 재결합하는 방법을 제시하였다. Xception CNN을 이용하여 입력 이미지의 특징 맵을 만들고 두 개로 분리된 패스를 통하여 전역과 지역 context 정보를 각각 생성한 후 결합하여 예측값을 만들어 낸다. 결합된 정보에 합성곱 이미지 픽셀의 잔차 결합과 채널 어텐션 기법을 추가하여 예측값의 정밀도를 높였다. 이 연구에서 제시한 전역 정보와 지역 정보를 분리하여 생성한 후 결합을 통하여 context 정보의 정밀도를 높이는 방법은 비전 트랜스포머의 단점을 보완하는데 사용할 수 있다.

기존의 비전 트랜스포머(ViT)는 입력 이미지가 다양한 스케일을 갖거나 해상도의 변화가 발생하면 잘 대응하지 못한다. 또한, 셀프 어텐션 실행을 모든 패치를 대상으로 하기 때문에 어텐션 코스트를 얻는데 연산량이 매우 커

지는 문제를 갖고 있다. [3] 연구는 이 문제 해결을 위해 패치를 일정크기로 묶어 윈도우라 칭하고 윈도우 범위에서 셀프 어텐션을 실행한다. 이 방법은 셀프 어텐션의 지역적 특징추출 성능을 강화하게 되고 윈도우 크기를 계층화하여 다양한 스케일의 입력 이미지에 대응할 수 있는 장점이 있다. 그러나 윈도우의 경계에 있는 지역적 특징은 무시될 수 있어서 그림 1처럼 cyclic shift 방법을 이용하여 윈도우 위치를 대각선 방향으로 이동함으로써 이웃한 윈도우와 겹쳐지는 특징을 얻는 방법도 개발하였다. [3]는 inductive bias가 약점인 비전 트랜스포머 구조에 locality inductive bias를 강화하는 대표적 연구가 되었다.

최근에는 [3]을 개선하는 연구가 지속으로 이루어지고 있는데 LG transformer[4]는 셀프 어텐션을 Local-Global(LG) 어텐션 블록으로 개선하였다. LG 블록은 shift 윈도우의 크기를 3단계로 계층화하여 local과 global 특징추출에 강점을 갖도록 하였다. PLG-ViT [5]는 Local-Global 셀프 어텐션을 병렬로 처리하여 실행시간을 줄이는 결과를 보였다. 이와같이 대부분의 연구가 shift 윈도우를 이용한 locality inductive bias 개선

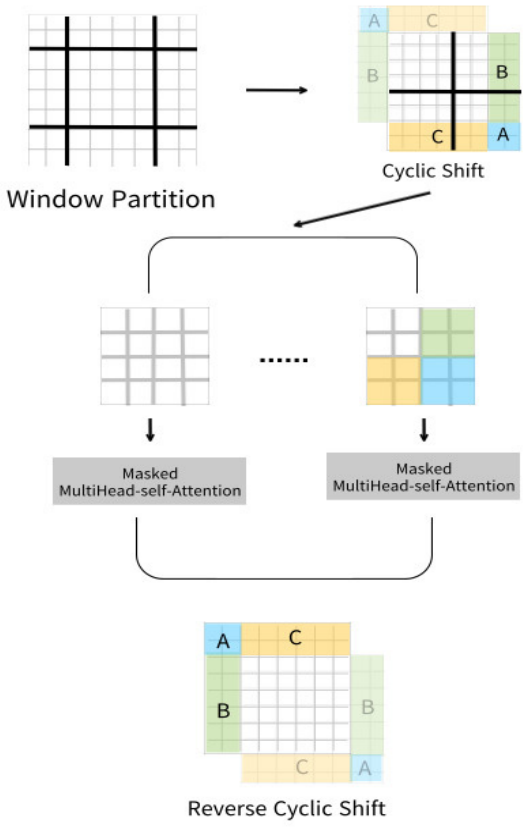

Fig. 1. Cyclic shifted window structure for attention.
그림 1. Cyclic shifted window구조 어텐션

과 윈도우 크기의 계층화를 통하여 local과 global 어텐션을 동시에 달성하는 방법을 제시하였다.

Ⅲ. 제안하는 비전 트랜스포머

기존 비전 트랜스포머의 셀프 어텐션이 갖는 약한 inductive bias의 단점을 보완하기 위해 주로 사용되는 시프트 윈도우 구조는 패치를 묶어 윈도우를 만들기 때문에 패치 분할과 결합 처리 과정이 필요하다. 또한, 지역적 특징과 전역적 특징을 모두 추출하기 위해 윈도우의 크기를 계층화 하면 계층마다 down sampling과 up sampling 과정이 필요하며 윈도우 시프트에 대한 처리 과정도 포함된다.

본 논문에서는 시프트 윈도우를 사용하지 않는 셀프 어텐션 구조에 neighborhood network를 활용하여 inductive bias 문제를 해결하는 새로운 비전 트랜스포머 구조를 제안한다.

1. 셀프 어텐션 구조

자연어처리에서 사용되던 시계열 순환신경망에서 어텐션은 오래전에 처리된 입력에 어텐션을 부여하여 긴 시간 범위의 특징값을 고려할 수 있는 구조이다. 비젼 트랜스포머에서는 어떤 점(query)에서 입력된 패치의 전 영역의 특징값을 어텐션 하되 다른 패치가 아닌 자기 자신의 패치 영역에서 어텐션 하기 때문에 셀프 어텐션 구조라 한다[6].

셀프 어텐션은 모든 입력 시퀀스에서 얻어진 global 정보를 각 시퀀스에 반영하여 전 영역의 특징값에 어텐션 하게 된다. 어텐션의 입력은 식(2)로 표현된다. 여기에서 n은 시퀀스를 구성하는 entity 수이며, d는 한 개 entity의 임베딩 길이이다.

$$X=(x_1,x_2,\cdots,x_n),\ X\in R^{n\times d} \tag{2}$$

결국, 시퀀스로 구성된 입력의 각 entity를 global 정보로 인코딩한 후 모든 entity간의 상호의존도를 계산하는 것이다. 입력에서 어텐션값을 구하기 위해 식(3), (4), (5)와 같이 세가지 학습 파라메터를 정의하여 식(6)을 계산하여 어텐션값을 출력한다. 이때, $d_q=d_k$이면 셀프 어텐션이 된다.

$$\text{Queries} = \left(W^Q\in R^{d\times d_q}\right),\ Q=XW^Q \tag{3}$$

$$\text{Keys} = \left(W^K\in R^{d\times d_k}\right),\ K=XW^K \tag{4}$$

$$\text{Values} = \left(W^V \in R^{d \times d_v}\right),\ V = XW^V \tag{5}$$

$$Z = softmax\left(\frac{QK^T}{\sqrt{d_q}}\right)V,\ Z \in R^{n \times d_v} \tag{6}$$

식(3), (4), (5), (6)의 연산과정은 입력(X)와 학습 파라메터 W^Q, W^K, W^V의 dot product로 이루어지며 아래 그림 2와 같은 구조로 실행된다.

2. 멀티 헤드 셀프 어텐션 구조

입력 시퀀스에 있는 entity간의 다중 의존관계는 다중 셀프 어텐션로 구성할 수 있는데 이것을 멀티 헤드 셀프 어텐션 구조라 한다. h개의 셀프 어텐션 블록으로 구성될 때 헤드의 수가 h되고 멀티 헤드 셀프 어텐션의 전체 학습 파라메터는 식(7)으로 표현된다. 셀프 어텐션 각 블록에서 출력되는 어텐션 값을 모두 결합을 하면 식(8)이 되며 h개로 구성된 멀티 헤드 셀프 어텐션 구조의 출력이 된다. 그림 2의 h에 해당 된다.

$$\left\{W^{Q_i}, W^{K_i}, W^{V_i}\right\}, i = 0 \cdots (h-1) \tag{7}$$

$$\left[Z_0, Z_1, \cdots, Z_{h-1}\right] \in R^{n \times h \cdot d_v} \tag{8}$$

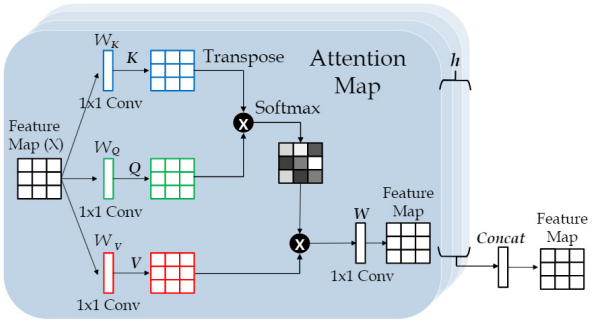

Fig. 2. Self Attention Score generation.
그림 2. 셀프 어텐션 스코어 생성

3. Neighborhood 어텐션 구조

윈도우 구조를 제안한 Swin에서는 그림 3 왼쪽과 같이 입력 이미지의 패치를 모두 일정크기로 묶어 윈도우라 정하고 윈도우를 대각선 방향으로 shift 시켜 주변 윈도우와 겹침 방법으로 윈도우 주변의 정보를 셀프 어텐션값 계산에 포함하여 local inductive biase를 해결하였다.

그러나 이 구조에서는 윈도우 단위로 나누어 셀프 어텐션을 연산하지만 윈도우가 shift 되면서 주변 윈도우들의 정보도 계산에 포함해야 하기 때문에 결국 그림 4의 상단과 같이 한 개의 query token은 윈도우의 모든 key와 value token과 연산을 해야 한다.

Neighborhood 어텐션[7]은 그림 3의 오른쪽과 같이 어텐션 포인트 주변을 묶어 neighborhood의 크기를 임의로 만들어 어텐션 연산을 하게 된다. 따라서, 그림 4의 하단과 같이 지정된 query의 주변(예를 들어, 3×3 token)에 해당하는 key, value와 어텐션 스코어를 계산하기 때문에 연산량이 많이 줄어든다. 즉, 식(6)의 셀프 어텐션을 지역화 하는 방법이다. 윈도우로 크기가 제한되지 않고 neighborhood의 영역 크기를 자유롭게 설정할 수 있기 때문에 local과 global 성분을 모두 포함 할 수도 있다. 즉, maximum neighborhood는 셀프 어텐션과 같아질 수 있다. 그림 4 하단의 연산과정은 식(3)~(5)와 동일 하지만 key, value의 크기가 윈도우 크기가 아닌 neighborhood 크기이다.

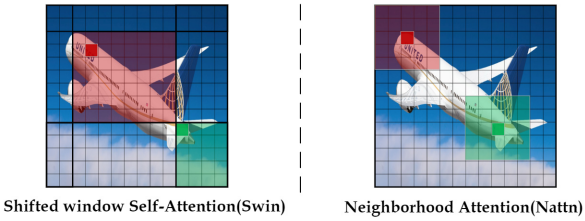

Fig. 3. Comparison of Attention Structures.
그림 3. Attention 구조 비교

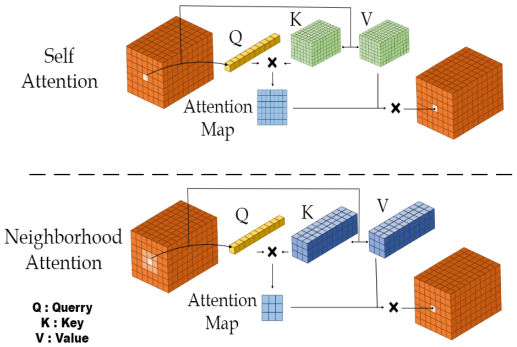

Fig. 4. Comparison of Attention Score Computation.
그림 4. 어텐션 스코어 연산 비교

4. 이중구조 셀프 어텐션 트랜스포머

트랜스포머는 입력으로 1차원 임베딩 데이터구조를 사용하기 때문에 입력 이미지 패치를 만들고 각 패치를 Linear Projection을 통하여 1차원 임베딩을 만들고 식(9)와 같이 표현된다. Linear Projection은 3×3 convolution 2개로 실행한다. 식에서 N은 트랜스포머의 시퀀스길이(패치수), P는 패치의 크기이다. H와 W는 원본 이미지 해상도이다.

$$x_p \in R^{N\times(P^2\times C)},\ N=\frac{HW}{P^2} \tag{9}$$

임베딩된 입력 이미지 패치의 1차원 벡터에는 패지 순서정보인 position embedding을 더한다.

제안하는 트랜스포머의 첫 번째 단계에서는 4×4 패치의 임베딩을 그림 5의 NAT 블록에서 neighborhood transformer 처리하여 local 특징에 attention된 score를 출력하도록 한다. NAT 블록은 4단계로 구성하고 3단계까지는 downsampler를 사용하여 특징맵의 크기를 반으로 줄이고 대신 채널 수를 2배를 늘리면 계층적으로 local 특징을 추출할 수 있다. neighborhood attention (NA)는 8개의 multi-head를 통하여 8개의 local attention 값을 얻어 낼 수 있다. NA의 앞과 뒤에는 Layer nomalization(LN)을 두어 작은 batch size에서도 효과적으로 성능을 높일 수 있다. 마지막 단계는 Multi Layer Perceptron(MLP)로 attention score를 분류하여 그림 5의 vision transformer encoder에서 global attention score를 출력하도록 한다. 여기에서는 neighborhood attention 대신 multi-head self attention이 적용된다.

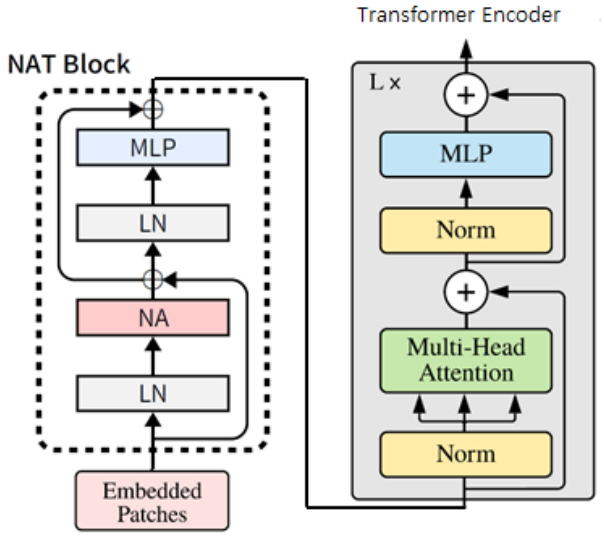

Fig. 5. Structure of a proposed self-attention.
그림 5. 제안하는 셀프 어텐션 구조

Ⅳ. 비젼 트랜스포머 성능측정

비젼 트랜스포머가 이미지 분류에서 좋은 성능을 내기 위해서는 이미지의 특징맵에 local 특징과 global 특징이 고루 담겨 있어야 한다. 그러나 기존의 ViT의 셀프 어텐션은 global 특징이 강조됨에 따라 locality inductive bias가 약화되어 이미지 분류에서 약점을 갖는다. 비젼 트랜스포머가 분류를 위해 사용하는 특징맵은 어텐션 메커니즘을 통하여 생성되기 때문에 어텐션 메커니즘의 구조에 따라 이미지 분류 성능이 결정된다.

본 논문에서는 local과 global 특징을 모두 고루포함하는 어텐션 메커니즘을 제안하고 이를 이용하는 비젼 트랜스포머가 이미지 분류에서 어떤 성능을 보이는지 몇 가지 다른 어텐션 메커니즘과 비교하였다.

이미지 분류에 사용되는 데이터 세트는 일반적으로 널리 알려진 CIFAR-10과 CIFAR-100으로 하였으며 비교에 사용된 어텐션 메커니즘은 NATTN +Non-Local Block(multiHead)는 global 특징추출에 장점이 있으면서 ViT에 비하여 연산 구조가 간단한 Non-Local 어텐션과 neighborhood를 결하고 multiHead 어텐션으로 설계된 메커니즘으로 본 논문에서 비교를 위하여 설계하였다. Pure -ViT[8]는 multiHead self-attention 이라는 기본적인 메커니즘을 적용한 비젼 트랜스포머이다. 본 논문에서 제안하는 메커니즘과 더불어 다양한 어텐션 구조를 비교한 결과 표 1과 표 2와 같이 본 논문에서 제안한 multiHead neighborhood +self-attention 이중 구조가 이미지 분류에서 Pure-ViT에 비하여 0.63% (CIFAR-10)와 1.57%(CIFAR-100) 개선되어 비교적 좋은 성능을 보였다. 성능 비교는 Top-1 accuracy로 하였으며 테스트 데이터를 이용한 학습 정확도는 그림 6과 그림 7에 나타내었다. 제안하는 방법은 다른 메커니즘에 비하여 학습에서 비교적 빠르게 수렴하는 것을 알 수 있다.

Table 1. Performance comparison of different attention mechanisms using CIFAR-10.
표 1. CIFAR-10을 사용한 다양한 어텐션구조에서 성능비교

Model	Top-1 acc. % (Data Set : CIFAR-10)
Proposed Attention (NATTN+Attention)	**94.42**
NATTN+Non-Local Block(multiHead)	92.24
Pure-ViT[8]	93.79

Table 2. Performance comparison of different attention mechanisms CIFAR-100.
표 2. CIFAR-100을 사용한 다양한 어텐션구조에서 성능비교

Model	Top-1 acc. % (Data Set : CIFAR-100)
Proposed Attention (NATTN+Attention)	**74.17**
NATTN+Non-Local Block(multiHead)	70.72
Pure-ViT[ ]	72.60

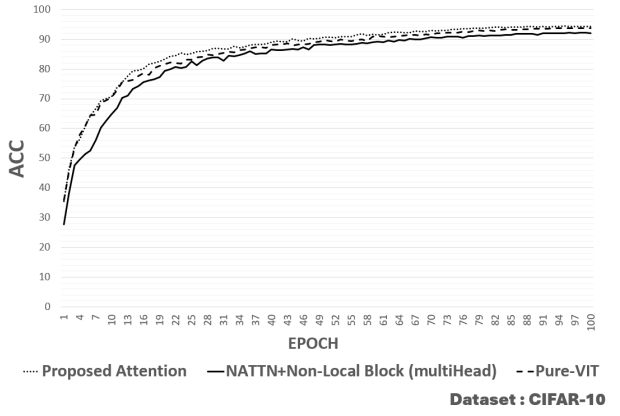

Fig. 6. Learning accuracy for CIFAR-10.
그림 6. CIFAR-10에 대한 학습 정학도

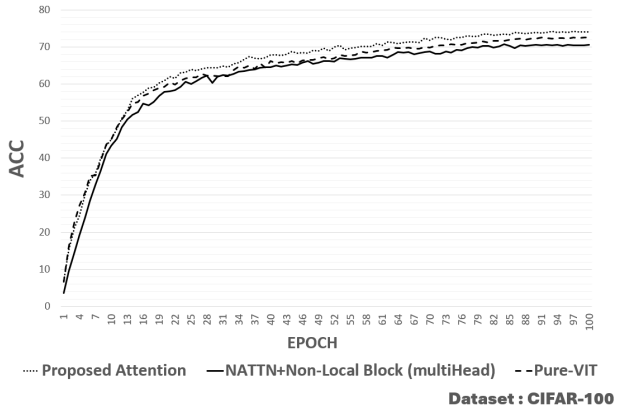

Fig. 7. Learning accuracy for CIFAR-100.
그림 7. CIFAR-100에 대한 학습 정학도

V. 결론

Attention is All You Need[1]논문이 나온 이후 어텐션 구조가 비젼 트랜스포머에서 혁신적인 모델이 되었다. 이후 비젼 트랜스포머에서 어텐션 메커니즘에 많은 연구가 되어왔으나 어텐션이 갖는 locality inductive bias문제가 존재하였으며 어텐션에서 locality를 향상하기 위한 다양한 연구가 지속되고 있다. Shifted 윈도우로 스캔하면서 윈도우내의 local 특징값을 셀프 어텐션과 결합하는 어텐션 메커니즘을 기반으로 하는 연구가 주류를 이루고 있으나 본 논문에서는 Query 주변에 이웃하는 적은 규모의 정보로 local 특징을 생성한 후 셀프 어텐션을 통과하면서 global 특징을 결합하는 어텐션 메커니즘을 제안하였다. 실험은 이미지 분류의 성능을 측정하였고 다양한 어텐션 메커니즘과 비교하였을 때 우수한 비젼 트랜스포머의 성능을 보였다.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, “Attention Is All You Need,” *31st Conf. on Neural Information Processing Systems(NIPS 2017),* 2017.
DOI: 10.48550/arXiv.1706.03762

[2] Chih-Yang Lin, Yi-Cheng Chiu, Hui-Fuang Ng, Timothy K. Shih, Kuan-Hung Lin, “Global-and-Local Context Network for Semantic Segmentation of Street View Images,” *Sensors,* Vol.20, No.10, 2020. DOI: 10.3390 /s20102907

[3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, “Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows,” *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.10012-10022, 2021.

[4] Jinpeng Li, Yichao Yan, Shengcai Liao, Xiaokang Yang, Ling Shao, “Local-to-Global Self-Attention in Vision Transformers,” 2021.
DOI: 10.48550/arXiv.2107.04735

[5] Nikolas Ebert, Didier Stricker, Oliver Wasenmuller, “PLG-ViT: Vision Transformer with Parallel Local and Global Self-Attention,” *Sensors,* Vol.23, No.7, 2023. DOI: 10.3390/s23073447

[6] B Yang, J Li, DF Wong, LS Chao, X Wang, Z Tu, “Context-aware self-attention networks,” *Proceedings of the AAAI conference on artificial intelligence,* 2019.
DOI: 10.48550/arXiv.1902.05766

[7] Ali Hassani, Steven Walton, Jiachen Li, Shen Li, Humphrey Shi, “Neighborhood Attention Transformer,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp.6185-6194, 2023.
DOI: 10.48550/arXiv.2204.07143

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, SylvainGelly, Jakob Uszkoreit, Neil Houlsby, “An image is worth 16×16 words: Transformersfor image recognition at

scale," *ICLR,* 2020.
DOI: 10.48550/arXiv.2010.11929

BIOGRAPHY

Kwang-Yeob Lee (Life Member)

1985 : BS degree in Electronics Engineering, Sogang University
1987 : MS degree in Electronics Engineering, Yonsei University.
1994 : PhD degree in Electronics Engineering, Yonsei University.

1989~1995.2 : Senior Researcher, Hyundai Electronics Inc.
1995.3~present : Professor, Dept. of Computer Engineering, Seokyeong University

Hwan-Hee Moon (Member)

2024 : BS degree candidate in Computer Engineering, Seokyeong University.

Tae-Ryong Park (Member)

1985 : Hangyang University, Dept. of Mathmatics(BS)
1987 : Hangyang University, Dept. of Mathmatics(MS)
1995 : Hangyang University, Dept. of Mathmatics(Ph.D)

1994~ : Seokyeong Univeristy, Dept. of Computer Engineering, Professor
〈Research interests〉 Crypto Algorithm, Computer Security, Computer Arithmetic, Recongnition Algorithm, Machine Learning