다중 모형 CNN 기반 응용 트래픽 분류

백의준, 김보선, 박지태, 신창의*, 김명섭
고려대학교, *국방기술품질원
{pb1069, boseon12, pjj5846, tmskim}@korea.ac.kr, superego99@dtaq.re.kr*

Multi-Shape CNN based Application Traffic Classification

Ui-jun Baek, Boseon Kim, Jee-Tae Park, Chang-Yui Shin*, Myung-Sup Kim
Korea University, *Defense Agency for Technology and Quality

요 약

인터넷 사용자의 수가 나날이 증가하고 점점 더 다양하고 복잡한 유형의 응용 또는 공격이 발생함에 따라 네트워크 트래픽 분류는 네트워크 관리 분야의 중요한 직무로 자리 잡았다. 그러나 가변 포트 번호 방식과 페이로드 암호화 기술의 도입은 전통적인 네트워크 트래픽 분류 기술에 명확한 한계점을 가져다주었다. 이에 통계적인 특징을 학습하여 응용 트래픽을 분류하는 머신러닝 기반 기술이 조명되었으며 최근에는 CNN을 이용한 딥러닝 기반 표현학습 방법이 주로 연구되고 있다. 기존 CNN 기반 응용 트래픽 분류에서는 입력 데이터로 패킷 원본 데이터를 가로와 세로의 비율이 같은 2차원 이미지 형태가 주로 고려되었으나 우리는 패킷 내 필드 값의 연속성과 패킷 데이터가 공간 정보가 없는 1차원 데이터임에 초점을 맞추었다. 우리는 패킷의 입력 벡터를 다양한 모형으로 변환하여 그에 대응하는 분류 모델을 생성하였으며 각 분류 모델의 성능을 비교한 결과 기존의 정사각형 모형의 입력 벡터보다 1차원 형태의 입력 벡터가 입력된 모델이 더 정확하게 응용 트래픽을 분류하였다. 또한 우리는 단일 입력 벡터를 다중 모형 입력 벡터로 변환하여 동시에 학습하고 앙상블 기반 응용 트래픽 분류 모델을 제안하였으며 제안한 모델은 단일 모형이 입력된 분류 모델보다 27% 높은 정확도를 보였다.

Ⅰ. 서 론

인터넷 사용자의 수가 나날이 증가하고 점점 더 다양하고 복잡한 응용 트래픽 또는 공격이 발생함에 따라 네트워크 트래픽 분류는 응용 타입 식별, 비정상 행위 탐지, QoS(Quality of Service) 보장 등 네트워크 관리 분야의 중요한 직무 중 하나로 자리매김하였다. 또한 네트워크 트래픽 암호화 기술이 인터넷 서비스 전 분야에 걸쳐 적용됨에 따라 전통적인 트래픽 분류 방법들을 통해 응용을 분류하는 것이 어려워졌고 이는 컴퓨터 네트워크 분야에서 정확하고 강인한 트래픽 분류 방법이 필요함을 강조하고 있다.

네트워크 트래픽 분류 분야의 대표적인 방법은 크게 포트 기반 분류, 페이로드 기반 분류, 머신러닝 기반 분류로 나눌 수 있다. 포트 기반 분류와 페이로드 기반 분류는 전통적으로 널리 쓰였던 분류 방법이나 가변 포트 번호의 도입과 트래픽 암호화 기술의 도입으로 인해 정확히 트래픽을 분류할 수 없다는 명확한 한계점을 가진다. 머신러닝 기반 분류는 현재 가장 많이 사용되고 있는 분류 방법의 하나며 트래픽 데이터로부터 추출된 통계적 특징을 학습하여 분류 모델을 생성한다. 그러나, 특징 데이터 차원 수가 증가하면 할수록 따라 학습이 제대로 수행되지 않는 한계점을 가지고 있으며 이러한 이유로 머신러닝 기반 분류에서는 분류 성능에 유의미한 영향을 미치는 특징이 무엇인지 선택하는 방법이 중요한 도전 과제로 남아있다. 최근에는 머신러닝 기반 방법 중 딥러닝 기반 방법이 많은 연구에서 고려되고 있으며 특히 컴퓨터 비전, 이미지 인식 분야에서 높은 성능을 발휘하는 CNN(Convolutional Neural Network) 기반 분류 방법이 가장 대표적이며 네트워크 플로우가 패킷의 시계열로 구성된 점을 이용하여 RNN(Recurrent Neural Network) 기반 분류 방법 또한 적용되는 추세이다. 일반적으로 CNN 기반 트래픽 분류에서는 네트워크 패킷 데이터를 2차원 이미지로 변환하여 분류 모델에 입력한다. 입력된 2차원 데이터는 여러 층의 Convolutional Layer와 Pooling Layer를 거치며 패킷의 특성을 포함하고 있는 특징 맵으로 변환된다. 그러나 네트워크 패킷은 1차원 데이터이며 일부 필드 값은 연속성을 지니고 있다는 점에서 2차원 이미지 구조와는 근본적인 차이가 있다. 우리는 입력되는 패킷의 필드 값이 패킷을 2차원으로 변환하는 과정에서 절단되어 해당 필드 값이 본래 가지고 있던 의미가 변형될 수 있다는 점에 주목했다. 그림 1은 패킷 데이터가 2차원 이미지로 변환하는 과정에서 필드 값이 손실되는 예시를 나타낸다.

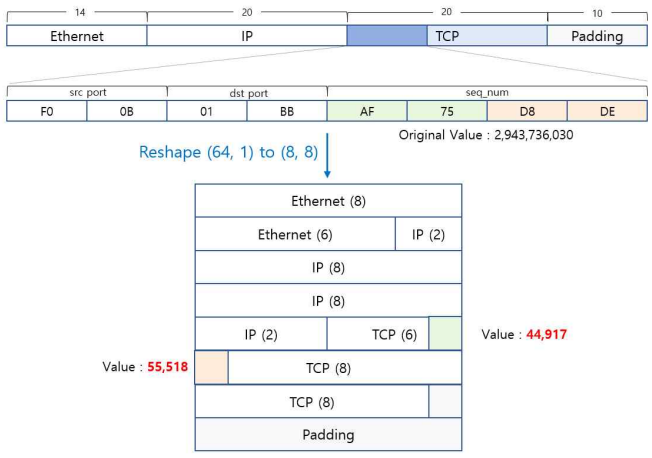

그림 1 패킷 데이터 변환 과정에서 필드 값 의미 손실 예시

본 논문은 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과 (2021RIS-004)이고 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구(NRF-2021S1A5C2A03097574)임

그림 1에서는 TCP 헤더의 필드 값의 의미가 손실되는 예시를 나타내었으나 유사하게 TCP 페이로드에서도 연속되는 문자열 중간이 절단되며 본래 의미에 손실이 발생할 수 있다. 또한, 인접한 픽셀값이 주로 유의미한 특징을 가지는 이미지와는 달리 패킷 데이터는 수직적인 관점에서 멀리 떨어진 필드 값과도 유의미한 관계가 있을 수 있다는 점도 고려 대상이다. 본 논문은 입력 데이터인 1차원 패킷 데이터를 다양한 모양의 2차원 데이터로 재구조화하고 이로부터 다양한 특징 맵을 생성하여 응용 트래픽을 분류하는 방법을 제안한다.

본 논문은 1장 서론에 이어 2장에서 관련 연구를 제시하고 3장에서 제안하는 방법을 설명한다. 4장에서는 실험 결과를 설명하고 5장에서 결론 및 향후 연구를 제시하는 것으로 본 논문을 마친다.

Ⅱ. 관련 연구

CNN은 많은 분야에서 사용되고 있는 가장 인기 있는 딥러닝 방법의 하나로 이미지 또는 텍스트 데이터의 지역적 정보 추출에 있어 강점을 가지며 네트워크 트래픽 분류 분야의 많은 연구에 활용되고 있다. [1]은 플로우 내 패킷들의 페이로드를 28*28의 2차원 이미지로 변환하여 CNN 모델에 입력하여 분류 모델을 생성하였다. 또한, 모델 생성에 필요한 하이퍼파라미터를 최적화하는 과정을 통해 10개의 웹 응용을 99.57%의 정확도로 분류하였다. [2]는 패킷을 전체 레이어(ALL)와 TCP 세션의 페이로드인 L7로 구분하여 서브데이터 세트를 생성하고 28x28의 2차원 이미지로 변환하여 분류 모델을 생성하였다. 결과적으로 10개 응용을 평균 99.41%의 정확도로 분류하였다. [3]은 [2]와 마찬가지로 패킷을 ALL과 L7로 구분하여 서브 데이터 세트를 생성하고 이를 784*1 모양으로 변환하고 학습하여 1D CNN 모델을 생성하였으며 28*28 형태의 이미지를 학습하여 2D CNN 모델을 생성하고 두 개 모델의 분류 성능을 비교하였다. 결과적으로 2D CNN 기반 모델의 분류 성능보다 1D CNN 기반 모델이 성능이 좋았으며 저자는 2D CNN의 2차원 공간 특징 학습이 1차원 암호화 트래픽 분류에서 명확하지 않다고 보고했다. [4]는 784*1 모양 데이터를 학습한 1D CNN과 28*28 모양 데이터를 학습한 2D CNN, 784*1 모양 데이터를 학습한 LSTM 모델을 생성하고 각 모델의 분류 성능을 비교하였으며 1D CNN 모델이 가장 좋은 분류 성능을 나타냈다. 또한 저자는 각 모델에서 분류한 결과들을 바탕으로 투표하여 최종 분류 결과를 생성하는 앙상블 방법을 제안했으며 제안한 앙상블 기반 모델의 분류 성능은 1D CNN 기반 분류 모델보다 6% 높은 분류 성능을 보였다. 대다수의 기존 연구들에서 1D CNN기반 분류 모델이 2D CNN 기반 분류 모델보다 트래픽 분류에 적합한 것을 확인할 수 있으나 현재까지도 많은 연구에서 주로 2D CNN 모델이 고려되고 있다. 이는 컴퓨터 비전, 이미지 객체 인식 등의 분야에서 CNN 모델의 발전을 선도하고 있으며 대다수의 첨단 연구가 2D CNN에 기반하고 있기 때문이라고 짐작해볼 수 있다. 컴퓨터 비전 분야의 이미지와 패킷 데이터의 구조 및 특성이 다른 만큼 패킷 데이터에 적합한 입력 모양을 정의하는 것은 필수적이다. 따라서 본 논문에서는 패킷 데이터의 입력 모양이 1차원인지 2차원인지 선택하는 문제를 더 확장하여 패킷 데이터의 어떤 모양이 정확한 분류에 있어 적합한지에 대한 비교실험을 수행하고 다양한 모양에 대응하는 모델의 분류 결과들을 집계하여 학습하는 앙상블 기반 응용 트래픽 분류 모델을 제안한다.

Ⅲ. 실험

A. 데이터 세트

본 실험을 위해 분류 실험을 위하여 50개 종류의 100개 플로우가 포함된 응용 트래픽 데이터 세트를 수집하였으며 수집한 응용의 종류는 표 1에 나타나있다.

표 1. 수집한 응용 리스트

Bithumb	Coineone	Upbit	Excel
Teams	Excel	PPT	Word
Onenote	WEB_Excel	WEB_PPT	WEB_Word
Daum	Gmail	Nate Mail	Navermail
Naver band	kakaotalk	Skype	Agoda
Airbnb	Goodchoice	Hotels.com	Netflix
Yanolja	11st	Coupang	Gmarket
Musinsa	Tmon	Wemakeprice	Facebook
Instagram	Kakaostory	Naverblog	Tistory
Twitter	Disney+	MelonMusic	NaverTV
Tving	Twitch	Wavve	NaverSeries
Youtube Music	Hotels Combined	Kakao webtoon	Naver Webtoon
Kakao page	Youtube		

B. 패킷 데이터 입력 방법

패킷 데이터 입력 방법은 단일 입력과 다중 입력으로 구분할 수 있으며 단일 입력은 분류 모델을 학습할 때 하나의 벡터만이 입력되는 경우를 말한다. 단일 입력은 플로우 내 첫 n bytes를 이어붙여 하나의 벡터로 만드는 것이 일반적이며 헤더 포함 여부를 결정할 수 있다. 우리는 헤더가 포함된 경우(ALL+L7)와 페이로드만 포함된 경우(L7)의 2가지로 나누어 서브 데이터 세트를 생성하였으며 이는 그림 2에 나타나있다.

다중 입력은 분류 모델을 학습할 때 여러 개의 데이터가 입력되는 경우를 말하며 모델은 플로우 내 첫 k개의 패킷 벡터를 동시에 입력받아 학습을 수행한다. [5]에 따르면 14개 이하의 패킷이 입력될 때 가장 좋은 성능

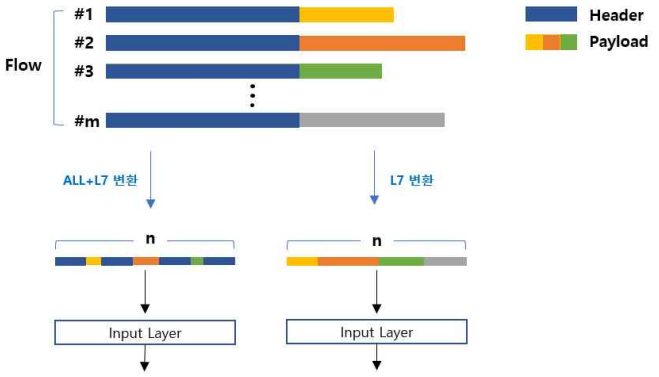

그림 2. 단일 입력 벡터 생성 방법

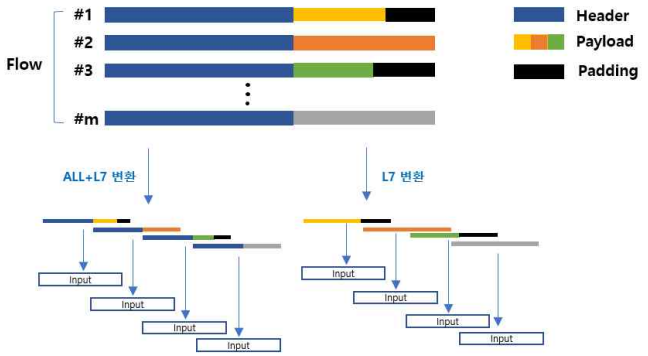

그림 3. 다중 입력 벡터 생성 방법

을 보이며 k에 따른 성능 변화 차이가 크지 않은 것으로 나타났다. 따라서, 우리는 k를 10으로 고정한다. 단일 입력과 마찬가지로 All+L7과 L7의 2가지 경우로 나누어 서브 데이터 세트를 생성하였으며 이는 그림 3에 나타나있다.

B. 입력 벡터 모양

기존 연구에서는 784*1(1D CNN), 28*28(2D CNN)의 2가지 모양이 주로 고려되었으며 이는 392*2, 196*4, 112*7 등의 다른 모양을 가진 입력으로 변환할 수 있다. 이때, 생성할 수 있는 모양의 개수는 입력 벡터의 크기의 약수의 개수와 같으며 784의 경우 15개의 모양을 생성할 수 있지만 마지막 모양은 1*784로서 첫 번째 모양과 동일하게 학습이 수행되므로 마지막 모양은 제외한다.

C. 분류 모델

분류 모델은 LeNet 구조와 유사하게 컨볼루션 레이어 2개와 풀링 레이어 2개로 이루어져 있고 마지막 풀링 레이어 후 2개의 덴스 레이어를 통해 분류 결과를 출력한다. 각 입력 벡터의 모양에 따라 모델의 하이퍼 파라미터를 최적화하기 위하여 Keras Tuner 중 Hyperband를 사용하여 각 모양에 적합한 모델 구조를 결정하였다. 모델 구조를 최적화하는 과정에서는 컨볼루션 레이어와 풀링 레이어의 필터 사이즈와 풀 사이즈만 변경하며 최적화를 수행했으며 고정한 하이퍼 파라미터는 표 2와 같다.

표 2. 모델 최적화 과정에서 고정한 하이퍼 파라미터

고정 하이퍼 파라미터	
컨볼루션 필터 개수	16
덴스 레이어 유닛 수	12

Ⅳ. 실험 결과

본 장에서는 최적의 하이퍼 파라미터가 적용된 분류 모델을 앞서 생성한 서브 데이터 세트에 적용하고 분류 결과를 비교한다.

표3은 L7 데이터 세트의 입력 모형별 최적화된 하이퍼 파라미터를 나타낸다. 주로 추출된 특징 맵을 요약하여 크기를 줄이는데 활용되는 풀링 레이어의 크기가 모두 (1, 1)로 최적화된 것은 패킷의 페이로드 데이터가 2차원 공간 정보가 아닌 1차원 공간 정보를 담고 있는 데이터인 것을 의미한다. 따라서, 2차원 공간 정보 기반 풀링 레이어를 적용하지 않는 것이 적합하다고 판단할 수 있다. 그림 4에서는 L7 데이터 세트의 모형별 분류 모델의 성능을 나타내며 일반적으로 사용했었던 정사각 모형(28*28)은 다른 모형들과 비교하여 높지 않은 것을 확인할 수 있다. 모형 중 패킷의 1차원적 특성을 고려한 392*2, 2*392, 112*7 모형이 가장 높은 성능을 보였다. 표4는 ALL 데이터 세트의 입력 모형별 최적화된 모델 구조를 나타내며 각 모형 별 분류 정확도는 그림 5에 나타나있다. L7 데이터 세트의 결과에서는 정사각 모형(28*28)이 평균적인 정확도를 나타냈으나 ALL 데이터 세트에서는 정사각 모형이 가장 낮은 정확도를 나타냈다. 또한, L7 데이터 세트와는 달리 ALL 데이터 세트에서 제일 높은 정확도를 나타냈던 모형의 경우 풀 사이즈를 (1, 2)를 적용하고 있으며 이는 패킷의 전체 데이터에서는 2차원 공간 정보가 활용되어야 함을 나타낸다. 표 5에서는 데이터 세트와 단일/다중 입력 여부 그리고 단일/다중 모형 여부에 따라 응용 트래픽을 분류한 결과를 나타낸다. L7 데이터 세트와 ALL 데이터 세트를 비교한 결과 ALL 데이터 세트를 적용하는 것이 평균 정확도 측면에서 23% 더 좋은 것을 알 수 있으며 다중입력, 다중모형이 적용된 분류 모델을 사용한 경우에는 27% 정도의 정확도 상승이 있었다. 1D CNN 기

표 3. 입력 모양별 최적화된 모델 구조(L7)

데이터셋	모형	필터 사이즈	풀 사이즈
L7	(784,1)	(43,1)	(1,1)
	(392,2)	(85,1)	(1,1)
	(196,4)	(11,1)	(1,1)
	(112,7)	(13,1)	(1,1)
	(98,8)	(8, 2)	(1,1)
	(56,14)	(7, 3)	(1,1)
	(49,16)	(4, 2)	(1,1)
	(28,28)	(5, 5)	(1,1)
	(16,49)	(2, 4)	(1,1)
	(14,56)	(1, 2)	(1,1)
	(8,98)	(1, 4)	(1,1)
	(7,112)	(1, 5)	(1,1)
	(4,196)	(1, 3)	(1,1)
	(2,392)	(1,43)	(1,1)

표 4. 입력 모양별 최적화된 모델 구조(All)

데이터셋	모양	필터 사이즈	풀 사이즈
ALL	(784,1)	(5, 1)	(1, 1)
	(392,2)	(3, 1)	(1, 1)
	(196,4)	(5, 1)	(1, 1)
	(112,7)	(2, 1)	(1, 1)
	(98,8)	(2, 1)	(1, 1)
	(56,14)	(2, 3)	(1, 2)
	(49,16)	(2, 2)	(1, 2)
	(28,28)	(2, 4)	(2, 1)
	(16,49)	(1, 2)	(1, 2)
	(14,56)	(2, 2)	(1, 2)
	(8,98)	(1, 7)	(1, 1)
	(7,112)	(1, 5)	(1, 1)
	(4,196)	(1, 7)	(1, 1)
	(2,392)	(1, 9)	(1, 1)

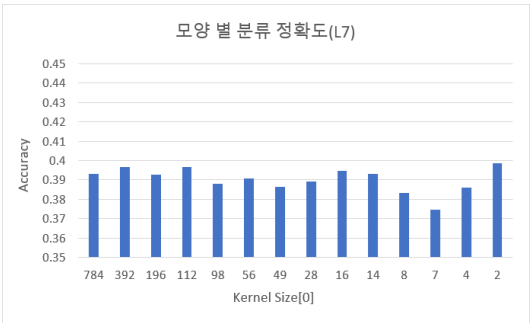

그림 4. 모양별 단일 입력 벡터에 대한 분류 정확도(L7)

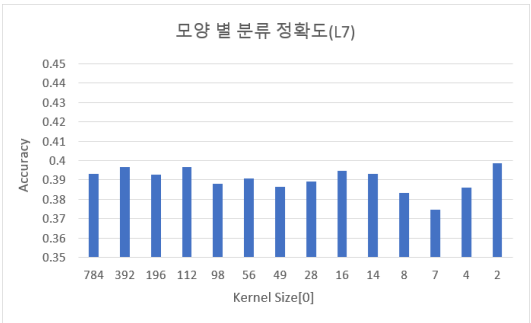

그림 5. 모양별 단일 입력 벡터에 대한 분류 정확도(All)

표 5. 단일/다중 입력, 단일/다중 모형 여부에 따른 분류 정확도

데이터 세트	모델	정확도
L7	단일 입력, 784*1 모형(1D CNN)	0.3933
	단일 입력, 28*28 모형(2D CNN)	0.3893
	단일 입력, 다중(784*1, 28*28) 모형	0.4146
	단일 입력, 다중 모형	0.4160
	다중 입력, 784*1 모형	0.4624
	다중 입력, 28*28 모형	0.4769
	다중 입력, 다중(784*1, 28*28) 모형	0.4792
	다중 입력, 다중 모형	0.4831
ALL	단일 입력, 784*1 모형	0.6047
	단일 입력, 28*28 모형	0.5820
	단일 입력, 다중(784*1, 28*28) 모형	0.6320
	단일 입력, 다중 모형	0.6567
	다중 입력, 784*1 모형	0.7366
	다중 입력, 28*28 모형	0.7158
	다중 입력, 다중(784*1, 28*28) 모형	0.7412
	다중 입력, 다중 모형	0.7526

반 모형(784*1)은 2D CNN 기반 정사각 모형(28*28)보다 모든 경우에서 0.5~2 정확도 상승이 있었으며 [4]에서 제안한 방법과 유사하게 1D CNN 분류 결과와 2D CNN 분류 결과를 앙상블한 분류모데릐 경우 1D CNN 모델보다 평균적으로 2% 정확도 상승이 있었다. 본 논문에서 제안했던 다중 모형별 분류 결과를 앙상블 하는 방법의 경우 L7 데이터 세트에서는 단일 입력일 경우와 다중 입력일 경우 모두 정확도가 소폭 상승했으며 ALL 데이터 세트를 활용한 실험 결과에서는 단일 입력일 때 2.5% 상승했으며 다중 입력일 때는 1.1%가 상승하였고 생성한 모든 분류 모델 중 가장 높은 정확도를 나타냈다.

Ⅳ. 실험 결과

본 논문은 패킷 입력 벡터를 다양한 모형으로 변환하고 각 모형 별 분류 결과를 앙상블하여 응용 트래픽을 분류하는 방법을 제안했다. 우리는 패킷 데이터가 2차원 이미지로 변환되는 과정에서 원래 가지고 있던 1차원 정보가 손실되어 정확히 학습이 되지 않는다는 가정 아래 패킷 데이터를 다양한 모형으로 변환하고 이를 앙상블하는 방법을 적용하여 패킷 데이터 변환 과정에서 발생하는 정보 손실을 최소화하고자 했다. 결과적으로, 제안한 방법은 모든 실험 결과에서 가장 높은 정확도를 나타내었으며 이는 응용 분류 분야뿐만 아니라 공격 탐지 분야에서도 효과적으로 활용할 수 있을것이라고 기대한다. 또한, 우리는 패킷 입력 벡터를 다양한 모형으로 변환하여 개별 분류 모델을 생성하여 분류 결과를 비교하였으며 그 결과 단순히 1차원적 공간정보를 활용하는 입력 벡터 모형(784*1)이나 기존에 주로 활용되었던 정사각 모형(28*28)보다 패킷을 더 잘 표현할 수 있는 모형이 있다는 것을 실험 결과를 통해 증명하였다. 하지만 다양하고 많은 응용 트래픽이 발생하는 환경에서 실용적으로 사용하기에는 정확도가 부족하다는 한계점이 있으며 우리가 변환한 다양한 모형들은 기존 컴퓨터 비전 분야 모델의 정사각 모형과 상이하기 때문에 쉽게 첨단 연구를 적용시키기 어렵다는 점도 큰 한계점으로 작용한다. 따라서 우리는 향후 연구로 기존 CNN 기반의 첨단 분류 모델을 응용 트래픽 분류 분야에 적합하게 변형하여 적용하는 방법에 대해 연구를 수행할 계획이며 본 논문에서 고려하지 않았던 응용 트래픽 플로우 내 패킷들의 시계열 특성을 고려한 분류 모델에 대해서 연구할 계획이다.

참 고 문 헌

[1] L. Xu, X. Zhou, Y. Ren, and Y. Qin, “A Traffic Classification Method Based on Packet Transport Layer Payload by Ensemble Learning,” in 2019 IEEE Symposium on Computers and Communications (ISCC), Jun. 2019, pp. 1–6. doi: 10.1109/ISCC47284.2019.8969702.

[2] S.-H. Ji, U.-J. Baek, M.-G. Shin, B.-M. Chae, H.-W. Moon, and M.-S. Kim, “Design of Web Application Traffic Classification Model Based on Convolution Neural Network,” kics, vol. 44, no. 6, pp. 1113–1120, Jun. 2019, doi: 10.7840/kics.2019.44.6.1113.

[3] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, “End-to-end encrypted traffic classification with one-dimensional convolution neural networks,” in 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Jul. 2017, pp. 43–48. doi: 10.1109/ISI.2017.8004872.

[4] W. Wang et al., “HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection,” IEEE Access, vol. 6, pp. 1792–1806, 2018, doi: 10.1109/ACCESS.2017.2780250.

[5] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, “Malware traffic classification using convolutional neural network for representation learning,” in 2017 International Conference on Information Networking (ICOIN), Jan. 2017, pp. 712–717. doi: 10.1109/ICOIN.2017.7899588.