GASF 이미지 인코딩 CNN 기반 데이터 효율적 봇넷 탐지: MLP·FT-Transformer와의 동일 조건 비교*

김도현°[1]† · 하상우[2]† · 안균승[3] · 조성제[3]
[1]단국대학교 인공지능융합학과, [2]단국대학교 컴퓨터학과, [3]단국대학교 소프트웨어학과
{dpotlight2, sangu1115, staran1227, sjcho}@dankook.ac.kr

Data-Efficient Botnet Detection Using a GASF Image-Encoded CNN: A Controlled Comparison with MLP and FT-Transformer

DoHyeon Kim°[1]† · SangWoo Ha[2]† · GyunSeung Ahn[3] · SeongJe Cho[3]
[1]Dept. of AI-based Convergence, [2]Dept. of Computer Science and Engineering,
[3]Dept. of Software Science, Dankook University
{dpotlight2, sangu1115, staran1227, sjcho}@dankook.ac.kr

요 약

네트워크 트래픽 기반 봇넷 탐지에서는 정상 트래픽으로 위장한 악성 행위를 제한된 학습 데이터만으로 안정적으로 식별하는 것이 중요하다. 본 논문은 CSE-CIC-IDS2018 데이터셋의 봇넷 트래픽을 대상으로, 동일한 25개 플로우 통계 피처를 사용하는 MLP, FT-Transformer, GASF 이미지 기반 CNN을 유사한 파라미터 규모에서 비교하고, 학습 데이터 비율을 5%에서 100%까지 단계적으로 축소하는 반복 실험을 통해 입력 표현 방식이 탐지 성능과 데이터 효율성에 미치는 영향을 분석하였다. 전체 데이터 조건에서는 세 모델 모두 F1 99.93% 이상의 높은 탐지 성능을 보였으며, MLP와 CNN-GASF가 F1 99.98%로 공동 최고를 기록하였다. 학습 데이터를 축소한 조건에서는 모델 간 차이가 두드러졌는데, MLP는 5% 조건에서 F1이 87.96%로 급격히 하락한 반면, CNN-GASF는 저장 용량은 증가하지만 동일 조건에서 99.87%를 유지하여, 적은 학습 샘플로도 높은 F1을 유지하는 샘플 효율성을 보였다.

1. 서 론

네트워크 보안 환경에서는 정상 트래픽으로 위장한 악성 행위를 조기에 탐지하는 기술이 중요하다[1]. 특히 봇넷(botnet)은 감염 호스트를 원격으로 제어하여 분산 서비스 거부, 정보 유출, 스팸 전송 등 다양한 공격에 활용될 수 있으므로 침입 탐지 시스템의 주요 탐지 대상이다. 그러나 실제 네트워크 트래픽은 클래스 불균형이 크고[2], 플로우 통계 피처(feature) 간 비선형 상호작용이 복잡하여 단순 규칙 기반 탐지나 얕은 분류 모델만으로는 안정적인 탐지가 어렵다[3].

딥러닝 기반 침입 탐지 연구에서는 플로우 피처를 1차원 수치 벡터로 입력하는 Tabular 기반 접근과 피처 간 관계를 이미지 형태로 변환하여 CNN으로 학습하는 이미지 기반 접근이 함께 연구됐다. 그러나 기존 접근법들은 대부분 서로 다른 실험 환경에서 개별적으로 평가되어 입력 표현 방식 자체가 성능에 미치는 영향을 공정하게 비교하기 어려웠다. 또한 실제 운영 환경을 고려했을 때, 전체 데이터 조건의 최종 성능뿐 아니라 제한된 학습 데이터 조건에서의 성능 안정성과 모델별 데이터 효율성을 동일 조건에서 정량적으로 비교 분석할 필요가 있다.

본 논문은 CSE-CIC-IDS2018 데이터셋[4]의 봇넷 트래픽을 대상으로 동일한 25개 플로우 피처를 사용하는 MLP, FT-Transformer, GASF 이미지 기반 CNN을 유사한 파라미터 규모로 비교한다. CSE-CIC-IDS2018은 CSE와 CIC가 공동 구축한 침입 탐지 데이터셋으로, Botnet을 포함한 여러 공격 시나리오를 제공한다. 본 연구의 주요 기여는 다음과 같다. 첫째, 동일 피처와 동일 데이터 분할 조건에서 1차원 수치 표현, 토큰 기반 어텐션 표현, 2차원 GASF 이미지 표현을 비교한다. 둘째, 모델 파라미터 규모를 유사하게 설정하여 입력 표현 방식의 영향을 중심으로 분석한다. 셋째, 학습 데이터 비율을 단계적으로 축소하여 GASF 기반 CNN의 학습 데이터 효율성(샘플 효율성)을 정량적으로 평가한다.

2. 관련 연구

딥러닝 기반 침입 탐지 연구는 네트워크 플로우를 수치형 테이블 데이터로 처리하는 방식과, 피처 또는 시간적 관계를 2차원 표현으로 변환하여 CNN을 적용하는 방식으로 발전해 왔다[5]. 기존 연구들은 주로 트래픽을 Tabular 데이터로 보고 MLP나 트리 기반 모델을 사용하는 방식[6]과 시계열 또는 상관관계 구조를 이미지로 변환한 뒤 합성곱 신경망을 적용하는 방식[7, 8]으로 나뉜다. 실제로 이미지 기반 침입 탐지 연구에서 Terzi(2022)는 GAF 변환과 CNN을 결합하여 CSE-CIC-IDS2017 데이터셋에서 공격 트래픽에 대해 F1 98.72%를 달성하였으며[5], Yan et al.(2024)은 MTF 변환과 앙상블 CNN을 CSE-CIC-IDS2018 일부 시나리오에 적용하여 F1 99% 이상을 보고하였다[11].

Tabular 모델은 입력 해석이 상대적으로 직관적이지만 피처 간 복잡한 비선형 관계를 충분히 반영하지 못할 수 있으며[9], 이미지 기반 모델은 지역 패턴과 상호관계를 더 잘 포착할 수 있으나 데이터 변환 과정에서 표현 손실의 영향을 받을 수 있다[10]. 이처럼 두 방식은 뚜렷한 기술적 장단점을 지님에도 불구하고, 선행 연구들의 수치는 서로 다른 피처 구성, 전처리 방식, 모델 규모 하에서 산출되어 입력 표현 방식 자체의 우열을 객관적으로 가리기 어려웠다. 특히, GAF·MTF 등 이미지 인코딩 기반 CNN

* 본 연구는 2026년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업 지원을 받아 수행되었음(2024-0-00035). 또한 본 연구는 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행되었음(IITP-2026-RS-2023-00259967).
† 공동 1저자임.

구조가 침입 탐지 분야에서 유효성이 확인된 바 있음에도 불구하고[5, 11], 본 데이터셋 환경에서 타 모델들과 동일한 조건으로 대조된 연구가 부재한 실정이다. 본 논문은 동일 피처, 유사 모델 규모, 동일 분할 조건에서 1차원 수치 벡터, 토큰 기반 어텐션, 2차원 GASF 이미지의 세 가지 표현 방식을 직접 대조함으로써 기존 연구의 파편화된 비교 한계를 극복한다는 점에서 차별화된다.

3. 제안 방법 및 실험 설정

3.1 데이터셋 및 전처리

본 논문은 CSE-CIC-IDS2018 데이터셋의 트래픽 캡처 파일 중, 봇넷 공격(Zeus, Ares C&C)의 플로우 메타데이터가 포함된 유일한 파일인 Friday-02-03-2018을 실험 대상으로 사용하였다. 해당 파일은 1,048,575개 플로우 레코드와 80개의 CICFlowMeter 기반 피처로 구성된다. 데이터 정제는 Timestamp 제거, 라벨 이진 인코딩, 결측치의 0 대체, 분산 0 피처 제거, 고상관 피처 ($|r| \geq 0.95$) 제거 등의 단계로 수행하여 44개의 특징을 선별했다.

피처 선택은 Random Forest 피처 중요도와 Mutual Information 점수를 결합한 방식으로 수행하였다. 학습 분할에서 계층적으로 추출한 200,000행에 한해 두 점수를 산출하여 데이터 누수를 방지하였으며, 각 점수를 자체 최댓값으로 정규화한 뒤 6:4 비율로 가중 합산하였다. 점수가 동일한 경우 봇넷의 C&C 주기 통신 특성을 반영하여 IAT → Active/Idle → Flow 통계 → 기타 순의 도메인 우선순위를 적용하여 최종 25개 피처를 선정하였다. 최종 선정된 25개 피처와 라벨은 1,048,575개 샘플로 구성된 약 159 MB 크기의 CSV 파일로 저장되었으며, 이는 4.2절의 학습 데이터 효율성 비교의 기준이 된다.

3.2 데이터 분할 및 학습 설정

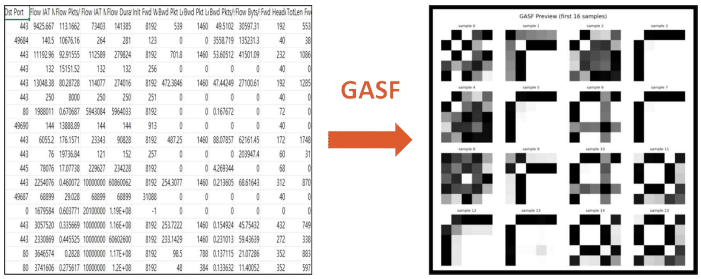
정제된 1,048,575개 샘플은 6:2:2 비율로 학습/검증/테스트 세트로 분할하여 각 분할에서 Bot/Benign 비율(27.29 : 72.71%)이 원본과 동일하게 유지되도록 하였다. 학습 세트에는 모델이 다수 클래스에 편향되지 않도록 Bot과 Benign을 1:1로 맞추는 언더샘플링을 적용하였고, 검증·테스트 세트는 원본 분포를 그대로 유지하여 실제 운영 환경의 클래스 분포에서 평가하였다. 이는 실제 네트워크 환경에서 정상 트래픽이 다수를 차지한다는 점을 반영한 것으로, 원본 분포 유지를 통해 모델 성능의 과대평가를 방지하였다.

공통 학습 설정으로는 클래스 불균형 환경에서 신뢰성 있는 지표인 검증 F1을 기준으로 한 Early Stopping(patience=5)를 적용하였으며, 손실 함수로는 BCEWithLogitsLoss를 사용하였다. 모델별 학습 설정은 사전 검증을 통해 각 구조에 적합한 값으로 결정하였으며, MLP와 CNN은 Adam 옵티마이저(MLP: 5×10^{-4}, CNN: 1×10^{-3})를, FT-Transformer는 AdamW 옵티마이저(학습률 3×10^{-4}, 가중치 감쇠 1×10^{-4})를 사용하였다.

3.3 GASF 이미지 변환

GASF(Gramian Angular Summation Field)는 본래 1차원 시계열 데이터의 시점 간 관계를 2차원 이미지로 인코딩하기 위해 제안된 기법이다. 본 연구에서 사용하는 25개 플로우 통계 피처는 시간 순서를 갖지 않으므로, GASF의 원래 시계열적 의미는 보존되지 않는다. 대신 본 연구에서는 피처 중요도 순으로 정렬된 입력 벡터에 GASF를 적용하여 피처 쌍 간의 결합 관계를 2차원 공간 패턴으로 표현하는 도구로 활용한다. 이를 통해 1차원 벡터에서는 직접 드러나지 않는 피처 간 결합 관계가 이미지의 공간 패턴으로 표현되어, CNN의 합성곱 연산으로 학습할 수 있게 된다. 이러한 적용의 유효성은 4장 실험 결과를 통해 사후 검증한다.

정규화 통계량은 데이터 누수를 방지하기 위해 학습 분할 샘플에서만 산출하였으며, 동일한 통계량을 검증·테스트 분할에 적용하여 GASF 이미지로 그림 1과 같이 변환하였다. 그 결과 1,048,575개 샘플이 (1,048,575, 25, 25, 1) 형상의 텐서로 변환되어 약 2.62 GB의 저장 공간을 차지하였다. 이는 1차원 CSV 데이터(약 159 MB) 대비 약 16.9배 큰 규모로, 4.2절의 학습 데이터 효율성 비교의 기준이 된다.



[그림 1]. GASF 이미지 변환 과정

3.4 비교 모델 구조

본 연구에서는 동일한 25개 피처에 대해 입력 표현 방식이 서로 다른 세 가지 딥러닝 모델을 비교한다. MLP(약 117만 파라미터)는 1차원 수치 벡터를 직접 입력받는 4층 신경망으로, 각 은닉층에 BatchNorm과 Dropout을 적용하여 과적합을 억제하였다. CNN(약 120만 파라미터)은 3.3절에서 변환된 25×25×1 GASF 이미지를 입력으로 받아 두 단계의 합성곱과 풀링을 거친 뒤 완전연결 계층으로 이진 분류를 수행하며, 피처 간 결합 관계를 공간적 패턴으로 학습한다. FT-Transformer(약 123만 파라미터)는 각 피처를 독립적인 토큰으로 임베딩하고 분류용 CLS 토큰을 추가한 뒤, 다층 Transformer 인코더의 Self-Attention을 통해 피처 간 상호작용을 직접 학습하는 테이블형 데이터 전용 구조이다. 세 모델은 입력 표현 방식만 달리하며, 모델 용량 차이에 따른 영향을 줄이기 위해 파라미터 수를 유사하게 설정하였다.

4. 실험 결과 및 분석

4.1 모델 성능 비교

표 1은 테스트 세트에 대해 서로 다른 시드 3회 반복 실험으로 얻은 평균을 나타낸다. 세 모델 모두 정확도 99.96% 이상, F1-Score 99.93% 이상을 기록하였다. ROC-AUC는 세 모델 모두 0.9999 이상으로 사실상 1에 수렴하였으며, 전체 약 21만 건 중 오분류는 MLP 약 21건, CNN 약 21건, FT-Transformer 약 73건에 불과하였다. 이처럼 단순 성능 지표만으로는 모델 간 우열을 가리기 어려우므로, 다음 절에서는 학습 데이터 양에 따른 성능 변화로 분석을 확장한다.

[표 1]. 테스트 세트 성능 비교

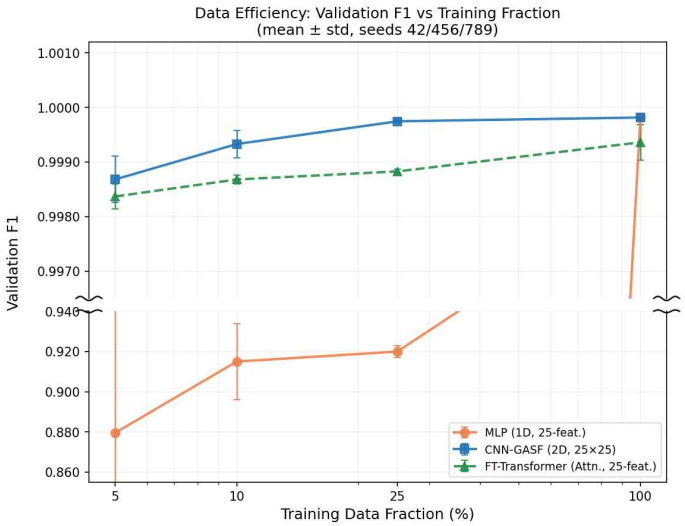
Metric \ Model	MLP	CNN-GASF	FT-Transformer
Accuracy (%)	99.990	99.990	99.965
Precision (%)	99.988	99.985	99.962
Recall (%)	99.974	99.978	99.911
F1-score (%)	99.981	99.982	99.936
오분류(건)	21	21	73

4.2 데이터 경량화 실험

학습 데이터 비율을 5%에서 100%까지 변화시키며 각 모델의 검증 F1과 학습 데이터 용량을 함께 측정하였다(표 2, 그림 2). 1차원 수치 데이터(1D CSV)로 학습하는 MLP, FT-Transformer는 샘플당 약 152바이트, 2차원 GASF 이미지로 학습하는 CNN은 샘플당 2,500바이트(25×25×4)를 차지한다.

[표 2]. 학습 데이터 비율별 검증 F1 및 학습 용량

	Training Data Ratio			
Category	5%	10%	25%	100%
1D 용량 (MB)	2.9	5.8	14.5	57.9
2D 용량 (MB)	46.6	93.2	233.0	931.6
MLP F1 (%)	87.961	91.503	91.999	99.981
CNN-GASF F1 (%)	99.868	99.933	99.975	99.982
FT-Transformer F1 (%)	99.837	99.868	99.883	99.936



[그림 2]. 학습 데이터 비율별 검증

F1 변화모델 간 데이터 효율성 차이는 5% 조건에서 두드러졌다. CNN-GASF는 균형 샘플링된 5% 학습 subset, 즉 약 15.7K건의 봇넷 공격 샘플만으로 F1 99.868%를 달성하였고, FT-Transformer 역시 99.837%를 유지한 반면, MLP는 동일 조건에서 F1이 87.961%로 급격히 하락하였다. 이는 1차원 수치 학습 방식이 학습 샘플 부족 환경에서 취약함을 보여주며, GASF 이미지 인코딩이 적은 샘플에서도 봇넷의 주기적 패턴을 안정적으로 학습함을 보여준다. 나아가 25% 학습 데이터 조건에서 CNN-GASF는 F1 99.975%를 달성하여, MLP가 전체 데이터로 학습했을 때의 99.981%와 사실상 동등한 결과를 약 1/4의 학습 샘플로 재현하였다. GASF 이미지 인코딩은 1차원 수치 데이터 대비 약 16.9배의 저장 용량을 요구하지만, 적은 학습 샘플로도 높은 F1을 안정적으로 유지하는 샘플 효율성을 보였다.

5. 결 론

본 논문은 GASF 이미지 인코딩이 레이블 확보가 어려운 네트워크 보안 환경에서 데이터 효율적 탐지 수단이 될 수 있음을 반복 실험을 통해 검증하였다. 전체 학습 데이터의 5%라는 동일 조건에서 CNN-GASF가 F1 99.868%를 달성한 반면 MLP는 87.961%에 그쳐, 이미지 기반 인코딩의 샘플 효율성이 두드러졌다. 본 연구에서 GASF 이미지는 1차원 수치 데이터 대비 약 16.9배의 저장 용량을 차지하였으나, 보다 정교한 피처 선택을 통해 25개보다 적은 수의 피처를 사용하면 GASF 이미지의 가로·세로 크기가 줄어들어, 샘플당 이미지 크기와 전체 저장 용량을 본 논문 대비 더욱 감소시킬 수 있을 것으로 기대된다. 다만 본 연구는 단일 데이터셋과 봇넷 단일 공격 유형에 한정되며, 세 모델 간 성능 차이가 근소한 만큼 다양한 데이터셋 및 모델과의 추가 비교 실험이 필요하다. 향후에는 본 방법론을 DDoS, 침투 공격 등 다양한 공격 유형으로 확장 적용함으로써, 새로운 공격이 등장하는 환경에서도 데이터 효율적인 탐지가 가능한지 검증할 계획이다.

참고문헌

[1] Kumar, A., et al. "Machine learning-based early detection of IoT botnets using network-edge traffic," Computers & Security, vol. 117, pp. 102693, 2022.

[2] 정윤경, 박기남, 김현주, 김종현, 현상원. "클래스 불균형 데이터에 적합한 기계 학습 기반 침입 탐지 시스템," 정보보호학회논문지, 제27권, 제6호, pp. 1385-1395, 2017.

[3] Allassak, N., Trichni, S., and Omary, F. "Federated Intrusion Detection Using TabTransformer-TCN-BiGRU-Attention: A High-Accuracy Hybrid Deep Learning Approach," Engineering, Technology & Applied Science Research, vol. 15, no. 6, pp. 29647–29654, 2025.

[4] Canadian Institute for Cybersecurity. "CSE-CIC-IDS2018 Dataset," University of New Brunswick, 2018.

[5] D. S. Terzi, "Gramian angular field transformation-based intrusion detection," Comput. Sci., vol. 23, no. 4, pp. 571–585, 2022.

[6] Sharafaldin, I., Lashkari, A. H., and Ghorbani, A. A. "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," **Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP)**, pp. 108-116, 2018.

[7] Wang, Z., and Oates, T. "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[8] Vandith Sreenivas, K., et al. "Classification of Arrhythmia in Time Series ECG Signals Using Image Encoding And Convolutional Neural Networks," 7th International Conference on Bio Signals, Images, and Instrumentation (ICBSII), pp. 1–6, 2021.

[9] Sandeep, S., Pratap, V., and Sravanthi, M. V. S. S. "Enhanced Intrusion Detection Using Stacked FT-Transformer Architecture," Journal of Cybersecurity and Information Management (JCIM), vol. 13, no. 2, pp. 16-26, 2023.

[10] Hatami, N., Gavet, S., and Chauvet, P. "Classification of Time-Series Images Using Deep Convolutional Neural Networks," Proceedings of the 10th International Conference on Computer Modeling and Simulation (ICCMS), pp. 242–246, 2018.

[11] Y. Yan, Y. Yang, S. Fang, M. Gao, and Y. Chen, "MUS Model: A Deep Learning-Based Architecture for IoT Intrusion Detection," Comput. Mater. Contin., vol. 80, no. 1, pp. 875–896, 2024.