

합성곱 신경망과 장단기 메모리를 이용한 사격음 분석 기법

Shooting sound analysis using convolutional neural networks and long short-term memory

강세혁,[†] 조지웅¹

(Se Hyeok Kang^{1†} and Ji Woong Cho¹)

¹육군 3사관학교 전자공학과

(Received February 16, 2022; accepted March 28, 2022)

초 록: 본 논문은 딥러닝기법 중 하나인 합성곱 신경망과 순환 신경망 중 하나인 장단기 메모리를 이용하여 사격시 발생하는 소음(이하 사격음)만으로 화기의 종류, 사격음 발생지점에 관한 정보(거리와 방향)를 추정하는 모델을 다루었다. 이를 위해 미국 법무부 산하 연구소의 지원하에 생성된 Gunshot Audio Forensic Dataset을 이용하였으며, 음향 신호를 멜 스펙트로그램(Mel-Spectrogram)으로 변환한 후, 4종의 합성곱 신경망과 1종의 장단기 메모리 레이어로 구성된 딥러닝 모델에 학습 및 검증 데이터로 제공하였다. 제안 모델의 성능을 확인하기 위해 합성곱 신경망으로만 구성된 대조 모델과 비교·분석하였으며, 제안 모델의 정확도가 90 % 이상으로 대조모델보다 우수한 성능을 보였다.

핵심용어: 사격음 분류, 합성곱 신경망, 장단기 메모리, 오디오 특징 추출, 오디오 전처리

ABSTRACT: This paper proposes a model which classifies the type of guns and information about sound source location using deep neural network. The proposed classification model is composed of convolutional neural networks (CNN) and long short-term memory (LSTM). For training and test the model, we use the Gunshot Audio Forensic Dataset generated by the project supported by the National Institute of Justice (NIJ). The acoustic signals are transformed to Mel-Spectrogram and they are provided as learning and test data for the proposed model. The model is compared with the control model consisting of convolutional neural networks only. The proposed model shows high accuracy more than 90 %.

Keywords: Gunshot classification, Convolutional Neural Network (CNN), Long Short-term Memory (LSTM), Audio feature extraction, Audio preprocessing

PACS numbers: 43.60.Bf, 43.60.Lq

1. 서 론

최근 사회는 다양한 영역에 걸쳐 초분산·초결합의 형태로 변모하고 있다. 이러한 추세에 발맞추어 군에서도 워리어 플랫폼^[1]과 같은 개인단위의 전투체계를 구현하기 위한 노력을 경주하고 있다. 개인단위의 전투체계를 운용하는 전투원들은 기본적인 임무 이외에도 많은 부수적 임무수행을 요구받을 것

이다. 이는 개인이 운용하는 장비의 종류 및 수가 증가하고, 임무수행에 대한 집중도가 떨어질 수 있음을 의미한다.

시대를 막론하고 개인 전투원에게 있어 중요한 임무 중 하나는 적 전투원에 관한 정보를 신속하고 정확하게 획득하는 것이다. 하지만 보통 상급부대에서 전체적으로 하달되는 적의 정보는 개략적인 정보일 가능성이 높기 때문에, 당장 내 눈앞에 있는 적에 대

[†]Corresponding author: Se Hyeok Kang (kangsehyeok0329@gmail.com)

Department of Electrical Engineering, Korea Army Academy, 495 Hoguk-ro, Gogyeong-myeon, Yeongcheon-si, Gyeongsangbuk-do 38900, Republic of Korea

(Tel: 82-54-330-4926)



Copyright©2022 The Acoustical Society of Korea. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

한 정보로서 활용하기에 제한되는 경우가 많다. 개인 전투원에게 있어 유의미한 적 정보는, 자신이 위치한 지점으로부터 어느 방향에 어느 거리만큼 떨어진 지점에 적이 존재하는지에 관한 정보이다. 개인 전투원이 주변의 적에 관한 정보를 획득하는 일반적인 방법은 적이 있을 것으로 예상되는 지점에 직접 이동하여 확인하는 것이다. 하지만 개인 전투원이 이동하는 동안 반대로 적에게 먼저 노출될 가능성이 있다.

따라서 본 연구팀은 개인 전투원이 직접 장비를 조작하지 않아도, 주변에서 들려오는 소리로부터 적의 유무와 위치(거리 및 방향)를 자동으로 분석·정보화하는 적 탐지센서가 미래의 개인단위 전투체계에 적용될 경우, 각 전투원이 다른 임무에 집중할 수 있는 여건을 보장하고 생존성을 증대시킬 수 있을 것이라 판단했다. 이에 본 논문에서는 적에 관한 다양한 소리 중 사격음을 분석하는 기법을 연구하였다.

사격음을 분석하는 연구는 다양한 기술을 적용하여 진행되어 왔다. 딥러닝이 대두되기 전까지 음성 인식, 객체 분류 등의 문제는 가우시안 혼합모델(Gaussian Mixture Model, GMM)^[23]과 히든 마르코프 모델(Hidden Markov Model, HMM)^[4,5] 등의 방법을 사용했다.

최근에는 딥러닝기술을 이용하여 음성이나 음향에서 객체를 분류하는 방향으로 많은 연구가 진행 중이다. 특히, Raponi *et al.*^[6]은 합성곱 신경망(Convolutional Neural Network, CNN)을 이용하여 사격음을 분석하여 구경, 화기 종류 등을 식별하는 모델을 제안하였다.

본 논문에서는 사격음을 분석하여 사격음 발생지점(방향과 거리)과 총기 종류를 식별하는 모델을 제안한다. 모델은 딥러닝기법 중 하나인 합성곱 신경망과 장단기 메모리(Long-Short Term Memory, LSTM)을 사용한다. 합성곱 신경망은 이미지와 같은 시각 데이터를 분석하는데 사용하는 인공신경망의 한 종류로, 컴퓨터 비전분야 등에서 주로 사용한다. 따라서, 음향 신호를 분석하기 위해서는 음향 데이터를 이미지와 같이 시각화 해주어야 한다. 이를 위해 음향 데이터를 멜 스펙트로그램으로 표현하여 사용한다.

장단기 메모리는 순환 신경망(Recurrent Neural Net-

work, RNN) 기법 중 하나로 시계열 데이터에 관하여 탁월한 성능을 보인다. 음향 데이터의 경우 데이터 간 시계열의 연속성을 가지고 있기 때문에 장단기 메모리를 활용한다면 효과적인 분석이 가능하다. 본 논문이 제안하는 모델은 멜 스펙트로그램 된 음향 데이터를 CNN과 LSTM이 결합된 신경망이 화기 종류(Pistol, Rifle), 사격 거리(10 m, 20 m, 40 m), 사격 방향(30°, 60°, 90°, 130°, 180°)을 식별한다.

본 논문의 구성은 다음과 같다. 2장에서는 사격음 분류와 관련하여 기존 연구에 대한 설명과 본 논문의 차이를 설명한다. 3장에서는 실험 설계에 관한 내용으로 제안 모델, 데이터셋, 실험환경에 대하여 설명한다. 4장에서는 실험 내용 및 결과를 정리하며, 5장에서는 결론과 기대효과를 제시한다.

II. 기존 연구

사격음을 분석에 관한 연구는 오래전부터 진행되어 왔다. 가우시안 혼합모델과 히든 마르코프 모델을 이용하여 화기 식별에 관한 연구가 진행되었는데, Morten Jr *et al.*^[4]는 음향 신호를 자기 회기(Auto Regressive, AR)를 갖는 히든 마르코프 모델로 포착된 사격음의 에너지와 스펙트럼 특성을 분석하여 화기 종류를 식별해 냈다. Djeddou *et al.*^[3]은 가우시안 혼합모델을 이용하여 두 종류의 화기(Simonov SKS, Kalashnikov AK)를 구분하는 모델을 제안하였다.

Raponi *et al.*^[6]은 합성곱 신경망(CNN)을 이용하여 사격음을 분석하여 화기의 종류, 구경, 화기의 분류(Pistol, Shotgun, Rifle)등을 수행하는 모델을 제안했다.

그러나, 앞서 언급한 References [2] ~ [5]의 연구는 실험환경 및 데이터와 일치한 조건에서만 성능을 보인다는 점에서 제한점이 있다. 또한, Reference [6]에서는 사격 지점 추정에 관한 연구는 실시되지 않았으며, 단순한 합성곱 신경망 구조로 시계열 데이터를 분석하는 데에는 한계가 있다. 이를 극복하기 위해서 본 논문에서는 다양한 화기와 조건으로부터 발생한 데이터를 활용하여 실험을 진행하였으며, CNN에 LSTM 레이어를 추가하여 시계열 데이터 분석에도 효과있는 모델을 구현하였다.

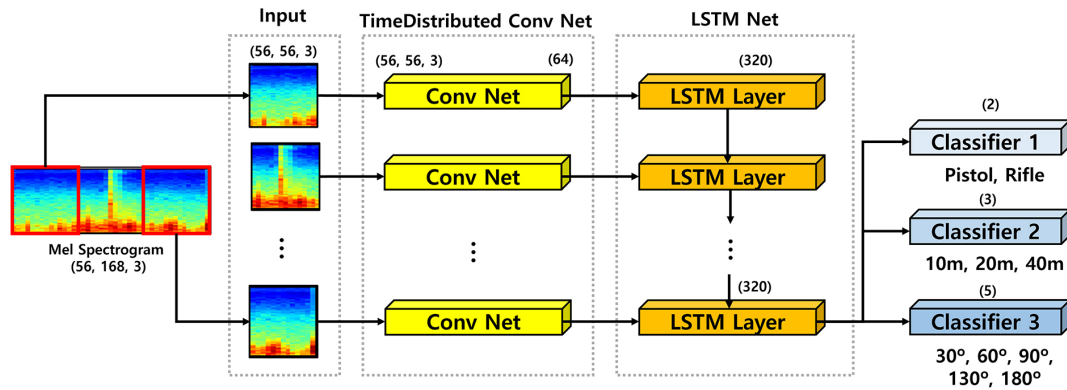


Fig. 1. (Color available online) The structure of proposed CNN + LSTM gunshot classifier model.

III. 실험 설계

3.1 제안 모델

본 논문에서 제안하는 모델의 구조는 Fig. 1과 같이 합성곱 신경망과 장단기 메모리가 결합된 구조이다. 입력으로 $[56 \times 168, 3]$ 크기의 멜 스펙트로그램 이미지를 입력 받는다. 데이터 전처리로 모델은 입력 이미지를 $[8, 56 \times 56, 3]$ 으로 샘플링한다.

CNN구조(Conv Net)는 시계열 데이터를 활용하기 위해 TimeDistributed 레이어로 구성된다. TimeDistributed 레이어는 각 스텝마다 오류를 계산해서 하위 스텝으로 오류를 전파하며, 신경망에 시계열 차원을 추가하여 시계열 분석이 가능하게 한다. CNN구조(Conv Net)의 세부 구조는 Fig. 2와 같다. Conv Net은 총 4개의 2차원 컨볼루션 레이어와 한 개의 전역 평균 풀링(Global Average Pooling, GAP) 레이어로 구성된다. 2차원 컨볼루션 레이어의 커널 크기는 $[3, 3]$ 으로, padding은 동일하게, strides는 Conv1에서만 1로, 나머지 Conv2 ~ 4에서는 2로 설정하였다. 각 레이어(Conv)에는 BatchNormalization을 실시후 활성화 함수로 LeakyReLU를 사용하였다(Fig. 3). 4번의 컨볼루션 레이어(Conv1 ~ 4)이후 전역 평균 풀링(GAP)을 사용하여 모델을 경량화시키고, 오버피팅을 방지하였다. GAP은 각 샘플들의 특징을 추출하기 위한 방법으로 평균 풀링을 반복하여 샘플의 크기를 축소시켜 샘플별로 $[1, 1]$ 크기의 필터크기의 대푯값을 추출한다. 따라서, Conv4에서 $[7 \times 7, 64]$ 사이즈의 샘플은 64개의 대푯값을 갖는 $[64]$ 크기의 레이어로 표현된다.

순환 신경망은 유닛간의 연결을 순환적인 구조로

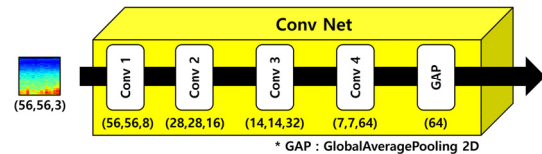


Fig. 2. (Color available online) The detailed structure of Conv Net.

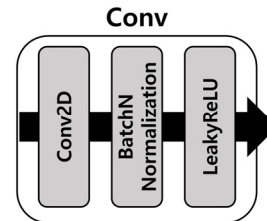


Fig. 3. The detailed structure of conv.

만들어 음성, 문자 등 시변적인 특징을 가진 데이터를 처리하는데 적합한 모델이다. 그러나, RNN의 경우 오차 역전파 과정에서 기울기가 소실되는 문제가 발생한다. 이를 극복하기 위해 제안된 방법 중 하나인 LSTM은 신경망에 망각 게이트를 추가하여 문제를 해결한다. 본 모델은 Conv Net을 통해 입력데이터의 특징이 추출된 샘플들의 시계열간 연관성을 파악하기 위해 LSTM을 이용한다. LSTM 레이어의 노드 수는 320개로 고정하였으며, dropout을 0.25로 설정하여 과적합을 방지하였다. 따라서, LSTM레이어는 입력되는 데이터의 순서에 따라 320개의 노드의 정보가 계승되고 320개의 노드의 정보는 업데이트되어 최종적으로 320개의 LSTM레이어 한 개로 합쳐진다. 320개의 LSTM레이어는 모델을 분류하기 위한 3개의 분류기의 신경망을 구성하는데 사용된다.

Table 1. Making up layers of the control model.

Layer Name	Scales
Input	[56×56, 3]
Conv 1	[56×56, 8]
Conv 2	[28×28, 16]
Conv 3	[14×14, 32]
Conv 4	[7×7, 64]
Dense(Flatten)	[3136]
Classifier 1 / 2 / 3	[2] / [3] / [5]

※ Each Conv has a BatchNormalization and a LeakyRELU activation function.

분류기는 총 3개로 구성된다. 3개의 분류기 모두 활성화 함수로 소프트맥스를 사용하여 분류를 수행한다. 분류기1(Classifier1)은 사격음에서 화기의 종류(Rifle, Pistol)을, 분류기2(Classifier 2)는 관측지점으로부터 사격지점까지 떨어진 거리(10 m, 20 m, 40 m)를 분류기3(Classifier 3)은 관측지점과 사격 지점 사이의 각도(30°, 60°, 90°, 130°, 180°)를 분류한다.

3.2 비교 모델(대조군)

본 논문에서 제안한 모델과 성능을 비교하기 위해 기존 연구에서 제시된 모델을 비교모델로 설정하여 성능을 비교한다. 비교 모델은 Reference [6]에서 제시한 CNN을 이용한 분류기 모델을 기본으로 하여 실험 목적에 맞게 일부 변형하였다. 비교 모델은 LSTM을 사용하지 않고 CNN만으로 구성되었으며, 세부적인 구조는 Table 1과 같다.

3.3 데이터 셋

본 실험에서 사용한 데이터 셋은 미국 법무부 산하 연구소(National Institute of Justice, NIJ)에서 지원한 실험에서 생성한 Gunshot Audio Forensics Dataset을 사용하였다.¹⁾ 데이터는 2017년 여름 애리조나에서 생성되었으며, 18종의 화기의 사격음을 20여 곳의 위치에서 4개의 서로 다른 녹음기로 녹음하였다. 사용된 18여 종의 총기는 Table 2와 같다. 본 논문에서는 Revolver와 Pistol은 권총으로, Rifle과 Carbine은 소

Table 2. Dataset : Gun model and category.

Gun Model	Category
High Standard Sport King	Pistol
S&W 34-1	Revolver
Ruger 10/22	Carbine
Remington 33 Bolt-Action	Rifle
Lorcin L380	Pistol
S&W 10-8	Revolver
Ruger Blackhawk	Revolver
Glock 19	Pistol
Sig P225	Pistol
M&P 40	Pistol
HK USP Compact	Pistol
Glock 21	Pistol
Colt 1911	Pistol
Kimber Tactical Custom	Pistol
M16A1 AR15	Rifle
WASR 10/63 AK47	Carbine
Winchester M14	Rifle
Remington 700	Rifle

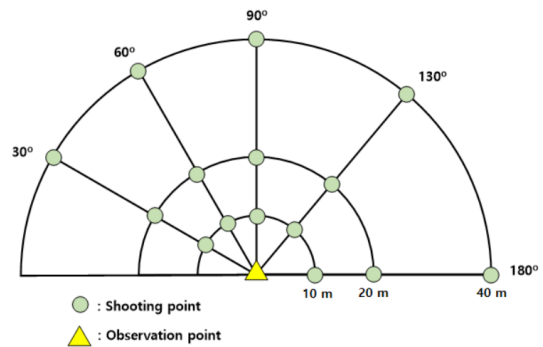


Fig. 4. (Color available online) Geometries of shooting points and an observation point.

총으로 분류하였다.

또한, 데이터를 관측지점을 기준으로 재구성하여 사격 지점을 관측지점으로부터의 거리(m)와 방향(°)으로 표현하였다. Fig. 4는 관측지점을 기준으로 표현한 사격 지점의 위치와 방향을 나타낸다. 거리는 10 m, 20 m, 40 m이며, 방향은 30°, 60°, 90°, 130°, 180°로 총 15개의 지점이다.

학습 및 테스트에 사용한 데이터 샘플은 1486개, 각 샘플의 시간은 1 s이며, 단일 채널 및 44 Hz, 100 Hz의 샘플링 레이트로 변환하였다.

1) Data collection was supported by NIJ grant 2016-DN-BX-0183 “Development of Computational Methods for the Audio Analysis of Gunshots.”

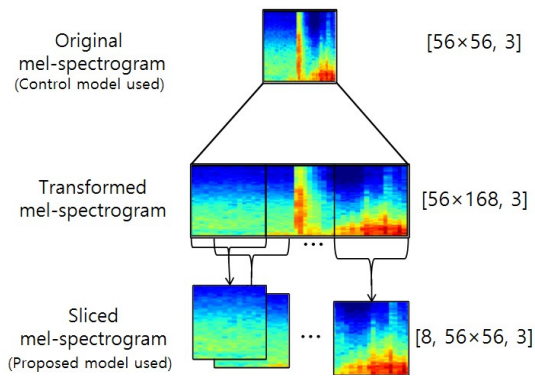


Fig. 5. (Color available online) The Pre-processing process of mel-spectrogram data for use in the proposed model.

CNN기반 제안 모델에 음향 신호를 사용하기 위해 서 데이터 샘플을 Short-time Fourier Transform(STFT) 한다. STFT와 멜 스펙트로그램은 python 오픈 소스 라이브러리 중 하나인 librosa^[7]를 이용하여 구현하였다. STFT시 프레임의 길이는 21 ms로 설정하였다. 프레임의 길이가 길면, 주파수 해상도가 높아지는 반면, 시간 해상도가 낮아진다. 일반적으로 음향 신호 처리에 20 ms ~ 30 ms의 해밍 윈도우를 사용하는 것을 고려하여 시간 해상도를 높이기 위하여 21 ms로 프레임 길이를 설정하였고, 프레임의 50%를 오버랩하여 시간 해상도를 높였다.

STFT를 토대로 멜 스펙트로그램을 추출하여 [56x168, 3] 크기의 이미지 파일로 변환한다. 대조 모델은 동일한 멜 스펙트로그램을 [56x56, 3] 크기의 이미지를 사용하였다.

[56x56, 3]의 멜 스펙트로그램은 [56x168, 3]로 리샘플링되는데 이때, y축의 값은 최단 인접 보간을 사용하여 값을 확장 시켰다. 필터링에 의해 입력값이 변조되는 것을 막기 위해 필터링은 추가하지 않았다.

[56x156, 3] 크기의 멜 스펙트로그램은 [56x56, 3] 크기를 갖는 8개의 프레임으로 샘플링되어 본 논문에서 제안하는 모델에서 사용되었다. 8개의 프레임으로 샘플링 시에는 [56x168, 3]의 이미지를 한 프레임당 윈도우 크기를 56으로 설정하여, [56x56, 3]크기로 잘랐으며, 홉 길이를 16으로 설정, 겹치는 구간이 40이 된다. LSTM 모델에 연속적인 데이터가 들어갈 때 전 후 데이터의 연속성을 모델이 분석하기 용이하게 하기 위하여 겹치는 구간의 길이를 길게 설정하였다.

Fig. 5는 멜 스펙트로그램 된 데이터를 제안 모델에서 사용하기 적합하게 전처리 하는 과정을 나타낸다.

입력데이터로 사용되는 멜 스펙트로그램의 경우 단일 채널(gray)로도 표현이 가능하지만, 본 실험에서는 3채널(RGB)의 스펙트로그램을 사용하였다. 멜 스펙트로그램의 x축은 시간, y축은 주파수이며, 색은 데시벨(dB)로 적색에 가까울수록 높은 값을, 청색에 가까울수록 낮은 값을 갖으며, 3채널을 사용한 이유는 딥러닝 모델이 단일채널의 데이터 보다 많은 데이터를 분석에 사용할 수 있을것이라 판단하였기 때문이다.

3.4 실험 환경 및 검증 방법

제안 모델의 훈련과 검증을 위해 데이터는 약 90% (1336 samples)의 학습 데이터와 약 10% (150 samples)의 테스트 데이터로 구분하였다. 학습데이터 중 10%는 검증용으로 사용하였다. 학습은 10폴드 교차 검증을 실시하였으며, 10개의 모델 중 정확도가 가장 높은과 낮은 모델을 제외한 8개 모델의 평균값을 구하였다.

학습시 최적화 함수는 Adam을 사용하였으며, 학습률은 0.001로, 미니 배치 크기는 32로 설정하였다. 에포크는 500으로 설정하였으나, 검증 손실이 75 epochs 동안 감소하지 않는다면 학습을 종료하는 조기 종료를 적용하였다. 제안 모델과 비교 모델의 손실함수는 분류기 1,2,3 모두 softmax 이후 Categorical Cross Entropy를 사용하였다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 - Score = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

* TP: True Positive, FP: False Positive, FN: False Negative, TN: True Negative

모델의 검증은 정확도와 F1-score를 함께 비교한다. F1-score는 정밀도, 재현도를 함께 비교한다. 정확

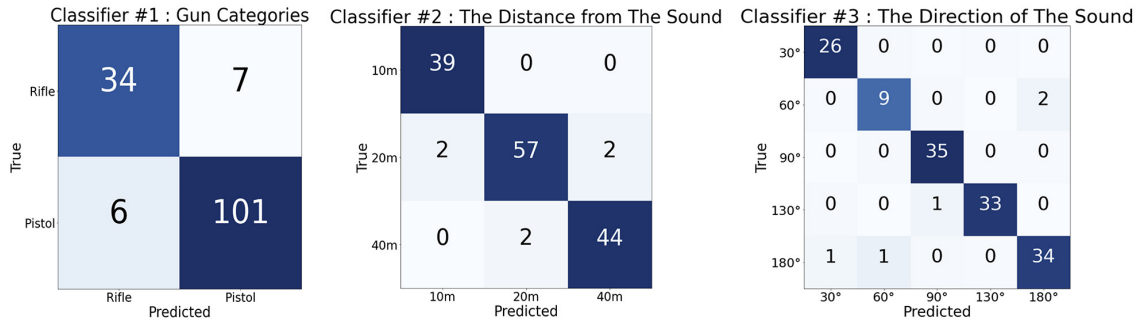


Fig. 6. (Color available online) Confusion matrices for each classifier of the proposed model.

Table 3. Comparison of the performance of the control model and the proposed model.

Category	Proposed model (LSTM+CNN)			Control model (CNN)		
	Acc (F1-score)	Precision	Recall	Acc (F1-score)	Precision	Recall
Rifle	90.3 %	83.7 %	81.3 %	87.3 %	80.0 %	72.6 %
Pistol	(87.9 %)	92.8 %	93.9 %	(83.7 %)	89.7 %	92.9 %
10m		94.9 %	96.7 %		81.5 %	82.0 %
20m	94.8 %	95.2 %	93.0 %	78.3 %	76.4 %	72.0 %
40m	(94.7 %)	93.7 %	95.2 %	(77.8 %)	76.4 %	81.9 %
30°		93.5 %	96.0 %		79.3 %	71.9 %
60°		85.7 %	75.0 %		89.5 %	53.1 %
90°	93.0 %	91.2 %	97.2 %	75.6 %	69.6 %	80.9 %
130°	(91.3 %)	96.8 %	93.8 %	(74.1 %)	75.4 %	71.2 %
180°		93.0 %	91.8 %		77.2 %	84.5 %

도는 총 테스트 데이터 중에서 정확하게 추정한 결과의 비로 정의되며, 정밀도는 모델이 정답이라 분류한 것 중 실제값이 정답인 비율이며, 재현도는 실제값이 정답인 것 중 모델이 정답이라 분류한 비율을 의미한다. 정확도, 정밀도, 재현도, F1-Score를 구하는 공식은 Eqs. (1)~(3)와 같다.

IV. 실험 결과

Table 3은 제안 모델과 대조 모델의 분류기 1,2,3의 정확도, 정밀도, 재현도, F1-score를 나타낸다. 제안 모델은 대조 모델에 비해 전반적으로 성능이 향상된 것을 알 수 있다. 분류 1의 경우 대조 모델과 비교했을 때 제안 모델은 정확도와 F1-score 모두 약 3% 증가하였다.

분류기 2와 3의 정확도는 각각 94.7%, 94.6%로 CNN만 사용하는 대조 모델과 비교했을 때 약 16% 이상 비약적으로 증가한 것을 알 수 있다. F1-score의 경우에도 90% 이상의 값을 나타냈다. 분류기 1에 비하여 분류기 2와 3은 성능의 향상 정도가 크다. 이는 LSTM 레이어를 사용함으로써, 시계열의 연관성을 갖는 거리와 방향에 대하여 더 정확한 추정을 한 것으로 볼 수 있다.

Fig. 6은 제안 모델의 테스트 데이터에 대한 예측 결과의 평균값이다.

V. 결론

본 논문에서는 CNN과 LSTM을 결합하여 사격음에서 화기의 종류, 관측지점에서 사격지점까지의 거리와 방향 등을 추정하는 모델을 제안하였다. 오디오 데이터를 시각화 하여 활용하기 위해 데이터를 멜 스펙트로그램 하여 이미지로 재구성 하였으며, 시계열 데이터를 활용하기 위해 LSTM 레이어를 사용하였다.

제안 모델은 세가지 분류기 모두 90% 이상의 정확도를 가졌으며, CNN만을 사용한 대조모델과 비교했을 때 분류기는 약 3%~18% 이상 정확도가 상승했다. 특히, 사격 지점의 위치와 방향을 분류하는 분류기 2와 3의 경우는 15% 이상의 두드러진 성능 향상을 보였다.

본 논문의 연구 결과는 디지털 포렌식과 군사 및 보안분야에서 사격음을 이용하여 현장을 파악 분석하는데 유용할 것으로 판단된다.

References

1. W. Kim, K. Shin, S. Cho, S. Baek, and Y. Kim, "Development of korean warrior platform architecture" (in Korean), J. of Convergence for Information Technology, **11**, 111-117 (2021).
2. S. Khan, A. Divakaran, and H. S. Sawhney, "Weapon identification using hierarchical classification of acoustic signatures," Proc. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VIII, 730510 (2009).
3. M. Djeddou and T. Touhami, "Classification and modeling of acoustic gunshot signatures," AJSE, **38**, 3399-3406 (2013).
4. K. D. Morton Jr, P. A. Torriane, and L. Collins, "Classification of acoustic gunshot signatures using a non-parametric bayesian signal model," Proc. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X, 80190T (2011).
5. E. Kiktova, M. Lojka, M. Pleva, J. Juhar, and A. Cizmar, "Gun type recognition from gunshot audio recordings," Proc. IEEE 3rd IWBF, 1-6 (2015).
6. S. Raponi, I. Ali, and G. Oligeri, "Sound of guns: digital forensics of gun audio samples meets artificial intelligence," arXiv preprint arXiv:2004.07948 (2020).
7. B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," Proc. 14th Python in Science Conf. 18-24 (2015).

저자 약력

▶ 강 세 혁 (Se Hyeok Kang)



2015년 2월 : 육군사관학교 정보과학 학사
 2020년 5월 : Arizona State University 컴퓨
 터공학 석사
 2021년 3월 ~ 현재 : 육군 3사관학교 전자
 공학과 교수

▶ 조 지 웅 (Ji Woong Cho)



2012년 2월 : 육군사관학교 응용물리학과
 학사
 2017년 2월 : 한국외국어대학교 일본어문
 학과 학사
 2021년 5월 : 일본방위대학교 정보수리학
 석사
 2021년 3월 ~ 현재 : 육군 3사관학교 전자
 공학과 교수