applied sciences

MDPI

Article

A Fast Identification Method of Gunshot Types Based on Knowledge Distillation

Jian Li [1], Jinming Guo [1], Xiushan Sun [1], Chuankun Li [1,*] and Lingpeng Meng [2]

[1] National Key Laboratory of Electronic Testing Technology, North University of China, Taiyuan 030051, China; lijian@nuc.edu.cn (J.L.); s2005018@st.nuc.edu.cn (J.G.); s2005019@st.nuc.edu.cn (X.S.)
[2] Hunan Vanguard Group Co., Ltd., Changsha 410100, China; menglingpeng@861china.com
* Correspondence: chuankun@nuc.edu.cn

Abstract: To reduce the large size of a gunshot recognition network model and to improve the insufficient real-time detection in urban combat, this paper proposes a fast gunshot type recognition method based on knowledge distillation. First, the muzzle blast and the shock wave generated by the gunshot are preprocessed, and the quality of the gunshot recognition dataset is improved using Log-Mel spectrum corresponding to these two signals. Second, a teacher network is constructed using 10 two-dimensional residual modules, and a student network is designed using depth wise separable convolution. Third, the lightweight student network is made to learn the gunshot features under the guidance of the pre-trained large-scale teacher network. Finally, the network's accuracy, model size, and recognition time are tested using the AudioSet dataset and the NIJ Grant 2016-DN-BX-0183 gunshot dataset. The findings demonstrate that the proposed algorithm achieved 95.6% and 83.5% accuracy on the two datasets, the speed was 0.5 s faster, and the model size was reduced to 2.5 MB. The proposed method is of good practical value in the field of gunshot recognition.

Keywords: gunshots; Log-Mel spectrum; knowledge distillation

Citation: Li, J.; Guo, J.; Sun, X.; Li, C.; Meng, L. A Fast Identification Method of Gunshot Types Based on Knowledge Distillation. *Appl. Sci.* **2022**, *12*, 5526. https://doi.org/10.3390/app12115526

Academic Editor: Ana Paula Betencourt Martins Amaro

Received: 28 April 2022
Accepted: 27 May 2022
Published: 29 May 2022

1. Introduction

Gun violence is a daily tragedy that affects the lives of individuals around the world. More than 500 people die every day as a result of gun violence, and 44% of worldwide homicides involve gun violence. Millions of people suffer from the severe and long-term psychological effects of gun violence or the threat of gun violence on individuals, families and their wider communities. Identifying the types of criminal firearms in time can effectively reduce casualties. The identification of the guns held by criminal suspects can help the police determine the threat level of criminal suspects and formulate effective arrest plans. For treating people with gunshot wounds, it can help medical staff to judge the patient's injury and implement targeted treatment by knowing the type of gun. Moreover, gunshot recognition systems can also be used as a complement to daily video surveillance to jointly monitor the occurrence of gun violence. In order to realize the above functions, the gunshot type recognition method needs to meet the requirements of real-time and lightweight. And unlike common audio monitoring and sound event recognition, sniper gunshots of various calibers and types are difficult to identify due to their similar sound spectrogram characteristics. Researchers have done a lot of work on gunshot recognition.

Begault and Beck [1] proposed a forensic audio gunshot analysis method which tries to derive, from the audio produced by live gunfire, important information about the caliber, model, and type of guns used by criminals. Busse et al. [2] proposed a gunshot recognition system based on support vector machine (SVM) to determine whether an audio signal carries a gunshot, achieving an accuracy rate of 70.39%. Ahmed et al. [3] combined Mel-Frequency Cepstral Coefficients (MFCC) with SVM, effectively improving the accuracy of gunshot recognition. Djeddou and Touhami [4] applied the GMM classifier to gunshot

classification, seeking to classify existing gunshot signals out of the noisy environment, and experiments demonstrated that the classification of gunshot signals of five different calibers can reach 96.29% in accuracy. Khan et al. [5] improved on the GMM classifier by embedding samples in it and registered an accuracy of 60–72% in an experiment covering a total of 100 shots of 20 different types. Kiktova et al. [6] proposed a decoding algorithm based on Hidden Markov Models (HMM) and Viterbi, which tries to obtain audio signals from urban noise monitoring systems and then determine the presence or absence of gunshots, and an 80% successful detection rate was achieved. With the development of deep learning technology, neural networks have also found applications in gunshot detection and recognition. Ryan Lilien [7] applied transfer learning to gunshot recognition—the 14-layer CNN network was pre-trained on the large Audioset, and then the resulting model was trained on 6000 individual gunshots of 18 types of guns. A 78.2% accuracy was registered. Raponi and Ali [8] proposed a gunshot classification method based on convolutional neural networks; 3655 pieces of audio from 59 different weapons were classified, giving a 90% accuracy.

In summary, with the continuous improvement in gunshot recognition accuracy, the network model becomes deeper, more training parameters are involved, and there is a greater demand on hardware computing power, leading to a decrease in the speed of gunshot recognition and to insufficient real-time performance. Consequently, it is difficult to achieve miniaturization. To address the above problems, this paper proposes a gunshot recognition method based on knowledge distillation. The scale of the network model is thus reduced without harming the accuracy of gunshot recognition, which lowers the requirement for the computing power of the network so that it can be used in embedded devices.

2. Principles Relating to Sniper Gunshot Recognition

A gunshot signal is complex in composition. It is composed of a muzzle blast [9], a shock wave [10], a secondary mechanical sound, and more. In the gun barrel the gunpowder burns in a narrow space to generate an expanding airflow, which propels the bullet out of the gun chamber. After the bullet escapes the muzzle, the air flow ejected out of the muzzle forms a muzzle blast, which is a very brief sound wave (measured in milliseconds) but of a high intensity (usually 120–160 dB) [11]. The muzzle blast supplies an important basis for interpreting the type and location of the gunshot. But the muzzle wave decays quickly with distance, and is hard to capture at a low signal-to-noise ratio. Therefore, it is difficult to identify the exact type of the gunshot relying on the muzzle wave only [12].

The warhead flying in the air rubs against the surrounding air and when the bullet travels faster than the speed of sound, this process generates a shock wave, which propagates outward from the bullet's path. Since the wave front of the shock wave is characterized by a rapidly rising positive pressure and an immediately dropping negative pressure, the sound wave propagating in the air is a longitudinal wave whose vibration direction is parallel to the propagation direction [13]. Therefore, the wave front of the shock wave features a rapid rise to the maximum amplitude, to be followed by a drop to the minimum amplitude, but it will eventually get back to the initial state of the waveform, as shown in Figure 1.

The form of the wave looks like the letter "N" [14]. Not all muzzle blasts look similar, and not all firearms produce shock waves. Most common pistol projectiles are subsonic and therefore generate no shock waves when flying in the air. As a feature specific to the bullet itself, the shock wave is an important basis for determining the caliber of the gun [15]. Therefore, it is an efficient way to identify the type and caliber of firearms by considering both the muzzle blast and the shock wave, two audio components of the gunshot.

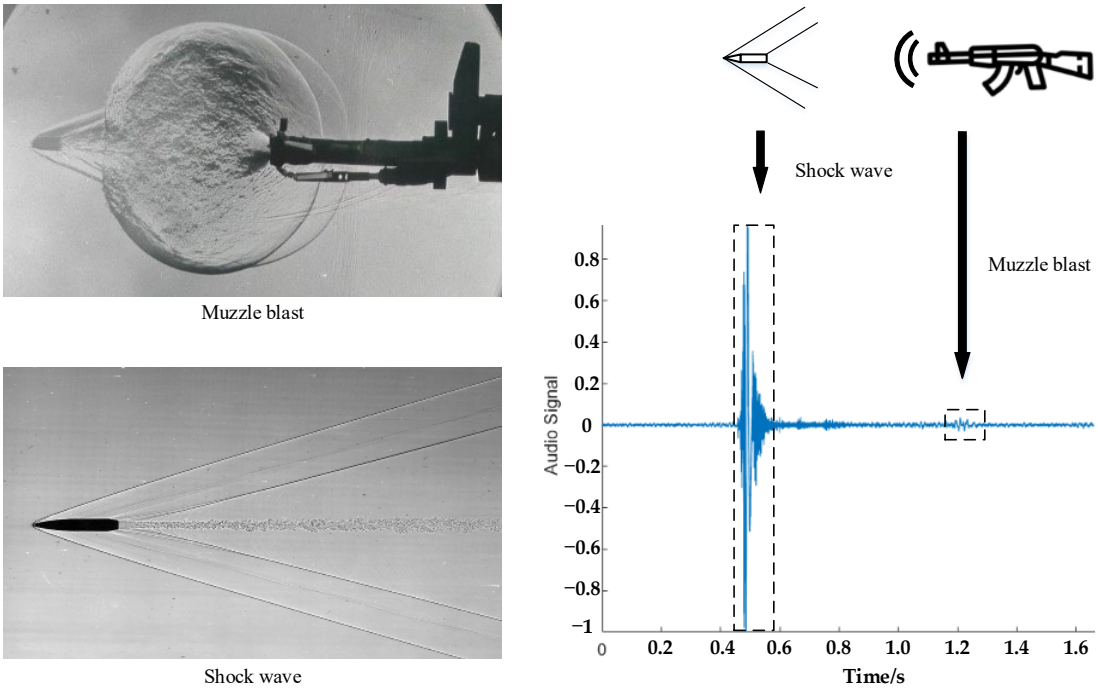

Figure 1. Schematic diagram of a muzzle sound wave.

3. Design of the Gunshot Recognition Network

The overall scheme of gunshot classification proposed in this paper is shown in Figure 2. This method consists of two parts: gunshot preprocessing and a gunshot recognition network. As the first step, the muzzle blast and the shock wave generated by the gunshot are preprocessed, in which the one-dimensional acoustic signal is converted into the corresponding two-dimensional Log-Mel image. A neural network model is then constructed by use of knowledge distillation [16]. Finally, the gun category is identified based on the gunshot information.

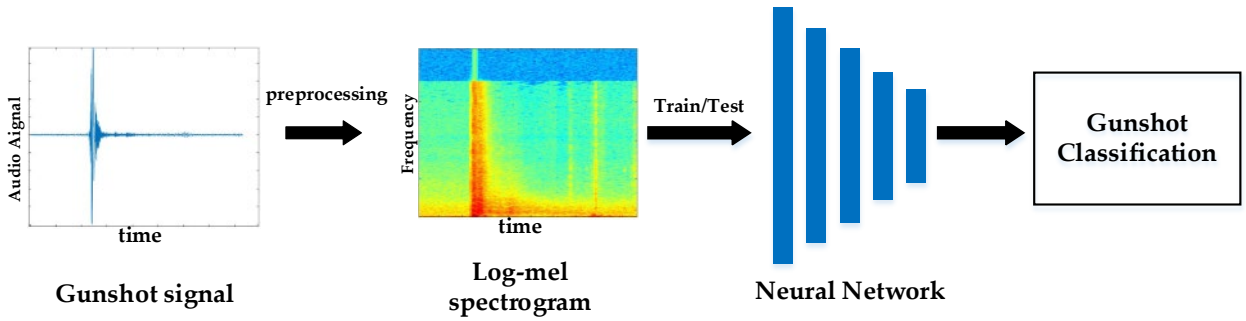

Figure 2. Schematic diagram of the overall scheme of gunshot classification.

3.1. Gunshot Preprocessing

Gunshots are pulse signals that are transient and non-stationary. Directly inputting this one-dimensional signal into a neural network is an impractical way to extract its frequency characteristics. The logarithmic Mel spectrum, as an important characteristic spectrum for sound recognition, possesses a strong feature expression ability. The logarithmic Mel spectrum maps the sound frequencies to a Mel scale using the formula below:

$$f_{mel} = 2595 \times log_{10}\left(1 + \frac{f}{700}\right) \quad (1)$$

where f_{mel} is the perceptual frequency domain in Mel (Mel frequency domain for short), and f is the actual speech frequency in Hz. The actual logarithmic Mel spectrum mapping is shown in Figure 3.

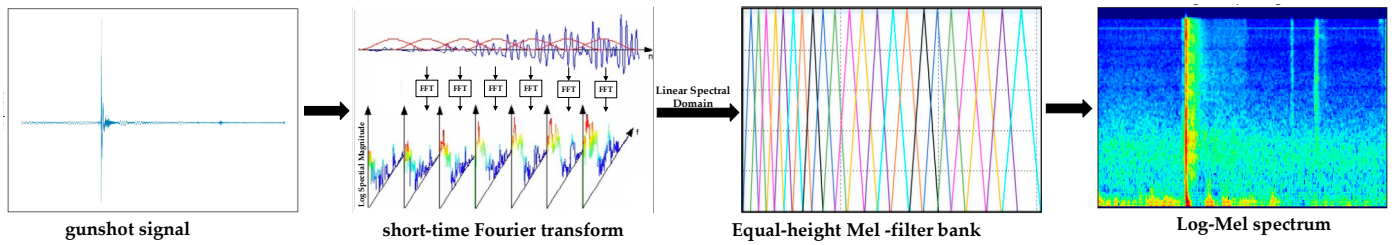

Figure 3. Schematic diagram of the Log-Mel extraction process.

The one-dimensional gunshot signal is first transformed into a short-time Fourier transform, and a power spectrum of size (128, 256) is obtained. The power spectrum is then passed through an equal-height Mel filter bank to obtain a Mel spectrogram. Finally, the logarithm of the energy value of the obtained Mel spectrogram is converted into decibels, and generates the logarithmic Mel spectrogram. The specific transformation parameters of feature map transformation are shown in Table 1.

Table 1. Log-Mel spectral transform specific parameters.

Short-Time Fourier Transform				Mel Transform
Frame Length	**Frame Shift**	**FFT Points**	**Window Type**	**Number of Mel Filters**
2048	1024	1024	Hamming	128

3.2. The Gunshot Recognition Network Based on Knowledge Distillation

Unlike the conventional convolutional neural network [17], knowledge distillation includes additional initialization of large dataset pre-training, and thus enhances the feature extraction ability of the network when addressing a small gunshot sample dataset. Unlike the neural network using transfer learning, knowledge distillation uses the teacher network to tutor the student network, which greatly downscales the network model without much loss to its recognition accuracy.

The learning method employed in knowledge distillation is shown in Figure 4. The network part consists of two networks: the teacher network and the student network. The teacher network is large in scale and complex in structure, and is unfit for portable devices. The student network poses lower computing requirements but gives higher real-time performance, making it suitable for various portable devices with a limited computing power.

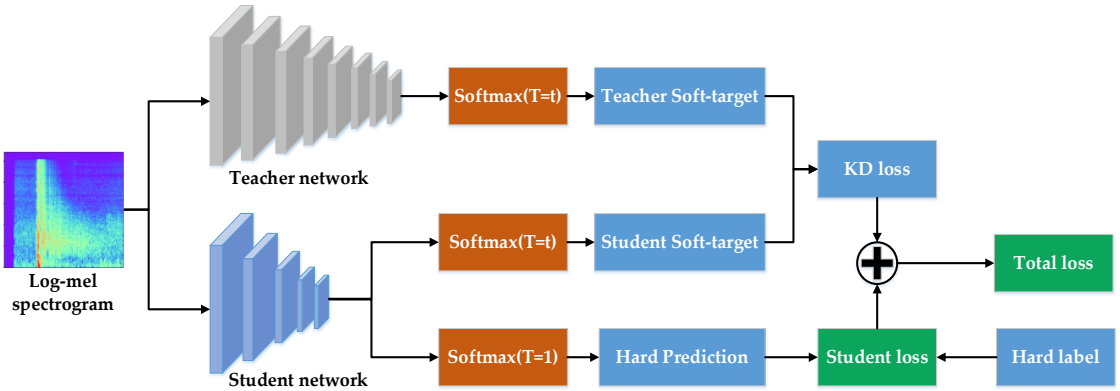

Figure 4. Schematic diagram of the knowledge distillation network.

First, the teacher network is trained on a large Log-Mel dataset to obtain more accurate gunshot recognition probability. Third, we need to make the teacher network, which has already been trained, and the small-scale, lightweight student network, learn a new small-sample dataset together. The teacher network guides the student network to iterate. In this iterative process, the teacher network keeps "distilling" the knowledge it has learned and then transfers the essence to the student network. Finally, the student network is trained.

3.2.1. Teacher Network

In terms of a knowledge distillation network, there are many teacher and student network models, but these are mostly used in image recognition, and are not fully applicable to gunshot category recognition. Ambient sound recognition belongs to acoustic event detection, and has high applicability in image recognition networks. The gunshot signal is different from the ambient sound, and the contours of their Log-Mel spectrograms are too similar. Therefore, textural features are more important than contour features for gunshot recognition. The deep convolutional neural network has been proved to have strong applicability in texture feature extraction, so this paper constructs a teacher network composed of stacks of 10 two-dimensional residual modules in the gunshot recognition task. Figure 5 shows its network structure. Two-dimensional residual modules play an effective role in increasing the depth of the network and precluding the gradient disappearing or exploding problem. After the Log-Mel spectrogram has gone through 10 residual modules for feature extraction, the features are fed into the global average pooling module for classification to produce the probability distribution of sound categories.

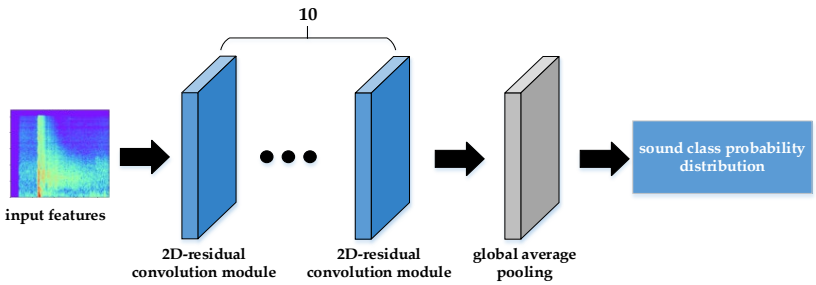

Figure 5. Schematic diagram of the teacher network structure.

The structure of each 2D residual network module is shown in Figure 6. The channel of the input signal first receives batch normalization layer processing (BN) and the Relu activation function. Then a 2D convolution operation is performed, with a convolution kernel size of 3 × 3. The above batch normalization layer, Relu activation function, and 2D convolution operation are repeated before giving the output. The other channel of signal is down-sampled by the average pooling layer, and is then connected in series with the output data of the first channel on the feature channel to give the output of this 2D residual module. The parameters of the teacher network are shown in Table 2.

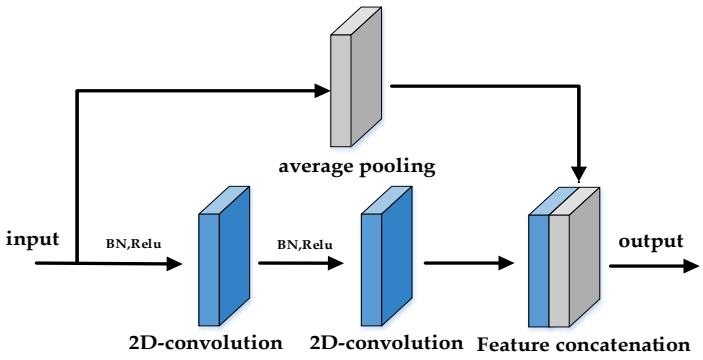

Figure 6. Schematic diagram of the 2D residual network module structure.

Table 2. Teacher network parameters.

Module	Operate	Kernel Size	Output Size
2D Residual Module 1	Conv_layer1 Conv_layer2	3 × 3 × 1 3 × 3 × 1	128 × 256 × 1 128 × 256 × 1
2D Residual Module 2	Conv_layer3 Conv_layer4	3 × 3 × 2 3 × 3 × 2	64 × 128 × 2 64 × 128 × 2
2D Residual Module 3	Conv_layer5 Conv_layer6	3 × 3 × 4 3 × 3 × 4	64 × 128 × 4 64 × 128 × 4
2D Residual Module 4	Conv_layer7 Conv_layer8	3 × 3 × 8 3 × 3 × 8	32 × 64 × 8 32 × 64 × 8
2D Residual Module 5	Conv_layer9 Conv_layer10	3 × 3 × 16 3 × 3 × 16	32 × 64 × 16 32 × 64 × 16
2D Residual Module 6	Conv_layer11 Conv_layer12	3 × 3 × 32 3 × 3 × 32	16 × 32 × 32 16 × 32 × 32
2D Residual Module 7	Conv_layer13 Conv_layer14	3 × 3 × 64 3 × 3 × 64	16 × 32 × 64 16 × 32 × 64
2D Residual Module 8	Conv_layer15 Conv_layer16	3 × 3 × 128 3 × 3 × 128	8 × 16 × 128 8 × 16 × 128
2D Residual Module 9	Conv_layer17 Conv_layer18	3 × 3 × 256 3 × 3 × 256	8 × 16 × 256 8 × 16 × 256
2D Residual Module 10	Conv_layer19 Conv_layer20	3 × 3 × 512 3 × 3 × 512	4 × 8 × 512 4 × 8 × 512
Global average pooling layer	Conv_layer21 Global average pooling	3 × 3 × classes -	4 × 8 × classes classes

3.2.2. Student Network

To improve the real-time performance of the network, the student network is designed based on depth wise separable convolution [18]. Different from regular convolution, depth wise separable convolution splits the convolution operation into two steps: depth wise convolution and pointwise convolution. Depthwise convolution only changes the size of the feature map, while pointwise convolution only changes the number of channels of the feature map. This operation involves fewer parameters and less computation than conventional convolution. Figure 7 gives the actual structure of the student network. It is composed of four depth wise separable convolution modules (same as Bottleneck residual block in MobileNetV2) stacked together. The last convolution changes the input feature shape to a 1 × 1 × category number, and the prediction result is thus obtained.

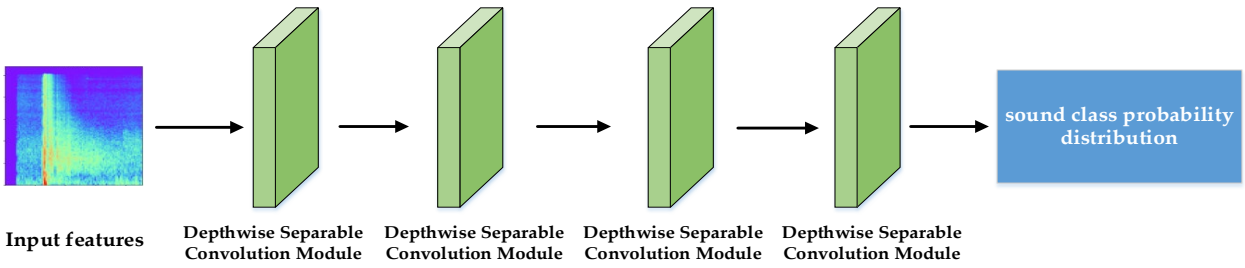

Figure 7. Schematic diagram of the student network structure.

3.2.3. Loss Function

The loss function is the key to ensure that students learn effective knowledge online. The loss function in this paper is divided into two parts: *KD loss* and *Student loss*. *KD loss* is a measure for the student network to follow the teacher network learning. The log output by the teacher network will help the student network to improve its ability (in

addition to the positive example Ground Truth, the negative example also carries a lot of valuable information). *KD loss* can be obtained by performing cross-entropy operation on the probability prediction values output by the teacher network and the student network after softmax-T. The formula for softmax-T is as follows.

$$q_i = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum_j \exp\left(\frac{z_j}{T}\right)} \quad (2)$$

where T is a temperature (This article $T = 4$), z is logits, and q is the mapping result. The temperature coefficient T needs to be divided by the logit before the softmax operation. However, this method will cause the probability of the largest possible category to be large, and the values of other categories to be small and lose some knowledge. However, the distribution of the results outputted by softmax-T is relatively flat, which plays a role in retaining similar information. *Student loss* is the cross entropy between the class probability output and the real label. This loss function ensures that the student network will not be misled by negative examples when learning from the teacher network. The overall loss function in this paper is as follows:

$$Total\ Loss = \ KD\ loss + 0.8 \times Student\ loss \quad (3)$$

3.2.4. Network Training Method

First, the teacher network is trained on the audioset dataset, and its feature extraction ability on the Log-Mel spectrogram is improved through multiple iterations. In this way, the teacher network can be more suitable for small datasets (similar to transfer learning). After obtaining the corresponding teacher network weight model, the knowledge distillation network can be run to help the student network improve the feature extraction ability.

4. Experimental Verification

4.1. Dataset

The AudioSet [19], YouTube Gunshots Dataset [20] and the NIJ Grant 2016-DN-BX-0183 project gunshot datasets [7] were used in this study. The AudioSet dataset was used to pre-train the teacher network, and the other two datasets were used to train the student network. The detailed parameters of these three datasets are shown in Table 3.

Table 3. Dataset.

Dataset	Sample Size
AudioSet	2.1 million audio samples, 527 sound categories
YouTube Gunshots Dataset	840 gun sound samples, 14 gun models
NIJ Grant 2016-DN-BX-0183 Project Gunshots Dataset	6000 gun samples, 18 gun models

The above sound signal was preprocessed and converted into the corresponding Log-Mel spectrogram, as shown in Figure 8 below.

4.2. Network Training

A deep learning environment was built with PyTorch, which incorporated two 1080Ti GPUs, used in combination with the parallel computing architecture CUDA to accelerate the entire training process. First, the teacher network was pre-trained on AudioSet. Eighty percent of the samples in each category were taken as the training set, with the remaining 20 percent as the validation set. The loss function of the network was in cross entropy, and the optimizer was SGD, which iterated 100,000 times.

In the training process, 128 batches of samples were taken at a time for training, at a learning rate of 0.001, to yield the corresponding teacher network model. Then, the NIJ Grant gunshot dataset was subjected to data enhancement (random cropping, mixup [21],

etc.) before being inputted into the teacher network and the student network knowledge distillation model. Each category of gunshots was sampled randomly at 10% to form a validation set. With the cross entropy of the teacher network and the student network as the loss function, and taking SGD as the optimizer, 128 batches of samples were taken at a time for training, at a learning rate of 0.001, and there were 120 iterations.

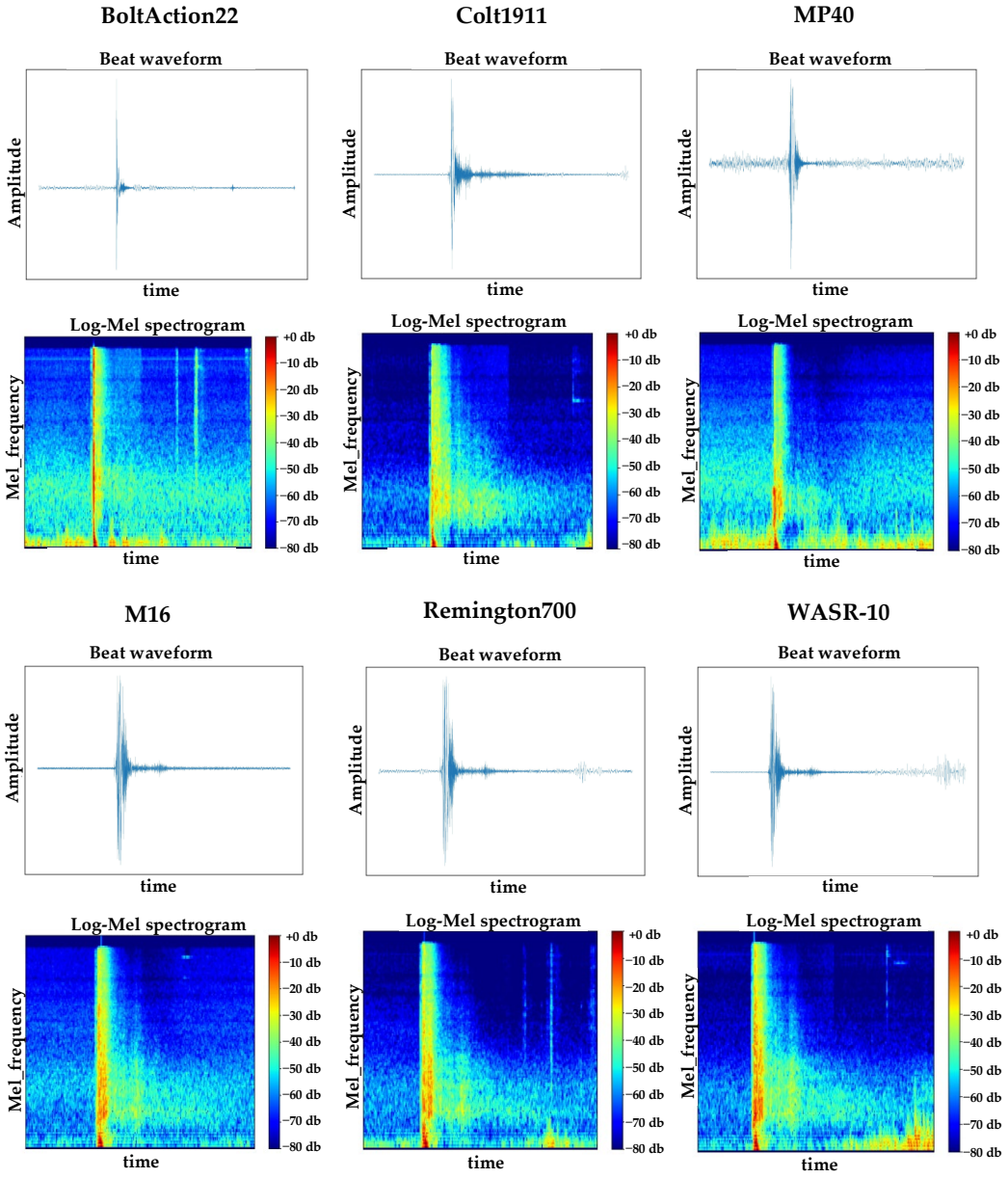

Figure 8. Waveform and Log-Mel spectrogram of the sound sample.

4.3. Experimental Results and Interpretation

4.3.1. Performance Analysis of the Knowledge Distillation Network

Figure 9 gives the loss and the accuracy curve of the knowledge distillation network at the end of 240 iterations. The accuracy of the network validation set was 0.83, and the weight model was reduced to 2.5 MB.

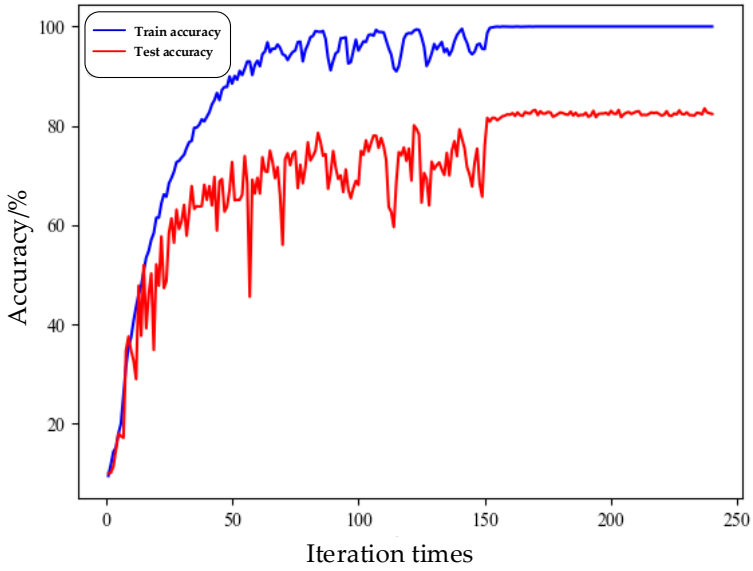

Figure 9. Training set and validation set accuracy of knowledge distillation network training.

4.3.2. Comparison of the Combined Performance of the Teacher Network and the Student Network

Knowledge distillation networks, composed of different teacher networks and student networks, were analyzed for their performance of gunshot recognition. Three networks, from resnet32 × 4 [22], wrn40 × 2 [23] and ours, were used as the teacher network to compare with the proposed network. Resnet8, Resnet8 × 4, WRN16 × 2, MobileNetV2 [24] and ShuffleV1 [25] were used as the student network to compare with the proposed network. Table 4 summarizes the comparison.

Table 4. Recognition accuracy of different teacher network and student network combinations.

Teacher Network	Student Network	Accuracy Rate (%)	Model Size
resnet32 × 4	Resnet8	5.64	337.8 KB
	Resnet8 × 4	80.44	4.9 MB
	WRN16 × 2	56.27	5.5 MB
	MobileNetV2	61.24	3.0 MB
	ShuffleV1	79.12	3.7 MB
	Ours	68.16	2.5 MB
wrn40 × 2	Resnet8	6.18	264 KB
	Resnet8 × 4	39.92	4.6 MB
	WRN16 × 2	71.03	5.5 MB
	MobileNetV2	79.52	4.0 MB
	ShuffleV1	82.16	3.5 MB
	Ours	75.56	2.5 MB
Ours	Resnet8	5.14	264 KB
	Resnet8 × 4	41.21	4.6 MB
	WRN16 × 2	75.24	5.5 MB
	MobileNetV2	72.25	4.0 MB
	ShuffleV1	81.24	3.5 MB
	Ours	83.49	2.5 MB

As can be seen from Table 4, the gunshot recognition ability of the entire network is not only dependent on the individual teacher network and student network, but also on their combinations. The teacher-student network combination proposed in this paper registered the highest accuracy of gunshot recognition, reaching 83.49%.

4.3.3. Ablation Experiment

This paper compares the recognition performance of the teacher network, the student network and the knowledge distillation network. Under the NIJ Grant 2016-DN-BX-0183 project gunshot data set, the teacher network and the student network were transferred, respectively. The initial learning rate was 0.04. After a total of 240 iterations, it decreased when the iteration reached 150, 180, and 210. The accuracy curve is obtained as shown in Figure 10. As can be seen, relative to the teacher network (transfer learning), the knowledge distillation network has narrowed down the oscillation amplitude. Relative to the student network (transfer learning), the knowledge distillation network has improved the accuracy.

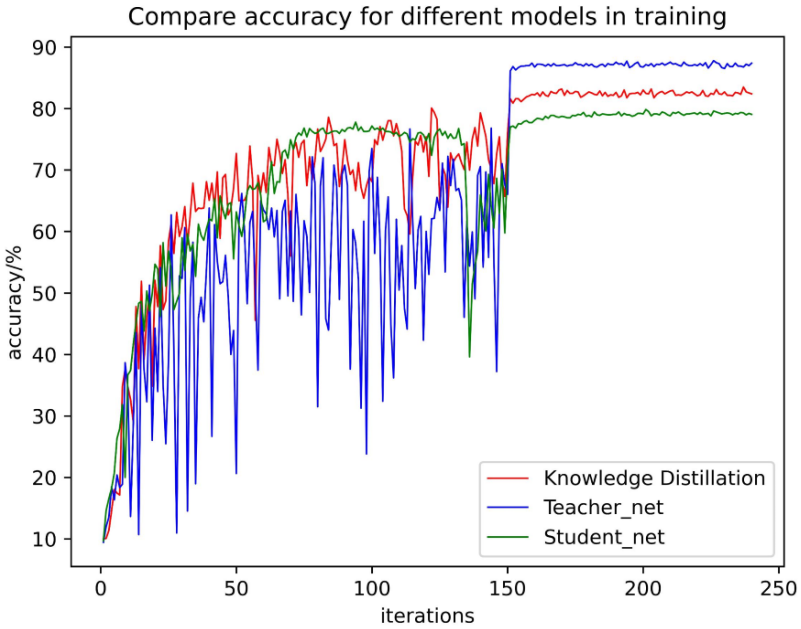

Figure 10. Ablation experiment curve.

Table 5 compares several sets of experiments under the NIJ Grant 2016-DN-BX-0183 project gunshot dataset and the YouTube gunshot dataset.

Table 5. Ablation experiment.

Method	YouTube Gunshot Dataset (Accuracy Rate%)	NIJ Grant Gunshot Dataset (Accuracy Rate%)	Model Size
Teacher network	92.48	82.14	114 MB
Student network	86.42	71.23	2.5 MB
Teacher network (Transfer Learning)	98.4	87.78	114 MB
Student network (Transfer Learning)	91.34	79.84	2.5 MB
Knowledge Distillation	95.6	83.49	2.5 MB

Table 5 suggests that after pre-training on a large dataset, the teacher network and the student network performed much better in gunshot recognition, but the network size was not reduced. Knowledge distillation was confirmed as able to improve the recognition ability of a small network (the student network) without changing the size of the network.

4.3.4. Performance Comparison with Other Methods

Methods for determining gun types by sound are many. However, in gunshot recognition there is no unified and authoritative dataset, and different datasets are applicable to

different studies. Therefore, the transfer learning method as applied to the same dataset was compared with the proposed method in this study. The comparative experiment was carried out using the NIJ Grant 2016-DN-BX-0183 project gunshot dataset and the YouTube gunshot dataset. The results are shown in Table 6 and Figure 11. As suggested by the table, the proposed algorithm achieved 95.6% and 83.5% accuracy on the two datasets, the speed was 0.5 s faster, and the model size was reduced to 2.5 MB.

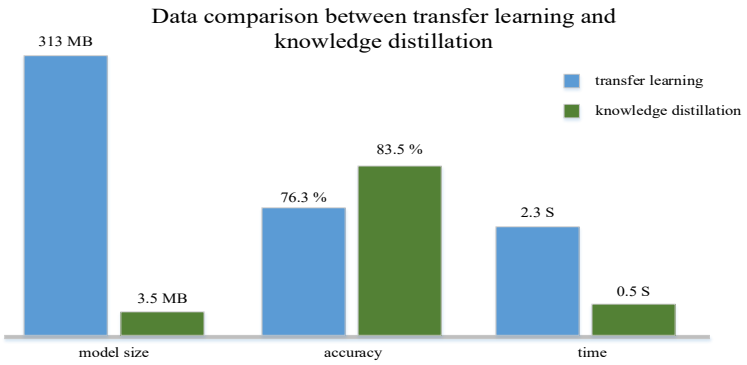

Figure 11. Comparison of transfer learning and knowledge distillation data.

Figure 12 shows the model's verification set confusion matrix. Kimber45 can be confused with Colt1991, and the same is true with Sig9 and Glock9 networks. The first reason is that they were of similar calibers, and the gunshot signals were extremely similar. Next, these pistol projectiles had not exceeded the speed of sound, so they did not carry features such as shock waves, resulting in the lack of some features, which led to a drop in recognition accuracy.

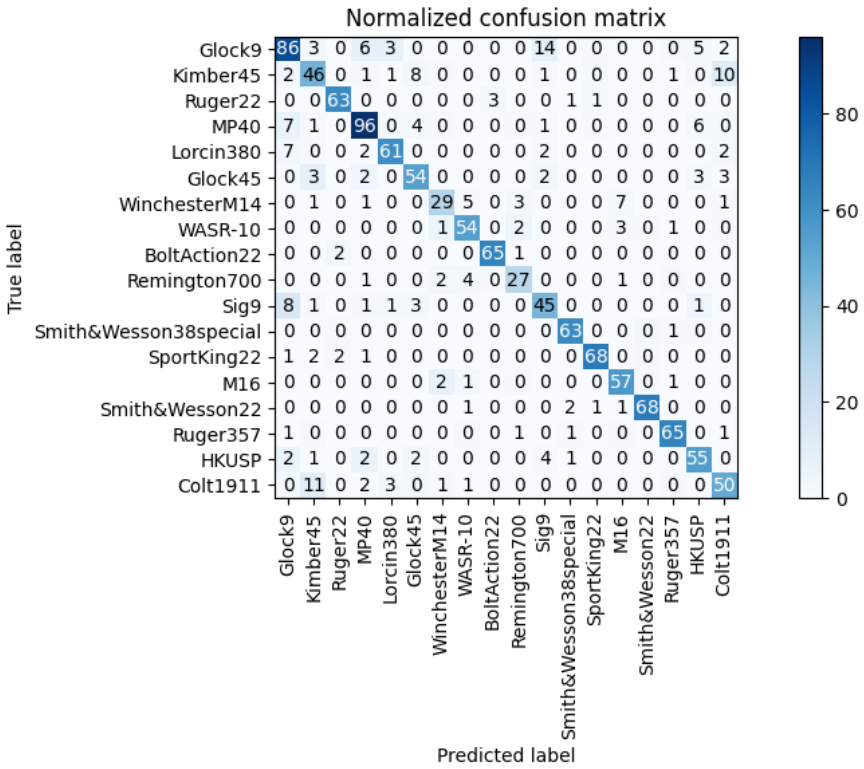

Figure 12. Confusion matrix diagram for the gunshot dataset.

Table 6. Results of identification accuracy test.

Method	YouTube Gunshot Dataset (Accuracy Rate%)	NIJ Grant Gunshot Dataset (Accuracy Rate%)	Model Size	Speed
Transfer Learning	92.4	76.3	312 MB	2.3 S
Knowledge Distillation	95.6	83.5	2.5 MB	0.5 S

5. Conclusions

A fast identification method of gunshot types based on knowledge distillation is proposed in this paper. A pre-trained large-scale teacher network was used to guide a lightweight student network to learn about gunshot features. This method reduced the size of the model, improved its recognition speed, and provided high-speed, concealed, and high-precision recognition ability in daily environments, thus conquering the problem of large model size and long recognition time in the current gunshot recognition field. After experimental verification on the AudioSet dataset and the NIJ Grant 2016-DN-BX-0183 project gunshot dataset, the proposed network was confirmed to as advantageous in accuracy, model size and recognition time. The gunshot recognition method has strong application value in police, medical and other fields.

Author Contributions: Conceptualization, J.L., J.G., X.S., C.L. and L.M.; methodology, J.L., J.G., X.S. and C.L.; software, J.L. and J.G.; investigation, J.L., J.G., X.S., C.L. and L.M.; writing, J.L. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the National Science Foundation of China (No. 61901419) and Fundamental Research Program of Shanxi Province (No. 20210302124031).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Begault, D.R.; Beck, S.D.; Maher, R.C. Overview of forensic audio gunshot analysis techniques. In Proceedings of the Audio Engineering Society Conference: 2019 AES International Conference on Audio Forensics, Porto, Portugal, 18–20 June 2019.
2. Busse, C.; Krause, T.; Ostermann, J.; Bitzer, J. Improved Gunshot Classification by Using Artificial Data. In Proceedings of the Audio Engineering Society Conference: 2019 AES International Conference on Audio Forensics, Porto, Portugal, 18–20 June 2019.
3. Ahmed, T.; Uppal, M.; Muhammad, A. Improving efficiency and reliability of gunshot detection systems. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013.
4. Djeddou, M.; Touhami, T. Classification and modeling of acoustic gunshot signatures. *Arab. J. Sci. Eng.* **2013**, *38*, 3399–3406. [CrossRef]
5. Khan, S.; Divakaran, A.; Sawhney, H.S. Weapon identification across varying acoustic conditions using an exemplar embedding approach. In Proceedings of the Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX, Orlando, FL, USA, 5–9 April 2010.
6. Kiktova, E.; Lojka, M.; Pleva, M.; Juhar, J.; Cizmar, A. Gun type recognition from gunshot audio recordings. In Proceedings of the 3rd International Workshop on Biometrics and Forensics (IWBF 2015), Gjovik, Norway, 3–4 March 2015.
7. Lilien, R. *Development of Computational Methods for the Audio Analysis of Gunshots*; NCJRS; 2016-DN-BX-0183; National Institute of Justice: Washington, DC, USA, 2018.
8. Raponi, S.; Ali, I.; Oligeri, G. Sound of Guns: Digital Forensics of Gun Audio Samples meets Artificial Intelligence. *arXiv* **2020**, arXiv:2004.07948. [CrossRef]
9. Arslan, Y. Impulsive Sound Detection by a Novel Energy Formula and Its Usage for Gunshot Recognition. *arXiv* **2017**, arXiv:1706.08759.
10. Hawthorne, D.L.; Horn, W.; Reinke, D.C. A system for acoustic detection, classification, and localization of terrestrial animals in remote locations. *J. Acoust. Soc. Am.* **2016**, *140*, 3182. [CrossRef]
11. Austin, M.E. On the Frequency Spectrum of N-Waves. *J. Acoust. Soc. Am.* **1967**, *41*, 528. [CrossRef]

12. Nimmy, P.; Rajesh, K.R.; Nimmy, M.; Vishnu, S. Shock Wave and Muzzle Blast Identification Techniques Utilizing Temporal and Spectral Aspects of Gunshot Signal. In Proceedings of the 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Thiruvananthapuram, India, 6–8 December 2018.
13. Maher, R.C. Modeling and Signal Processing of Acoustic Gunshot Recordings. In Proceedings of the 2006 IEEE 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop, Teton National Park, WY, USA, 24–27 September 2006; Volume 4, pp. 257–261.
14. Libal, U.; Spyra, K. Wavelet based shock wave and muzzle blast classification for different supersonic projectiles. *Expert Syst. Appl.* **2014**, *41*, 5097–5104. [CrossRef]
15. Aguilar, J. Gunshot Detection Systems in Civilian Law Enforcement. *J. Audio Eng. Soc.* **2015**, *63*, 280–291. [CrossRef]
16. Hinton, G.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. *Comput. Sci.* **2015**, *14*, 38–39.
17. Li, C.; Li, S.; Gao, Y.; Zhang, X.; Li, W. A Two-stream Neural Network for Pose-based Hand Gesture Recognition. *IEEE Trans. Cogn. Dev. Syst.* **2021**, 1–10. [CrossRef]
18. Howard, A.G.; Zhu, M. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
19. Gemmeke, J.F.; Ellis, D.P.; Freedman, D.; Jansen, A.; Lawrence, W.; Moore, R.C.; Plakal, M.; Ritter, M. Audio Set: An ontology and humanlabeled dataset for audio events. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 776–780.
20. Sánchez-Hevia, H.A.; Ayllón, D.; Gil-Pita, R.; Rosa-Zurera, M. Maximum likelihood decision fusion for weapon classification in wireless acoustic sensor networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 1172–1182. [CrossRef]
21. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. Mixup: Beyond empirical risk minimization. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2016**, arXiv:1512.03385.
23. Zagoruyko, S.; Komodakis, N. Wide Residual Networks. *arXiv* **2016**, arXiv:1605.07146.
24. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2018**, arXiv:1801.04381.
25. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.