CLAIMS

1. An artificial intelligence-based audio processing method, comprising:

obtaining an audio clip of an audio scene, the audio clip comprising noise;

performing audio scene classification processing based on the audio clip to obtain an audio scene type corresponding to the noise in the audio clip; and

determining a target audio processing mode corresponding to the audio scene type, and applying the target audio processing mode to the audio clip according to a degree of interference caused by the noise in the audio clip,

wherein the target audio processing mode comprises a noise reduction processing mode; and

the determining a target audio processing mode corresponding to the audio scene type comprises:

determining a noise type matching the audio scene type based on the audio scene type corresponding to the audio scene; and

querying a correspondence between different candidate noise types and candidate noise reduction processing modes based on the noise type matching the audio scene type, to obtain a noise reduction processing mode corresponding to the audio scene type,

noise types matching different audio scene types being not exactly the same.

2. The method according to claim 1, wherein the applying the target audio processing mode to the audio clip according to a degree of interference caused by the noise in the audio clip comprises:

determining the degree of interference caused by the noise in the audio clip; and

applying the noise reduction processing mode corresponding to the audio scene type to the audio clip in response to the degree of interference being greater than an

interference degree threshold.

3. The method according to claim 1, wherein the applying the target audio processing mode to the audio clip comprises:

performing matching processing on the noise type matching the audio scene type and the noise in the audio clip; and

suppressing the noise successfully matched with the noise type to obtain a suppressed audio clip,

a ratio of a speech signal strength to a noise signal strength in the suppressed audio clip being lower than a signal-to-noise ratio threshold.

4. The method according to claim 1, wherein

the target audio processing mode comprises a bitrate switching processing mode; and

the determining a target audio processing mode corresponding to the audio scene type comprises:

querying a correspondence between different candidate audio scene types and candidate bitrate switching processing modes based on the audio scene type corresponding to the audio scene, to obtain a bitrate switching processing mode corresponding to the audio scene type.

5. The method according to claim 1, wherein

the target audio processing mode comprises a bitrate switching processing mode; and

the determining a target audio processing mode corresponding to the audio scene type comprises:

comparing the audio scene type corresponding to the audio scene with a preset audio scene type; and

determining a bitrate switching processing mode associated with the preset audio

scene type as a bitrate switching processing mode corresponding to the audio scene type in response to determining through the comparison that the audio scene type is the preset audio scene type.

6. The method according to claim 4 or 5, wherein the applying the target audio processing mode to the audio clip comprises:

obtaining a communication signal strength of the audio scene;

reducing an audio bitrate of the audio clip according to a first set ratio or a first set value in response to the communication signal strength of the audio scene being less than a communication signal strength threshold; and

increasing the audio bitrate of the audio clip according to a second set ratio or a second set value in response to the communication signal strength of the audio scene being greater than or equal to the communication signal strength threshold.

7. The method according to claim 4 or 5, wherein the applying the target audio processing mode to the audio clip comprises:

determining jitter information of a strength of a communication signal in the audio scene based on strengths of the communication signal that are obtained by multiple times of sampling in the audio scene; and

reducing an audio bitrate of the audio clip according to a third set ratio or a third set value in response to the jitter information indicating that the communication signal is in an unstable state.

8. The method according to claim 4 or 5, wherein the applying the target audio processing mode to the audio clip comprises:

reducing an audio bitrate of the audio clip according to a fourth set ratio or set value in response to a type of a communication network for transmitting the audio clip being a set type.

9. The method according to claim 1, wherein

the audio scene classification processing is implemented by a neural network

model, and the neural network model learns an association between the noise comprised in the audio clip and the audio scene type; and

the performing audio scene classification processing based on the audio clip to obtain an audio scene type corresponding to the noise in the audio clip comprises:

calling the neural network model based on the audio clip to perform the audio scene classification processing to obtain an audio scene type having an association with the noise comprised in the audio clip.

10. The method according to claim 9, wherein

the neural network model comprises a mapping network, a residual network, and a pooling network; and

the calling the neural network model based on the audio clip to perform the audio scene classification processing comprises:

performing feature extraction processing on the audio clip through the mapping network to obtain a first feature vector of the noise in the audio clip;

performing mapping processing on the first feature vector through the residual network to obtain a mapping vector of the audio clip;

performing feature extraction processing on the mapping vector of the audio clip through the mapping network to obtain a second feature vector of the noise in the audio clip;

performing pooling processing on the second feature vector through the pooling network to obtain a pooled vector of the audio clip; and

performing non-linear mapping processing on the pooled vector of the audio clip to obtain the audio scene type having an association with the noise comprised in the audio clip.

11. The method according to claim 10, wherein

the mapping network comprises a plurality of cascaded mapping layers; and

the performing feature extraction processing on the audio clip through the mapping network to obtain a first feature vector of the noise in the audio clip comprises:

performing feature mapping processing on the audio clip through a first mapping layer in the plurality of cascaded mapping layers;

outputting a mapping result of the first mapping layer to a subsequent mapping layer in the plurality of cascaded mapping layers, and continuing to perform feature mapping and mapping result outputting through the subsequent mapping layer, until an output is provided to a last mapping layer; and

determining a mapping result outputted by the last mapping layer as the first feature vector of the noise in the audio clip.

12. The method according to claim 10, wherein

the residual network comprises a first mapping network and a second mapping network; and

the performing mapping processing on the first feature vector through the residual network to obtain a mapping vector of the audio clip comprises:

performing mapping processing on the first feature vector through the first mapping network to obtain a first mapping vector of the audio clip;

performing nonlinear mapping processing on the first mapping vector to obtain a non-mapping vector of the audio clip;

performing mapping processing on the non-mapping vector of the audio clip through the first mapping network to obtain a second mapping vector of the audio clip; and

determining a summation result of the first feature vector of the audio clip and the second mapping vector of the audio clip as the mapping vector of the audio clip.

13. The method according to claim 1, further comprising:

constructing audio samples respectively corresponding to a plurality of different audio scenes based on a noise-free audio signal and background noise respectively corresponding to the plurality of different audio scenes; and

training a neural network model based on the audio samples respectively corresponding to the plurality of different audio scenes to obtain a neural network model for audio scene classification.

14. The method according to claim 13, wherein the constructing audio samples respectively corresponding to a plurality of different audio scenes based on a noise-free audio signal and background noise respectively corresponding to the plurality of different audio scenes comprises:

performing following processing for any one of the plurality of different audio scenes:

fusing the background noise of the audio scene and the noise-free audio signal based on a fusion ratio of the background noise of the audio scene and the noise-free audio signal to obtain a first fused audio signal of the audio scene;

fusing the background noise of the audio scene corresponding to a first random coefficient in the first fused audio signal to obtain a second fused audio signal of the audio scene; and

fusing the noise-free audio signal corresponding to a second random coefficient in the second fused audio signal to obtain an audio sample of the audio scene.

15. The method according to claim 13, wherein the training a neural network model based on the audio samples respectively corresponding to the plurality of different audio scenes to obtain a neural network model for audio scene classification comprises:

performing audio scene classification processing on the audio samples respectively corresponding to the plurality of different audio scenes through the neural network model, to obtain predicted audio scene types of the audio samples;

constructing a loss function of the neural network model based on the predicted audio scene types of the audio samples, audio scene annotations of the audio samples, and weights of the audio samples; and

updating a parameter of the neural network model until the loss function converges, and determining the updated parameter of the neural network model that causes the loss function to converge as a parameter of the neural network model for audio scene classification.

16. The method according to claim 1, further comprising: before the performing audio scene classification processing based on the audio clip,

performing framing processing on a time domain signal of the audio clip to obtain a multi-frame audio signal;

windowing the multi-frame audio signal, and performing a Fourier transform on the windowed audio signal to obtain a frequency domain signal of the audio clip; and

performing logarithmic processing on a Mel band of the frequency domain signal to obtain the audio clip for the audio scene classification.

17. An audio processing apparatus, comprising:

an obtaining module, configured to obtain an audio clip of an audio scene, the audio clip comprising noise;

a classification module, configured to perform audio scene classification processing based on the audio clip to obtain an audio scene type corresponding to the noise in the audio clip; and

a processing module, configured to determine a target audio processing mode corresponding to the audio scene type, and apply the target audio processing mode to the audio clip according to a degree of interference caused by the noise in the audio clip,

wherein the target audio processing mode comprises a noise reduction processing mode, and the processing module is configured to:

determine a noise type matching the audio scene type based on the audio scene type corresponding to the audio scene; and

query a correspondence between different candidate noise types and candidate noise reduction processing modes based on the noise type matching the audio scene type, to obtain a noise reduction processing mode corresponding to the audio scene type,

noise types matching different audio scene types being not exactly the same.

18. An electronic device, comprising:

a memory, configured to store executable instructions; and

a processor, configured to execute the executable instructions stored in the memory to implement the artificial intelligence-based audio processing method according to any one of claims 1 to 16.

19. A computer-readable storage medium, storing executable instructions, the executable instructions, when executed by a processor, implementing the artificial intelligence-based audio processing method according to any one of claims 1 to 16.

20. A computer program product, comprising a computer program or instructions, the computer program or instructions causing a computer to execute the artificial intelligence-based audio processing method according to any one of claims 1 to 16.