AudioSoundEvent v1.0.2 Model card

Label_list:
0_backgroundnoise
1_babycry
2_glassbreak
3_gunshot

Preprecess:
Wav Sample_rate: 8000
FBANK Dim 24(mean average)

Postprecess:
softmax(4 classes)

Model Structure

```
Layer (type:depth-idx)                             Output Shape              Param #
==========================================================================================
ResNetSE                                           [1, 4]                    --
├──Conv2d: 1-1                                     [1, 4, 24, 192]           36
├──BatchNorm2d: 1-2                                [1, 4, 24, 192]           8
├──ReLU: 1-3                                       [1, 4, 24, 192]           --
├──Sequential: 1-4                                 [1, 8, 24, 192]           --
│    └──SEBottleneck: 2-1                          [1, 8, 24, 192]           --
│    │    └──Conv2d: 3-1                           [1, 4, 24, 192]           16
│    │    └──BatchNorm2d: 3-2                      [1, 4, 24, 192]           8
│    │    └──ReLU: 3-3                             [1, 4, 24, 192]           --
│    │    └──Conv2d: 3-4                           [1, 4, 24, 192]           144
│    │    └──BatchNorm2d: 3-5                      [1, 4, 24, 192]           8
│    │    └──ReLU: 3-6                             [1, 4, 24, 192]           --
│    │    └──Conv2d: 3-7                           [1, 8, 24, 192]           32
│    │    └──BatchNorm2d: 3-8                      [1, 8, 24, 192]           16
│    │    └──SELayer: 3-9                          [1, 8, 24, 192]           25
│    │    └──Sequential: 3-10                      [1, 8, 24, 192]           48
│    │    └──ReLU: 3-11                            [1, 8, 24, 192]           --
│    └──SEBottleneck: 2-2                          [1, 8, 24, 192]           --
```

```
│    │    └──Conv2d: 3-12                    [1, 4, 24, 192]           32
│    │    └──BatchNorm2d: 3-13               [1, 4, 24, 192]           8
│    │    └──ReLU: 3-14                      [1, 4, 24, 192]           --
│    │    └──Conv2d: 3-15                    [1, 4, 24, 192]           144
│    │    └──BatchNorm2d: 3-16               [1, 4, 24, 192]           8
│    │    └──ReLU: 3-17                      [1, 4, 24, 192]           --
│    │    └──Conv2d: 3-18                    [1, 8, 24, 192]           32
│    │    └──BatchNorm2d: 3-19               [1, 8, 24, 192]           16
│    │    └──SELayer: 3-20                   [1, 8, 24, 192]           25
│    │    └──ReLU: 3-21                      [1, 8, 24, 192]           --
├──Sequential: 1-5                          [1, 16, 12, 96]           --
│    └──SEBottleneck: 2-3                    [1, 16, 12, 96]           --
│    │    └──Conv2d: 3-22                    [1, 8, 24, 192]           64
│    │    └──BatchNorm2d: 3-23               [1, 8, 24, 192]           16
│    │    └──ReLU: 3-24                      [1, 8, 24, 192]           --
│    │    └──Conv2d: 3-25                    [1, 8, 12, 96]            576
│    │    └──BatchNorm2d: 3-26               [1, 8, 12, 96]            16
│    │    └──ReLU: 3-27                      [1, 8, 12, 96]            --
│    │    └──Conv2d: 3-28                    [1, 16, 12, 96]           128
│    │    └──BatchNorm2d: 3-29               [1, 16, 12, 96]           32
│    │    └──SELayer: 3-30                   [1, 16, 12, 96]           82
│    │    └──Sequential: 3-31                [1, 16, 12, 96]           160
│    │    └──ReLU: 3-32                      [1, 16, 12, 96]           --
│    └──SEBottleneck: 2-4                    [1, 16, 12, 96]           --
│    │    └──Conv2d: 3-33                    [1, 8, 12, 96]            128
│    │    └──BatchNorm2d: 3-34               [1, 8, 12, 96]            16
│    │    └──ReLU: 3-35                      [1, 8, 12, 96]            --
│    │    └──Conv2d: 3-36                    [1, 8, 12, 96]            576
│    │    └──BatchNorm2d: 3-37               [1, 8, 12, 96]            16
│    │    └──ReLU: 3-38                      [1, 8, 12, 96]            --
│    │    └──Conv2d: 3-39                    [1, 16, 12, 96]           128
│    │    └──BatchNorm2d: 3-40               [1, 16, 12, 96]           32
│    │    └──SELayer: 3-41                   [1, 16, 12, 96]           82
│    │    └──ReLU: 3-42                      [1, 16, 12, 96]           --
├──Sequential: 1-6                          [1, 32, 6, 48]            --
│    └──SEBottleneck: 2-5                    [1, 32, 6, 48]            --
│    │    └──Conv2d: 3-43                    [1, 16, 12, 96]           256
│    │    └──BatchNorm2d: 3-44               [1, 16, 12, 96]           32
```

```
│       │       └──ReLU: 3-45                  [1, 16, 12, 96]           --
│       │       └──Conv2d: 3-46                [1, 16, 6, 48]
2,304
│       │       └──BatchNorm2d: 3-47           [1, 16, 6, 48]            32
│       │       └──ReLU: 3-48                  [1, 16, 6, 48]            --
│       │       └──Conv2d: 3-49                [1, 32, 6, 48]            512
│       │       └──BatchNorm2d: 3-50           [1, 32, 6, 48]            64
│       │       └──SELayer: 3-51               [1, 32, 6, 48]            292
│       │       └──Sequential: 3-52            [1, 32, 6, 48]            576
│       │       └──ReLU: 3-53                  [1, 32, 6, 48]            --
│       └──SEBottleneck: 2-6                   [1, 32, 6, 48]            --
│       │       └──Conv2d: 3-54                [1, 16, 6, 48]            512
│       │       └──BatchNorm2d: 3-55           [1, 16, 6, 48]            32
│       │       └──ReLU: 3-56                  [1, 16, 6, 48]            --
│       │       └──Conv2d: 3-57                [1, 16, 6, 48]
2,304
│       │       └──BatchNorm2d: 3-58           [1, 16, 6, 48]            32
│       │       └──ReLU: 3-59                  [1, 16, 6, 48]            --
│       │       └──Conv2d: 3-60                [1, 32, 6, 48]            512
│       │       └──BatchNorm2d: 3-61           [1, 32, 6, 48]            64
│       │       └──SELayer: 3-62               [1, 32, 6, 48]            292
│       │       └──ReLU: 3-63                  [1, 32, 6, 48]            --
│       └──SEBottleneck: 2-7                   [1, 32, 6, 48]            --
│       │       └──Conv2d: 3-64                [1, 16, 6, 48]            512
│       │       └──BatchNorm2d: 3-65           [1, 16, 6, 48]            32
│       │       └──ReLU: 3-66                  [1, 16, 6, 48]            --
│       │       └──Conv2d: 3-67                [1, 16, 6, 48]
2,304
│       │       └──BatchNorm2d: 3-68           [1, 16, 6, 48]            32
│       │       └──ReLU: 3-69                  [1, 16, 6, 48]            --
│       │       └──Conv2d: 3-70                [1, 32, 6, 48]            512
│       │       └──BatchNorm2d: 3-71           [1, 32, 6, 48]            64
│       │       └──SELayer: 3-72               [1, 32, 6, 48]            292
│       │       └──ReLU: 3-73                  [1, 32, 6, 48]            --
│       └──SEBottleneck: 2-8                   [1, 32, 6, 48]            --
│       │       └──Conv2d: 3-74                [1, 16, 6, 48]            512
│       │       └──BatchNorm2d: 3-75           [1, 16, 6, 48]            32
│       │       └──ReLU: 3-76                  [1, 16, 6, 48]            --
```

```
│    │    └──Conv2d: 3-77                    [1, 16, 6, 48]
2,304
│    │    └──BatchNorm2d: 3-78               [1, 16, 6, 48]            32
│    │    └──ReLU: 3-79                      [1, 16, 6, 48]            --
│    │    └──Conv2d: 3-80                    [1, 32, 6, 48]            512
│    │    └──BatchNorm2d: 3-81               [1, 32, 6, 48]            64
│    │    └──SELayer: 3-82                   [1, 32, 6, 48]            292
│    │    └──ReLU: 3-83                      [1, 32, 6, 48]            --
├──Sequential: 1-7                          [1, 64, 3, 24]            --
│    └──SEBottleneck: 2-9                    [1, 64, 3, 24]            --
│    │    └──Conv2d: 3-84                    [1, 32, 6, 48]
1,024
│    │    └──BatchNorm2d: 3-85               [1, 32, 6, 48]            64
│    │    └──ReLU: 3-86                      [1, 32, 6, 48]            --
│    │    └──Conv2d: 3-87                    [1, 32, 3, 24]
9,216
│    │    └──BatchNorm2d: 3-88               [1, 32, 3, 24]            64
│    │    └──ReLU: 3-89                      [1, 32, 3, 24]            --
│    │    └──Conv2d: 3-90                    [1, 64, 3, 24]
2,048
│    │    └──BatchNorm2d: 3-91               [1, 64, 3, 24]            128
│    │    └──SELayer: 3-92                   [1, 64, 3, 24]
1,096
│    │    └──Sequential: 3-93                [1, 64, 3, 24]            2,176
│    │    └──ReLU: 3-94                      [1, 64, 3, 24]            --
│    └──SEBottleneck: 2-10                   [1, 64, 3, 24]            --
│    │    └──Conv2d: 3-95                    [1, 32, 3, 24]
2,048
│    │    └──BatchNorm2d: 3-96               [1, 32, 3, 24]            64
│    │    └──ReLU: 3-97                      [1, 32, 3, 24]            --
│    │    └──Conv2d: 3-98                    [1, 32, 3, 24]
9,216
│    │    └──BatchNorm2d: 3-99               [1, 32, 3, 24]            64
│    │    └──ReLU: 3-100                     [1, 32, 3, 24]            --
│    │    └──Conv2d: 3-101                   [1, 64, 3, 24]
2,048
│    │    └──BatchNorm2d: 3-102              [1, 64, 3, 24]            128
│    │    └──SELayer: 3-103                  [1, 64, 3, 24]
```

1,096
│ │ └──ReLU: 3-104 [1, 64, 3, 24] --
├──TemporalAveragePooling: 1-8 [1, 192] --
├──BatchNorm1d: 1-9 [1, 192] 384
├──Linear: 1-10 [1, 64]
12,352
├──BatchNorm1d: 1-11 [1, 64] 128
├──Linear: 1-12 [1, 4] 260
==========
==========
Total params: 61,574
Trainable params: 61,574
Non-trainable params: 0
Total mult-adds (M): 10.73
==========
==========
Input size (MB): 0.02
Forward/backward pass size (MB): 7.19
Params size (MB): 0.25
Estimated Total Size (MB): 7.46

Accuracy:
loss: 0.12032, accuracy: 0.96124