UrbanSound8k sound classification

Completed by Serhii Syrota

Model architecture

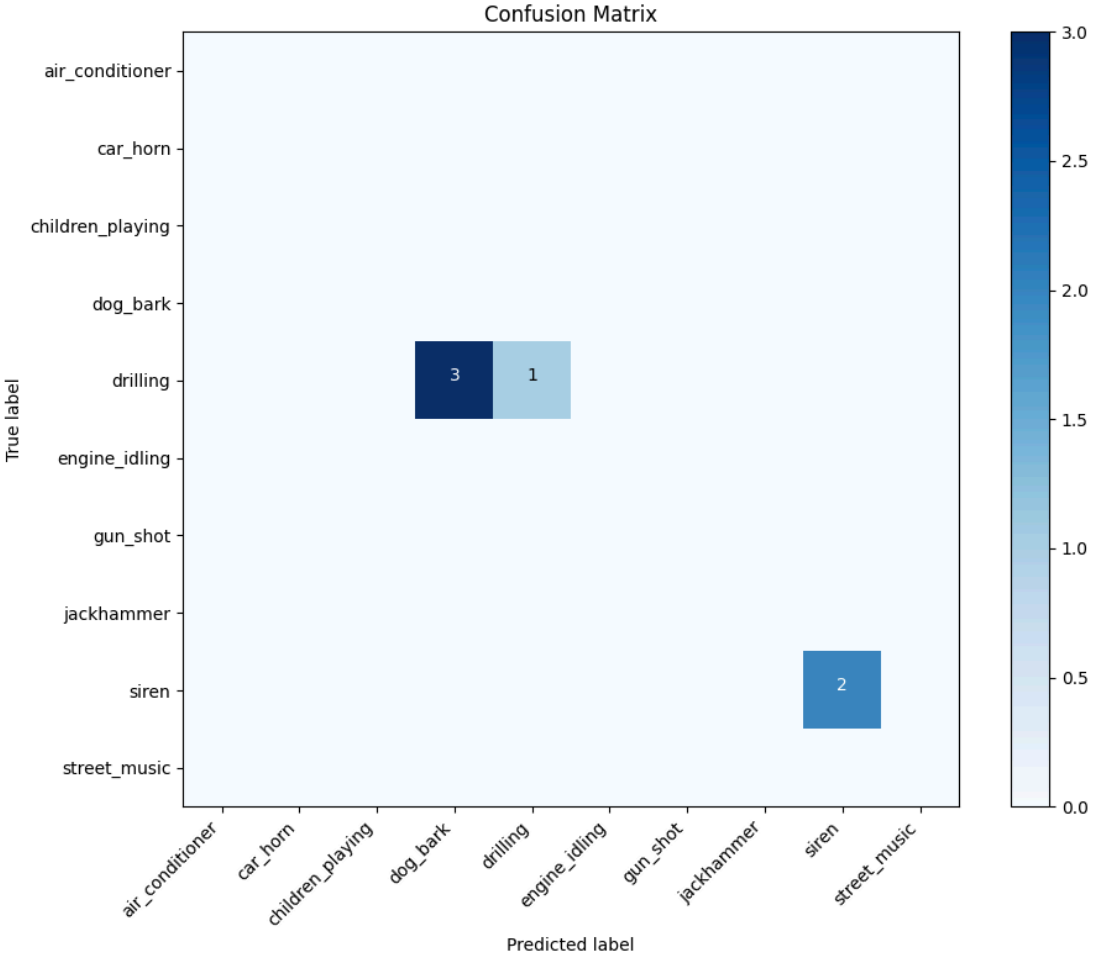
```
model = keras.Sequential([
            keras.layers.Conv2D(32, (12, 12), activation='relu',
input_shape=self.input_shape),
            keras.layers.MaxPooling2D((8, 8)),
            keras.layers.Conv2D(64, (3, 3), activation='relu'),
            keras.layers.MaxPooling2D((2, 2)),
            keras.layers.Conv2D(128, (3, 3), activation='relu'),
            keras.layers.MaxPooling2D((2, 2)),
            keras.layers.Flatten(),
            keras.layers.Dense(128, activation='relu'),
            keras.layers.Dropout(0.1),
            keras.layers.Dense(self.num_classes, activation='softmax')
        ])
        model.compile(optimizer='adam',
                      loss='categorical_crossentropy',
                      metrics=['accuracy'])
```

The generated images has size 256x256, but the size can be reduced to 64x64(with network first layer tuning) without accuracy loss.

Results

```
55/55 ━━━━━━━━━━━━━━━━━━━━ 2s 41ms/step - accuracy: 0.9962 - loss: 0.0093
Epoch 47/50
55/55 ━━━━━━━━━━━━━━━━━━━━ 2s 41ms/step - accuracy: 0.9964 - loss: 0.0118
Epoch 48/50
55/55 ━━━━━━━━━━━━━━━━━━━━ 2s 42ms/step - accuracy: 0.9916 - loss: 0.0269
Epoch 49/50
55/55 ━━━━━━━━━━━━━━━━━━━━ 2s 41ms/step - accuracy: 0.9951 - loss: 0.0211
Epoch 50/50
55/55 ━━━━━━━━━━━━━━━━━━━━ 2s 42ms/step - accuracy: 0.9659 - loss: 0.0956
128/128 ━━━━━━━━━━━━━━━━━━━━ 8s 57ms/step - accuracy: 0.8825 - loss: 1.2941
```

Real data confusion


Confusion Matrix
air_conditioner
car_horn
children_playing
dog_bark
drilling
engine_idling
gun_shot
jackhammer
siren
street_music
True label
3
1
2
Predicted label
3.0
2.5
2.0
1.5
1.0
0.5
0.0