Urban Sound Classification

How to classify urban noise in order to improve environmental monitoring, noise pollution control, or urban planning in smart cities?

Gwendoline Hays-Valentin, Hugo Michel, Thomas Rigou, Charaf Zguiouar

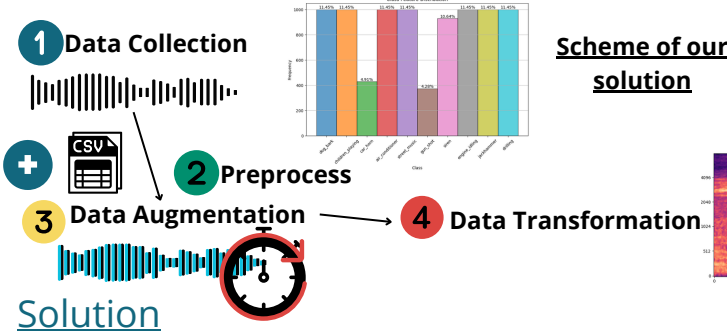
Link to our GitHub

Summary

Urban noise pollution is a significant concern in modern cities. It can have negative health impacts, reduce quality of life, and hinder productivity. To address this challenge, researchers are exploring the use of sound classification in smart city initiatives. This poster will explore how deep learning techniques can be applied to classify urban sounds, paving the way for improved environmental monitoring, noise pollution control, and urban planning.

Problem

Our goal is to develop a system that can accurately categorize various sounds present in a city environment. These sounds can be diverse, ranging from traffic noise and construction activities to animal sounds and human voices. By leveraging deep learning, we can create models capable of distinguishing between these different sound classes.



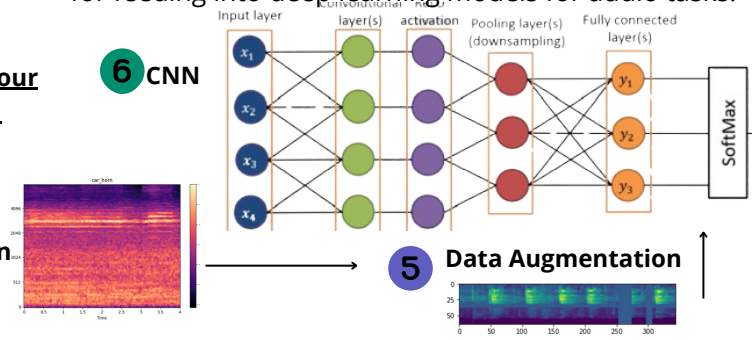
Scheme of our solution

Solution

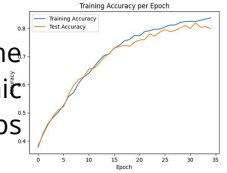
Han et al. (2021) demonstrated the effectiveness of CNNs for sound classification, achieving high accuracy in identifying different sound sources. Similarly, **Piczak (2015)** showed the use of Mel Spectrograms as a robust feature representation for audio data, which facilitated the training of deep learning models. Additionally, integrating extensive metadata has been proven to enhance model performance by providing context for the audio signals, as discussed by **Khamparia et al. (2019)**.

Building on these findings, we follow an approach based on this process :

1. **Data Collection**: The dataset contains 8732 labeled sound in WAV format excerpts (<=4s) of urban sounds from 10 classes (*dog_bark*, *drilling*, *engine_idling*,...). The classes are drawn from the urban sound taxonomy. **J. Salamon et al.**, *"A Dataset and Taxonomy for Urban Sound Research".*
2. **Preprocessing**: To prepare the audio for our model, we standardize it by:
   - Converting mono to stereo (duplicating the channel)
   - Setting all audio to the same sampling rate
3. **Data Augmentation (time):** We augment the raw audio by randomly shifting it slightly in time (Time Shift).
4. **Data Transformation:** We transform the audio into Mel Spectrograms. These capture the key features ideal for feeding into deep learning models for audio tasks.



5. **Data Augmentation (audio) :** We can further augment the Mel Spectrogram (rather than the raw audio) using SpecAugment. This technique temporarily hides portions of the frequency or time domain to improve the model's robustness to variations.
6. **CNN:** The actual model architecture is similar to standard image classification. We use a CNN with convolutional blocks to extract features from the spectrograms, followed by a softmax layer for final classes prediction. We also integrate batch normalization to stabilize the training and add dropout to prevent the overfitting.
7. **Training :** We define functions for the optimizer (Adam), loss, and a dynamic learning rate scheduler. This typically helps the model converge faster.



We train the model by iterating through batches of data, tracking accuracy on a validation set (simulating unseen data for this demo). Finally, we disable training and use the model to make predictions on the validation data.

8 Results

The end-to-end model achieves 82% accuracy in classifying urban sounds, validated with high precision, recall, and F1-scores. "Jackhammer" and "gun_shot" have high recall, important for urban monitoring. The confusion matrix reveals common misclassifications between "engine_idling" and "drilling", as well as "children_playing" and "dog_bark", due to similar acoustic features. Overall, the model demonstrates reliable performance across diverse urban sounds.

Precision : 0.82	Recall : 0.82	F1-Score : 0.82

	precision	recall	f1-score
air_conditioner	0.78	0.91	0.84
car_horn	0.79	0.80	0.79
children_playing	0.74	0.74	0.74
dog_bark	0.84	0.74	0.79
drilling	0.82	0.78	0.80
engine_idling	0.86	0.76	0.81
gun_shot	0.79	0.96	0.87
jackhammer	0.90	0.92	0.91
siren	0.85	0.88	0.86
street_music	0.79	0.77	0.78