Urban Sound Propagation	This dataset is assembled for research into urban sound propagation, comprising 25,000 data points across 10 diverse cities. Each city is represented by 2,500 locations, offering a comprehensive analysis of various urban configurations. The dataset utilizes OpenStreetMap (OSM) imagery to detail the urban layout within a 500m x 500m area for each location, where buildings are delineated with black pixels and open spaces with white pixels. Supplementing the urban structural images, the dataset includes sound distribution maps at resolutions of 512x512 and 256x256. These maps are precisely generated through the NoiseModelling v4.0 framework, an advanced simulation tool engineered for accurate modeling of sound dynamics within urban environments. For researchers and experts interested in exploring the intricacies of sound simulation, additional insights can be obtained from the NoiseModelling framework documentation.
DATASET LINK	DATA CARD AUTHOR(S)
Provide a link to the dataset:	*Select **one role per** Data Card Author:* *(Usage Note: Select the most appropriate choice to describe the author's role in creating the Data Card.)*
https://doi.org/10.5281/zenodo.10609793	**Name, Team:** Anonymous, Anonymous for Blind Review **Name, Team:** Anonymous, Anonymous for Blind Review **Name, Team:** Anonymous, Anonymous for Blind Review

Authorship

Publishers

PUBLISHING ORGANIZATION(S)	INDUSTRY TYPE(S)	CONTACT DETAIL(S)
Provide the names of the institution or organization responsible for publishing the dataset:	*Select **all applicable** industry types to which the publishing organizations belong:*	*Provide publisher contact details:*
Organization Name	Anonymous for Blind Review	**Publishing POC:** Anonymous for Blind Review **Affiliation:** Anonymous for Blind Review **Contact:** Anonymous for Blind Review **Mailing List:** Anonymous for Blind Review **Website:** Anonymous for Blind Review

Dataset Owners

TEAM(S)	CONTACT DETAIL(S)	AUTHOR(S)
Provide the names of the groups or team(s) that own the dataset:	*Provide pathways to contact dataset owners:*	*Provide the details of all authors associated with the dataset:* *(Usage Note: Provide the affiliation and year if different from publishing institutions or multiple affiliations.)*
Anonymous for Blind Review	**Dataset Owner(s):** Anonymous for Blind Review **Affiliation:** Anonymous for Blind Review **Contact:** Anonymous for Blind Review **Group Email:** Anonymous for Blind Review **Website:** Anonymous for Blind Review	Anonymous for Blind Review

Funding Sources

INSTITUTION(S)	FUNDING OR GRANT SUMMARY(IES)
Provide the names of the funding institution(s):	*Provide a short summary of programs or projects that may have funded the creation, collection, or curation of the dataset.* *Use additional notes to capture any other relevant information or considerations.*
Anonymous for Blind Review Anonymous for Blind Review	Anonymous for Blind Review

Dataset Overview

DATA SUBJECT(S)	DATASET SNAPSHOT	CONTENT DESCRIPTION
*Select **all applicable** subjects contained the dataset:*	*Provide a snapshot of the dataset:* *(Use the additional notes to include relevant information, considerations, and links to table(s) with more detailed breakdowns.)*	*Provide a short description of the content in a datapoint.*
Data about natural phenomena Data about places and objects	<table><tr><td>Size of Dataset</td><td>~5 GB</td></tr><tr><td>Number of Instances</td><td>~100,000</td></tr><tr><td>Training</td><td>19908 x 4</td></tr><tr><td>Evaluation</td><td>3732 x 4</td></tr><tr><td>Test</td><td>1244 x 4</td></tr></table> **Additional Notes:** The dataset is segmented into four distinct subsets, each tailored to explore specific aspects of sound propagation in urban environments: Baseline, Reflection, Diffraction, and Combined.	A data point in this dataset consists of two main components: an urban layout image from OpenStreetMap and a corresponding sound distribution map. The urban layout image is a 500m x 500m area depiction where buildings are marked in black and open spaces in white. The sound distribution map generated using NoiseModelling v4.0, illustrates the sound dynamics within that urban environment at resolutions of 512x512 or 256x256.

Dataset Version and Maintenance

MAINTENANCE STATUS	VERSION DETAILS	MAINTENANCE PLAN
*Select **one**:*	*Provide details about **this** version of the dataset:*	*Summarize the maintenance plan for the dataset:* *Use additional notes to capture any other relevant information or considerations.*
Regularly Updated (New versions of the dataset have been or will continue to be made available.)	**Current Version:** 2.0 **Last Updated:** 02/2024 **Release Date:** 02/2024	**Feedback:** Anonymous for Blind Review

Example of Data Points

PRIMARY DATA MODALITY	SAMPLING OF DATA POINTS	DATA FIELDS
*Select **one**:*	*Provide link(s) to data points or exploratory demos:*	*List the fields in data points and their descriptions.* *(Usage Note: Describe each field in a datapoint. Optionally use this to show the example.)*

Multimodal

- Image Data
- Geospatial Data
- Tabular Data

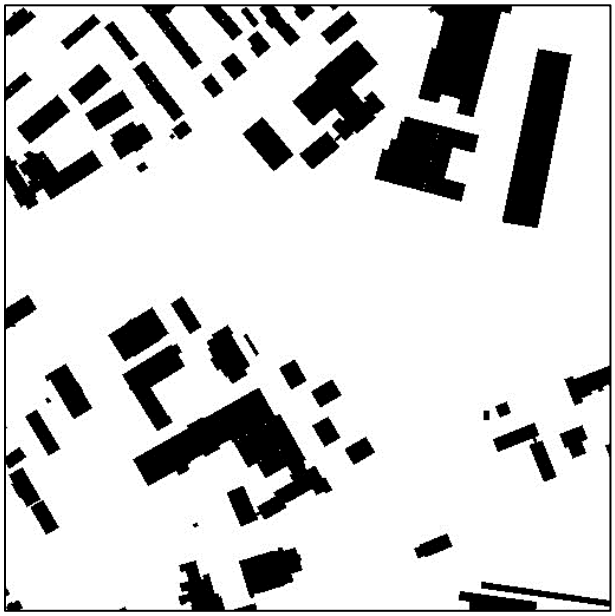
Field Name	Field Value	Description
lat	float	Latitude of the sound measurement location.
long	float	Longitude of the sound measurement location.
db	Object	Key-value pairs of sound levels in decibels for a given frequency (lwd{fqz}).
soundmap	string	Path to 256x256 resolution sound distribution image.
soundmap_512	string	Path to 512x512 resolution sound distribution image.
osm	string	Path to Open Street Map image showing urban layout.
temperature	float	Temperature (°C) at the location.
humidity	float	Humidity (%) at the location.
yaw	float	Orientation of the noise source. Can be empty.
sample_id	int	Unique identifier for the data point.

TYPICAL DATA POINT	EXAMPLE OF DATA POINT
Provide an example of a typical data point and describe what makes it typical. ***Use additional notes to capture any other relevant information or considerations.***	

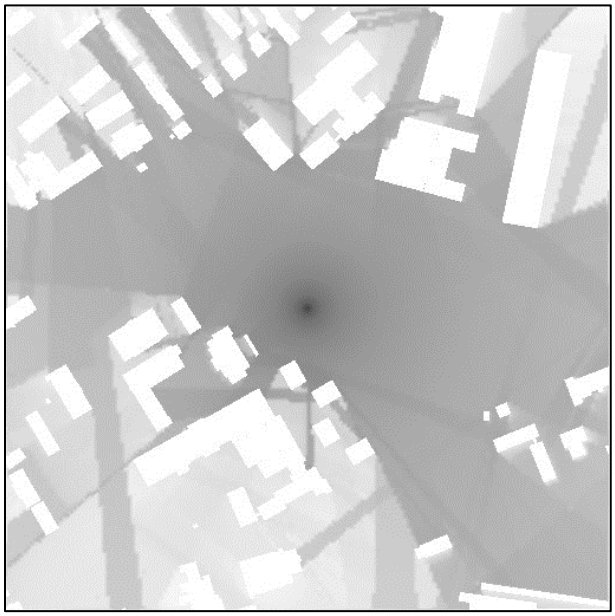

```
``` {
  "lat": 48.030229082138526,
  "long": 11.367773397906852,
  "db": {"lwd500": 69},
  "soundmap":
  "./soundmaps/256/0_LEQ_256.png",
  "soundmap_512":
  "./soundmaps/512/0_LEQ_512.png",
  "osm":
"./buildings/osm_23747.png",
  "temperature": 12,
  "humidity": 35,
  "yaw": None,
  "sample_id": "23747"
  } ```
```


Below is an example of an OSM and Simulated Sound Propagation pair:

OSM:



Simulated Sound Propagation:


```

Motivations & Intentions

Motivations

PURPOSE(S)	DOMAIN(S) OF APPLICATION	MOTIVATING FACTOR(S)
*Select **one**:*	*Provide a list of key domains of application that the dataset has been designed for:* *(Usage Note: Use comma-separated keywords.)*	*List the primary motivations for creating or curating this dataset:* *(Usage Note: use this to describe the problem space and corresponding motivations for the dataset.)*
Research	*`Generative Models`, `1-step Physic Simulation`, `Sound Propagation`, `Machine Learning`*	Generative models, through their capacity to learn from complex datasets, hold significant potential in understanding the intricate physics behind sound propagation. By training on data that encompasses various urban layouts and the corresponding sound distribution maps, these models can uncover the underlying patterns and principles governing how sound travels and interacts with different obstacles, such as buildings and open spaces. This capability enables the creation of predictive models that can simulate sound behavior in any urban environment, thereby offering valuable insights for urban planning, acoustic design, and noise mitigation strategies, all rooted in a deep understanding of the physical laws of sound propagation.

Intended Use

DATASET USE(S)	SUITABLE USE CASE(S)	UNSUITABLE USE CASE(S)
*Select **one**:*	*Summarize known suitable and intended use cases of this dataset.* *Use additional notes to capture any specific patterns that readers should look out for, or other relevant information or considerations.*	*Summarize known unsuitable and unintended use cases of this dataset.* *Use additional notes to capture any specific patterns that readers should look out for, or other relevant information or considerations.*
Safe for research use	**Sound propagation**: Enhancing models for predicting how sound travels in densely built areas.	**Predicting indoor noise levels**: The dataset is designed for outdoor urban sound distribution, not for indoor environments. **Traffic flow or congestion analysis:** It focuses on sound distribution and ignores vehicle movements or traffic patterns.

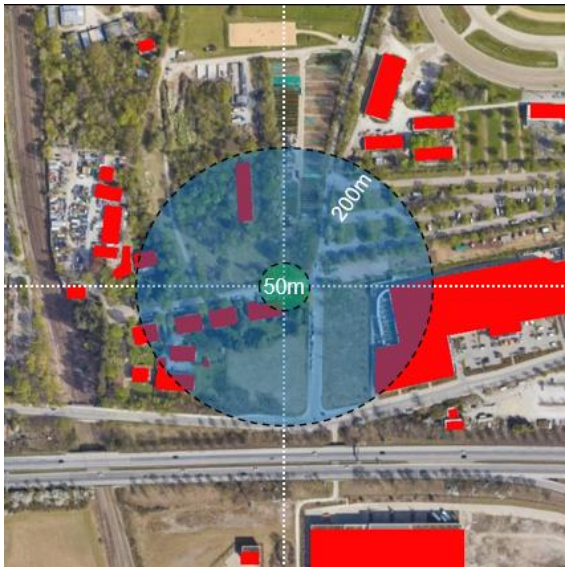
	RESEARCH AND PROBLEM SPACE(S)	CITATION GUIDELINES
	Provide a description of the specific problem space that this dataset intends to address.	*Provide guidelines and steps for citing this dataset in research and/or production work.* *Use additional notes to capture any specific patterns that readers should look out for, or other relevant information or considerations.*
	The dataset specifically addresses the problem space of outdoor urban noise propagation. It is intended for developing models that can predict sound distribution around buildings within urban environments.	**Guidelines & Steps:** <Summarize here. Include links where necessary.> **BiBTeX:** ``` Anonymous for Blind Review ``` **Additional Notes:** <Add here>

Provenance

Collection

METHOD(S) USED	METHODOLOGY DETAIL(S)	SOURCE DESCRIPTION(S)
*Select **all applicable** methods used to collect data:*	*Provide a description of each collection method used.* *Use additional notes to capture any other relevant information or considerations.* *(Usage Note: Duplicate and complete the following for collection method type.)*	*Provide a description of each upstream source of data.* *Use additional notes to capture any other relevant information or considerations.*
- API - Physical Simulation Framework	**Overpass API** **Source:** The Overpass API is a read-only API that serves up custom selected parts of the OSM map data. It acts as a database over the web: the client sends a query to the API and gets back the data set that corresponds to the query. **Platform:** https://overpass-api.de/ **Is this source considered sensitive or high-risk?** [Yes / **No**] **Dates of Collection:** [10 2023 - 12 2024] **Primary modality of collected data:** Geospatial Data **NoiseModelling v4.0** **Source:** An advanced simulation tool engineered for accurate modeling of sound dynamics within urban environments. **Platform:** https://github.com/Universite-Gustave-Eiffel/NoiseModelling **Is this source considered sensitive or high-risk?** [Yes / **No**] **Dates of Collection:** [10 2023 - 12 2024] **Primary modality of collected data:** Geospatial Data	**OSM Buildings:** This source provides images from Open Street Map (OSM) that depict urban layouts, specifically focusing on buildings within cities. In these images, black pixels represent buildings, and white pixels indicate open spaces. **Sound Propagation:** This component of the dataset involves simulated sound distribution images around urban centers, where the noise source is placed at the center.

Collection Criteria

DATA SELECTION	DATA INCLUSION	DATA EXCLUSION
Summarize the data selection criteria. *Use additional notes to capture any other relevant information or considerations.*	*Summarize the data inclusion criteria.* *Use additional notes to capture any other relevant information or considerations.*	*Summarize the data exclusion criteria.* *Use additional notes to capture any other relevant information or considerations.*
Location Sampling: The locations are randomly sampled across 10 cities/areas: ["Hamburg", "Hannover", "Augsburg", "Bonn", "Muenchen", "Schwerin", "Berlin", "Paris", "Stuttgart", "Aachen"]	**Enough Obstacles:** At least 10 Buildings within a circle r=200m around the sound source. **No Obstacle to close:** No Building within a r=50m circle around the sound source. **Additional Notes:** 	**Additional Notes:** If the Data Inclusion criteria is not met, the data is excluded. No additional exclusion criteria are introduced.