

TRAFFIC NOISE ESTIMATION FROM SATELLITE IMAGERY WITH DEEP LEARNING

Leonardo Eicher, Michael Mommert, Damian Borth

AIML Lab, School of Computer Science, University of St. Gallen

ABSTRACT

Road traffic noise represents a global health issue. Despite its importance, noise data are unavailable in many regions of the world. We therefore propose to approximate noise data from satellite imagery in an end-to-end Deep Learning approach. We train a U-Net segmentation model to estimate road noise based on freely available Sentinel-2 satellite imagery and existing road traffic noise estimates for Switzerland. We are able to achieve an RMSE of 8.8 dB(A) for daytime traffic noise and 7.6 dB(A) for nighttime traffic noise with a spatial resolution of 10 m. In addition to identifying major road networks, our model succeeds to predict the spatial propagation of noise. Our results suggest that this approach provides a pathway to estimating road traffic noise for areas for which no such measures are available.

Index Terms— Remote Sensing, Deep Learning, Traffic Noise.

1. INTRODUCTION

Vehicle road noise constitutes the second most harmful environmental stressor in Europe, with 82 million people in urban areas being constantly exposed to dangerous noise levels [1], exceeding the limits recommended by the WHO. Affected individuals are at an elevated risk of adverse health effects, including ischaemic heart disease, hypertension, and strokes [2, 3]. Since on-site noise measurements by ground sensors are expensive and impractical to conduct to scale, numerous models have been developed that estimate traffic noise based on proxy variables.

Today, noise mapping is widely achieved by physical models that formulate a relationship between multiple input variables and the resulting noise. However, since physical models like sonROAD18 [4], which is currently employed in Switzerland, rely on parameters related to traffic, road, and environmental conditions as inputs, they require the collection of substantial amounts of data. Alternative approaches utilize Artificial Neural Networks in a regression setup [5, 6, 7], but still rely on the manual collection of numerous input variables over large areas.

In this work, we investigate the feasibility of estimating road traffic noise from satellite imagery only, based on road noise estimates for Switzerland derived with the sonROAD18

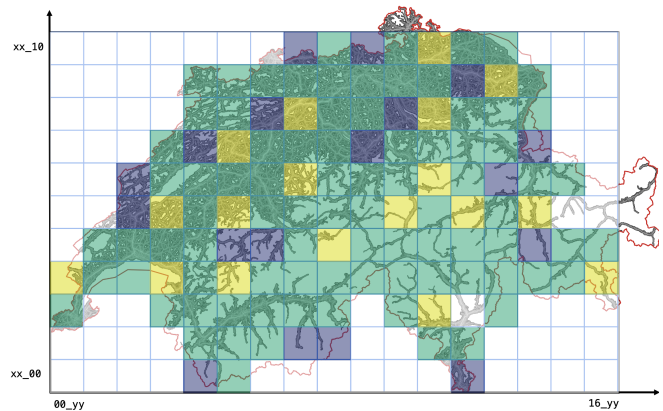


Fig. 1. Daytime road traffic noise map based on sonROAD18 [4] and tiling scheme utilized for noise data and satellite imagery; colored tiles indicate their assignment to the training (green), validation (blue) or test (yellow) dataset.

model [4]. This approach is inspired by methods for the extraction of road networks from aerial images [8, 9], but extends that work by also learning the noise characteristics of different areas and road types. Our method enables the estimation of road traffic noise on a global scale, including areas where no such estimates are currently available.

2. DATA

We estimate road traffic noise from Sentinel-2 satellite imagery, utilizing only the highest resolution bands 2, 3, 4 and 8 (B, G, R, and NIR) with a spatial resolution of 10×10 m on the ground. We download imaging data obtained between 2020-07 to 2020-09 of Switzerland without cloud coverage and crop the images into 19×19 km tiles (see Figure 1) for better handling.

We utilize road traffic noise data provided by the Swiss Federal Office for the Environment (FOEN) as ground-truth for this study. Based on the sonROAD18 [4] physical model, two separate datasets are provided for daytime (6am - 10pm) and nighttime (10pm - 6am) road traffic noise, both of which are used in this work independently. Each noise dataset consists of a raster image containing the modeled noise in units of dB(A) with a spatial resolution of 10 m, covering Switzerland.

Pixel-wise noise in the daytime dataset ranges between 0 and 96 dB(A) (0–89 dB(A) in the nighttime dataset); the average noise level across the whole dataset and country corresponds to 10.68 dB(A) (7.94 dB(A), respectively). Missing values in the noise maps (typically caused by large buildings) are replaced by median values calculated within a 25×25 pixel sliding window. Finally, a Gaussian filter is applied to smoothen the noise map (see Section 3).

For training and evaluation, spatially aligned satellite image crops and noise map crops are provided to our model; prediction is solely based on satellite imagery. From each 19×19 km tile we generate ~ 100 crops (see Section 3), resulting in a total of more than 14,000 image/noise map pairs. We split this dataset into a training, validation and test set in such a way as to account for noise statistics on the national level: utilizing the same tiling scheme that is used for the satellite imagery, we derive the average road noise level for each 19×19 km tile and assign each tile to either the training, validation and test dataset in such a way as to reflect the noise statistics across the entire country (see Figure 1). 70% of the tiles are assigned to the training dataset, 15% each to the validation and test datasets.

3. METHODS & EXPERIMENTAL SETUP

We frame the task of estimating road traffic noise as a pixel-wise regression task and utilize a U-Net architecture [10] implemented in PyTorch. We input a satellite image and obtain a noise estimate for every pixel as output. For optimization, we use an MSE loss function in combination with an Adam optimizer and a cosine annealing learning rate scheduler.

The following experiments are conducted: (1) For data augmentation purposes, random flipping & rotations, as well as Gaussian noise are applied to both datasets. (2) Image and noise map crop sizes in the range of $\{100, 224, 300, 500\}$ pixels are applied to assess the effect of increasing the amount of context the model is able to observe when making a noise estimation. (3) Lastly, the degree of Gaussian smoothing on the ground truth noise data (see Section 2) with different kernel sizes is investigated.

4. RESULTS

Our best-performing model (for both daytime and nighttime noise data) is a U-Net trained with a batch tile size of 224×224 pixels and smoothing with a Gaussian kernel with $\sigma = 1$ pixel applied on the ground truth and a learning rate of 1^{-04} with a cosine annealing schedule (no warm restarts). Horizontal & vertical flips and random rotations with a probability of 0.5 are applied as data augmentations. We train the model for 60 epochs using the MSE loss function with a batch size of 16.

We achieve an average RMSE of 8.79 ± 0.06 dB(A) on the daytime test set and 7.60 ± 0.07 dB(A) for the nighttime test

set; uncertainties are standard deviations of RMSE over five runs with different random seed values. These performance metrics underline our ability to estimate noise levels properly. Figure 2 also shows that we are also able to properly identify road networks and estimate their impact on noise patterns.

5. DISCUSSION

As expected, our model performs best in settings with clearly visible roads, especially in rural environments, and for isolated highways. Predictions in urban environments turn out to be more challenging, as the **limited spatial resolution** of the satellite images (10×10 m per pixel) impedes the detection of roads against a crowded urban background. This results in a lower level of detail in the predicted noise maps and an elevated background noise level (see Figure 2). Nevertheless, major roads and their contribution to the noise pattern is typically detected at a high signal-to-noise level. As a further result of the limited imaging resolution, noise protection measures, such as walls around highways in urban areas, are too small to be resolved and therefore ignored by our model. Furthermore, we point out that our approach only predicts **temporally static** noise estimates (ignoring diurnal variations) as provided by the underlying sonROAD18 physical model.

While the RMSE metric is comparable in both **urban and rural settings**, predictions in rural environments (see, e.g., Figure 2) turn out to be more precise, whereas the distribution of predicted noise values in urban environments is more uniform and thus distributed closer to the mean. This results in a lower standard deviation of values in our prediction than in the ground truth (compare plot titles in columns 2 and 3 in Figure 2). The more robust performance in rural areas is likely in part also attributable to the composition of the dataset, which mainly contains rural areas. Since most areas are not in close proximity to roads, the national median value in the daytime and nighttime noise datasets is 0 dB(A). To compensate for this, the MSE loss function is used as it penalizes outliers stronger, putting stronger weights on areas with high noise levels.

We investigate whether our model correctly predicts different **noise profiles**. Figure 3 displays cross-sections of our predictions and the ground-truth noise data for three sample locations. Our model correctly predicts noise profiles of different shapes for different road types: noise profiles of highways are generally wider than those of smaller roads; the number of lanes in highway is accounted for. Therefore, in addition to simply identifying and extracting road networks, our model also learns road-specific noise characteristics.

We note that our model performs slightly better on the **nighttime** noise map, which might be due to the fact that nighttime traffic noise levels are generally lower, resulting in a narrower spectral distribution. Therefore, the nighttime model is more likely to make reasonable predictions, reducing RMSE.

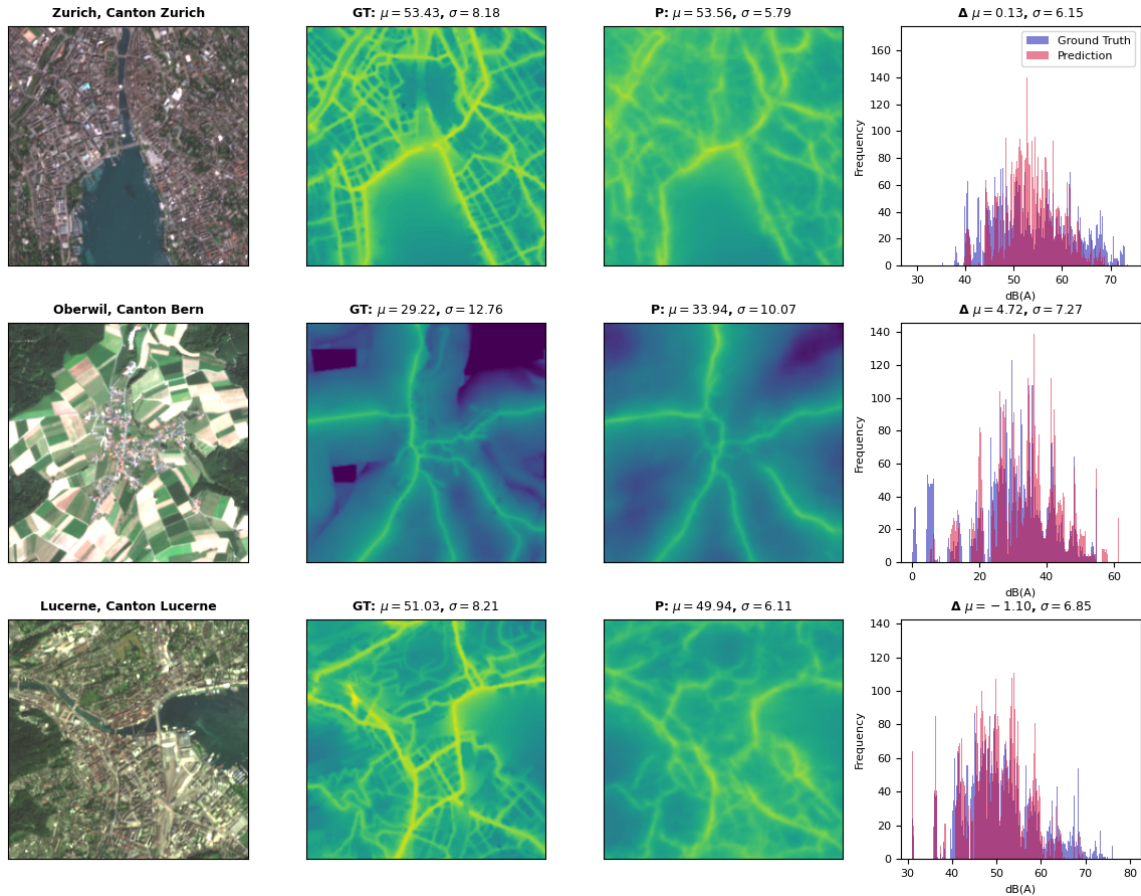


Fig. 2. Comparison of our daytime model predictions to ground-truth noise data. The columns from left to right show: a true-color image of the location, the ground-truth (GT) noise data, the noise estimation predicted (P) by our model, and the distribution of noise values from prediction (red) and ground truth (blue). The plot titles show average noise and standard deviation (columns two and three) and average difference and standard deviation between prediction and ground truth (column four) in units of dB(A).

Our approach is subject to a number of **limitations**. In contrast to the sonROAD18 physical model [4], our approach does not account for elevation information. Also, by being trained on data for Switzerland, the applicability of our models to geographically different regions might be limited. By providing more geographically diverse training data, we can improve our model to generalize well on a global scale.

6. CONCLUSIONS

Road noise mapping is traditionally conducted with physical models, requiring the labor-intensive collection of multiple input parameters. This work demonstrates the feasibility of approximating road noise from satellite imagery only.

Utilizing the U-Net architecture, our Deep Learning model achieves promising results across various geographical settings. Road noise estimation works especially well in rural areas and on highways, where roads are easier to con-

trast from their surroundings. Furthermore, our approach is capable of accurately modeling the propagation of road noise.

The dataset we used solely consists of data of Switzerland. Future studies could incorporate data from other countries, extending the applicability of the Deep Learning models to other geographies and countries, enabling road noise mapping in locations that previously lacked sufficient data.

7. REFERENCES

- [1] European Environment Agency, “Environmental noise in Europe - 2020,” <https://www.eea.europa.eu/publications/environmental-noise-in-europe>, accessed 2021-12, 2020.
- [2] E. van Kempen, M. Casas, G. Pershagen, and M. Foraster, “WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Envi-

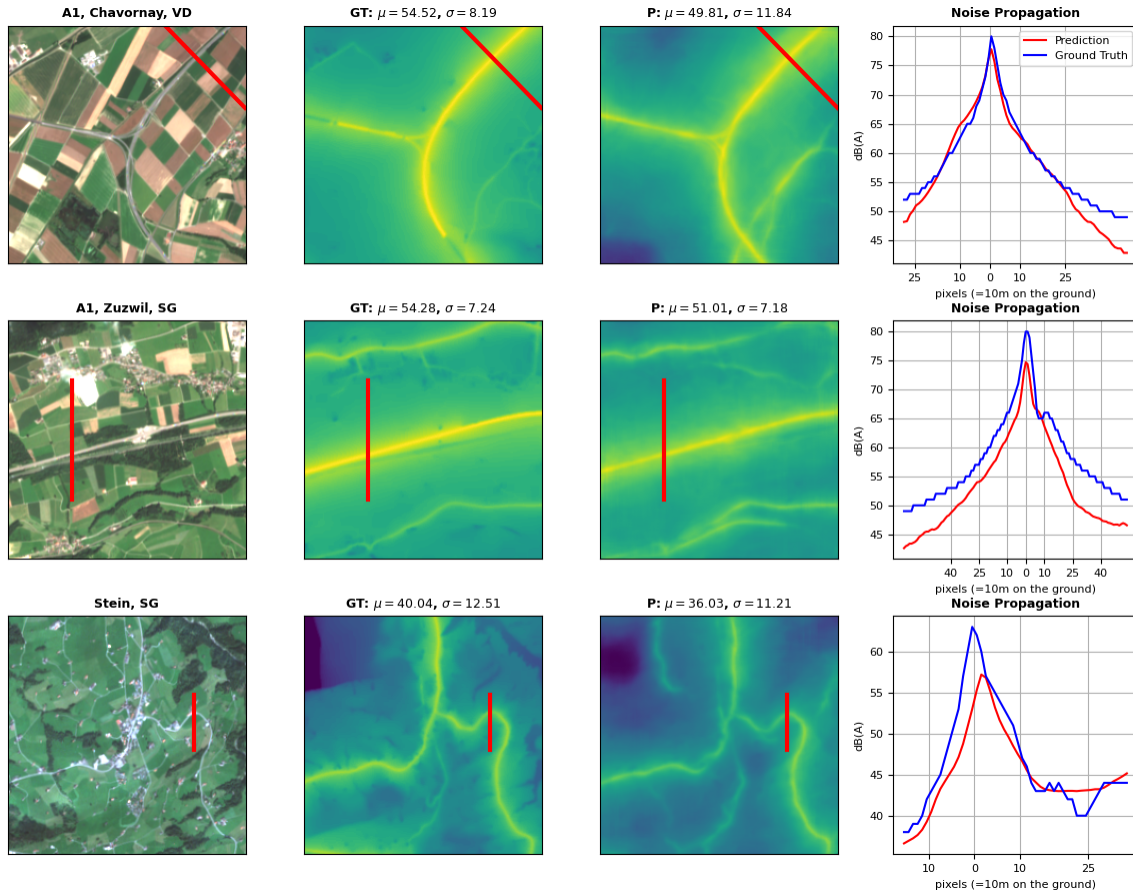


Fig. 3. Road noise cross-section profiles for three different locations. Read straight lines indicate the location of the cross-section. Columns from left to right show: a true-color image of the location, ground-truth noise data, noise estimation predicted by our model and the comparison of the noise profiles. The width of both the ground-truth and the predicted noise profiles decreases from the top location to the bottom location, which show, in this order: a three-lane freeway, a two-lane freeway, and a rural road. The plot titles of columns two and three show average noise and standard deviation.

- ronmental Noise and Cardiovascular and Metabolic Effects: A Summary,” vol. 15, no. 2, pp. 379, 2018.
- [3] E. van Kempen and W. Babisch, “The quantitative relationship between road traffic noise and hypertension,” vol. 30, no. 6, pp. 1075–1086, 2012.
- [4] K. Heutschi and B. Locher, “sonROAD18 - Berechnungsmodell für Strassenlärm,” Swiss Bundesamt für Umwelt (BAFU), 2018.
- [5] P. Kim, H. Ryu, J.-J. Jeon, and S. I. Chang, “Statistical Road-Traffic Noise Mapping Based on Elementary Urban Forms in Two Cities of South Korea,” *Sustainability*, vol. 13, no. 4, 2021.
- [6] J. Staab, A. Schady, M. Weigand, T. Lakes, and H. Taubenböck, “Predicting traffic noise using land-use regression – a scalable approach,” *Journal of Exposure Science & Environmental Epidemiology*, 2021.
- [7] G. Rey Gozalo, E. Suárez, A. L. Montenegro, J. P. Arenas, J. M. Barrigón Morillas, and D. Montes González, “Noise Estimation Using Road and Urban Features,” *Sustainability*, vol. 12, no. 21, 2020.
- [8] Fayyen Bastani, Songtao He, Sofiane Abbar, Mohammad Alizadeh, H. Balakrishnan, Sanjay Chawla, Samuel Madden, and David J. DeWitt, “RoadTracer: Automatic Extraction of Road Networks from Aerial Images,” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4720–4728, 2018.
- [9] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang, “Road Extraction by Deep Residual U-Net,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
- [10] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *CoRR*, vol. abs/1505.04597, 2015.