Using CNNs on Sentinel-2 data for noise modelling

Jeroen Staab[a, b*], Thomas Stark[a], Michael Wurm[a], Kathrin Wolf[c],
Marco Dallavalle[c, d], Arthur Schady[e], Tobia Lakes [b, f] and Hannes Taubenböck[a, g]

[a]German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany
[b]Geography Department, Humboldt-University Berlin, 10099 Berlin, Germany
[c]Institute of Epidemiology, Helmholtz Zentrum München,
German Research Center for Environmental Health (HMGU), 85764 Neuherberg, Germany
[d]Institute for Medical Information Processing, Biometry, and Epidemiology,
Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[e]Institute of Atmospheric Physics (IPA), German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany
[f]Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), 10099 Berlin, Germany
[g]Department of Urban Remote Sensing, Institute of Geography and Geology,
University of Würzburg, 97074 Würzburg, Germany

*Email: jeroen.staab@dlr.de

***Abstract*—Urbanisation and road traffic noise go hand in hand. While the WHO and the European Environmental Agency are concerned about high noise levels and the respective adverse effects on health, detailed exposure maps are scarce. Utilizing highly accurate sound propagation models is expensive and scalable Land-Use Regressions (LUR) are often limited by the lack of available training data. Also, the portfolio of statistical models of LURs so far has not been extended towards deep learning despite their recent contributions in urban remote sensing. By challenging a semantic segmentation network with the noise mapping problem, we aimed to test their capabilities. Different input channels, scoping road data, Sentinel-2 images, topographical data and a building model are compared against each other. The best performing model utilizes all eleven features available and has an overall accuracy of 0.89. Therewith a methodical cornerstone is laid. We suggest that future studies shall intensify experiments on input channels, learning strategy and spatial application.**

***Index Terms*—traffic noise, exposure mapping, deep learning, semantic segmentation**

I. Introduction

In 2018, 4.2 billion people lived in urban areas [1]. With ongoing urbanization, traffic volumes and noise levels increase. However, it often remains unclear for the WHO whether the long-term day-evening-night road traffic noise level (also referred to as L_{den})) exceeds a critical threshold of 53 dB(A) [2]. In Europe alone, an estimated 12,000 premature deaths are attributed to high noise exposure every year [3]. Despite the fact that we are aware of the negative impacts that noise may have, spatial data is scars and prone to inconsistencies [4].

This research was mainly funded by the German Federal Environmental Foundation (DBU, funding code 20017/500) and the project " Noise2NAKO^{AI}" (funding code ZT-I-PF-5-42) granted by the Helmholtz Association. Additional funds were provided through the "Digitaler Atlas 2.0" project funded by the German Aerospace Center

Engineers use sophisticated software simulating the physical mechanics of acoustics to produce accurate L_{den} noise maps. However, the large expenditure for respective mapping noise restricts their application. Therefore, the European Noise Directive (END) obliges member states to map noise only in urban agglomerations with a population of more than 100,000 people and along main highways with more than 3 million vehicle crossings per year.

A complementary approach for mapping noise over large areas are Land-Use-Regression (LUR) models. Here, data on traffic, surrounding land cover and further spatial predictors are used to train a multiple regression model. Once introduced for mapping traffic related air pollution [5] - and being used for global predictions in this context today [6] - they have soon found their way into the noise mapping domain [7]. While typically, linear- (e.g. [7]) or generalized additive models (e.g. [8]) are used, recent developments involve random forests [9] and artificial neural networks [10], too. To the best of our knowledge though, Convolutional Neural Networks (CNNs) have not been utilized in this domain yet. This may not surprise, as CNNs require large amounts of training data and so far, LURs are trained using *in-situ* measurements. Targeting the subproblem, that such *in-situ* measurements also do not reflect source specific L_{den} values as required by the END, we previously demonstrated that European noise maps can be used as training data for LURs [11].

In the field of earth observation and especially urban remote sensing, CNNs are recently en vogue [12]–[15]. Inspired by their capabilities to cope with diverse problems, in this study, we want to test CNNs for noise mapping. Due to similarities of the problem statements, we propose to utilize an image data stack containing information on traffic roads and the surrounding environment to the network. Building upon [11]'s findings, we use European noise maps of 70 German cities as reference. They cover approximately 12,250 km^2, are ordinally

scaled and show urban noise levels above 55dB(A) with 5 dB(A) increments. Choosing a semantic segmentation model, our study aims at exemplifying the suitability of CNNs for noise mapping.

II. Data and Methods

A. Area of Interest

As stated above, in Europe, larger urban agglomerations ($\geqq$ 100k population) are obligated to report noise above a critical threshold of 55 dB(A). With respect to Germany, this concerns 70 cities, having a population ranging from 3,613,495 (Berlin) to 103,949 (Moers). Overall, 75% of them have a population below 330,786.

Working at a 10 meter resolution, we generated a grid of 96x96 pixels covering urban areas of Germany. Each grid cell - further also referred to as tile - must contain at least 80% reference data. In total, 11.888 tiles are available for our experimental setup.

B. Reference Noise Data

The strategic noise maps in Europe are produced using highly accurate engineering simulations [16] and depict a yearly averaged noise level indicator with a penalty for evening and nights known as L_{den}. According to the END, only critical levels above 55 dB(A) are mapped with 5 dB(A) increments. The individual communities are free to choose an output format, but vectorized choropleth maps are most common [17]. We downloaded the most recent road traffic noise data available (*id est* 2017) and harmonized it by rasterizing it to 10x10 meter pixels.

C. Input Data

We consider the presence of noise as a function of both - noise emission and sound propagation. Correspondingly, we curated eleven input features (see Figure 1) covering both aspects. After laying them out below, they are grouped into three experimental feature sets within the next section.

Starting on the emission side - European noise maps are source type-specific such that our reference shows road traffic noise only - roads are an important asset first of all. We retrieved *OpenStreetMap* data for our reference year 2017. With respect to different road types emitting different magnitudes of noise (as a function of speed limit, traffic volume and fraction heavy vehicles), six road types *motorway*, *trunk*, *primary*, *secondary*, *tertiary* and *residential* are differentiated and are used as respective proxy [11]. In order to inform the network about roads outside of the tile as well, we used *gdal_proximity* [18] to calculate the distance to the nearest road of each type for each pixel.

Propagating soundwaves interact with the ground. While soft and porous materials like vegetation have sound absorbing effects [19], plain and solid surfaces help noise traveling over long distances. Here, we rely on a remote-sensing-derived Sentinel-2 (Level 1C) median mosaic which [20] previously used as input as well. Scoping the year 2017, the multi-temporal image collection was aggregated to the median pixel value for bands *2*, *3*, *5* and *8*.

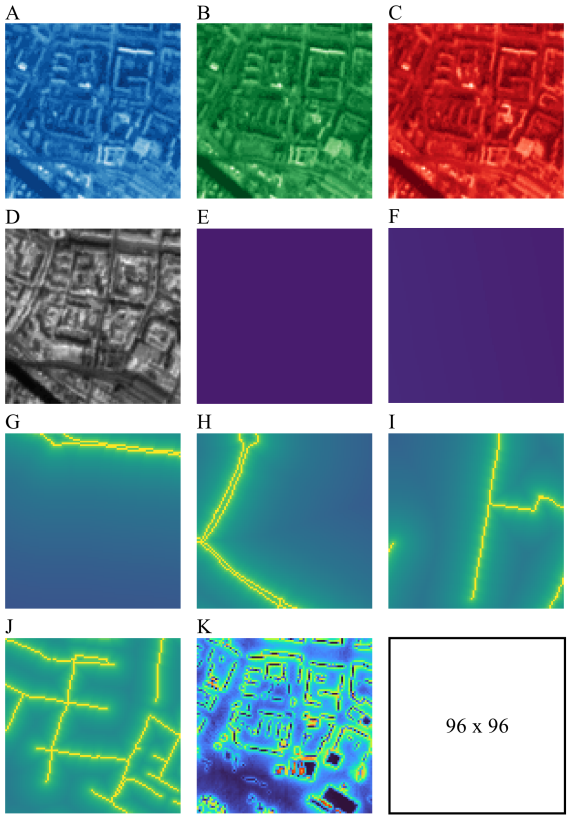

Fig. 1. Representation of all input features for one exemplary tile (t502082). A-D showing Sentinel-2 channels *2*, *3*, *5*, and *8*. E-J *OpenStreetMaps* roads types *motorway*, *trunk*, *primary*, *secondary*, *tertiary* and *residential*. K depicts Sky-View-Factor derived from land surface model.

Last, propagating sound waves also interact with topography and buildings. In an urban context, [21] for example, stressed the unfavorable effects of dense street canyons. Therefore, we summed up Copernicus DEM (30 meter resolution) and a building model (LoD1) (rasterized to 10 meters) to a surface model and derived the sky-view-factor (SVF) using the *horizon R* package [22]. The maximum search radius for obstacles is 1000 meters. Low values close to 0 (colored red in Figure 1) depressed locations with less air volume available and many high barriers around, while values close to 1 depict pixels (colored dark blue in Figure 1) allowing free sound propagation into all directions.

Prior handing over the data to our experimental setup, all input features were scaled between 0 (global minimum) and 255 (global maximum). Additionally, the global mean and standard deviation were were computed for centering the data (also referred to as whitening).

III. Experimental Setup

Testing the capabilities of the CNN, we defined three kinds of input channels: 1) *Roads* includes proximities to the

six different road types only; complementary 2) *RGB* only includes the three Sentinel-2 bands *2*, *3* and *5*; As well as 3) *All Features* embracing all eleven features as presented above.

We chose Resnet50 [23] as an encoder, because of its performance and efficiency, and for the decoder we chose the Unet [24] architecture. The model consists of 32.53 million trainable parameters.

The toolbox of [25] was altered to the needs of our objectives, e.g., for the number of potential input channels to permit a data input of three to eleven input channels. The models are trained on an NVIDIA Quadro RTX 4000 using a batch size of 128 images and the Adam optimizer. The training period was set at 100 epochs. As loss function, we used the soft cross entropy loss with label smoothing and modified class weights, where the first class weight is decreased to .1 and all other classes stay at 1. During the training procedure, an exponentially decreasing learning rate with a factor of 0.5 is applied. The learning rate is initially set to 0.0001. Augmentations are a common technique in machine learning to increase the amount of available training data [26]. Analogous to this, we applied the following augmentations to our image tiles: horizontal and vertical flip, random crop, and, sharpen.

To test the resulting model on unseen data, the dataset is randomly split into 80% training and 20% testing. The quantitative assessment of the performance of the different input channel combinations is evaluated based on various types of evaluation criteria. We present the Overall accuracy (OA), F1 score, Precision, Recall, and Intersection over Union (IoU; also known as the Jaccard Index).

IV. Results and Discussion

A. Performance and Efficiency

Table I shows that the *Roads* data set alone, helped explaining large proportions of the variance. Vice versa, using the three *RGB* channels of Sentinel-2 only, no satisfactory scores were achieved. The combination of all features available though - road proximites in conjunction with information on the surrounding built-up and topographical environment does increase all accuracy measures to a maximum (highlighted bold). This is particularly visible for Precision, Recall and its harmonic mean, the F1 score.

TABLE I
Results for three input channel combinations for the overall accuracy, F1 score, Precision, Recall, and the IoU. The highest metrics of the three input channels is bold.

	Overall Accuracy	F1 score	Precision	Recall	IoU
Roads	0.8863	0.6582	0.6592	0.6571	0.4247
RGB	0.8102	0.4019	0.4153	0.3910	0.2251
All Features	**0.8930**	**0.6781**	**0.6791**	**0.6771**	**0.4387**

Also, in the context of green AI though, it is very interesting to mention that training with *Roads* did not further improve from the 57th epoch on, while *RGB* and *All Features* was only interrupted after 91 and 96 epochs respectively.

With respect to the uncertainties, we want to stress potential inconsistencies in the reference data though. About 3% of our input tiles contain reference data from more than one city although the European noise maps may be produced by different planning bureaus. As a consequence, the legal degrees of freedom defined within the European Good-Practice Guide [27] - e.g. guessing building height where no LoD1 model is available - might have been applied. Usually such inconsistencies are only visible along shared borders when comparing the produced maps against each other [4]. While it is impossible to tell from a distance, which city's reference data is the most accurate, our experimental setup comprises these federal inconsistencies.

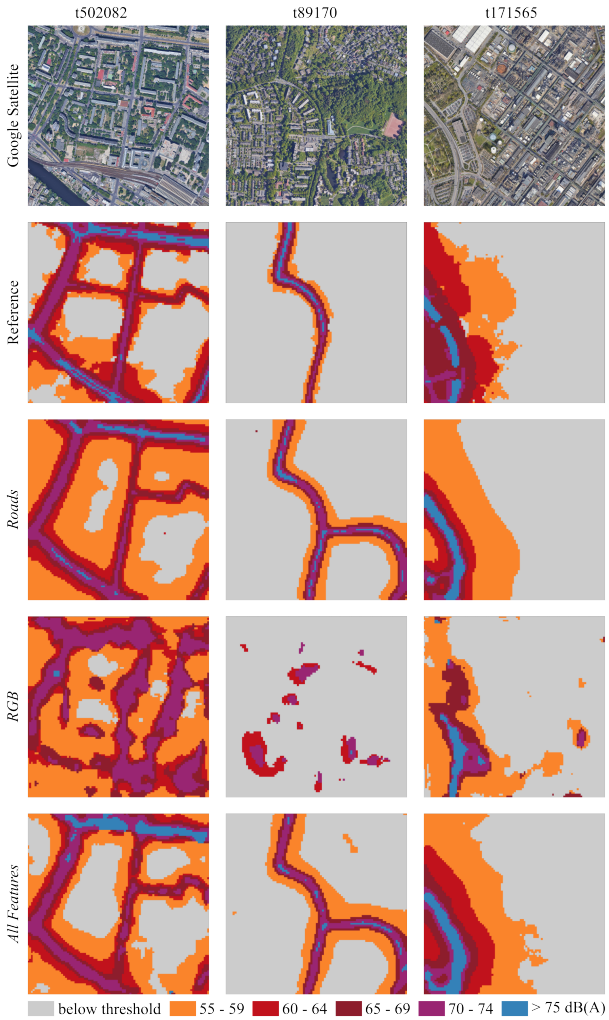

Fig. 2. Predictions produced with our experimental setup for three complementary examples. The first row shows high resolution imagery (Google Satellite) supporting interpretation of results. Colored according to DIN 18005-2, the second row depicts the reference data, while the rows below presents predictions made with models trained using *Roads*, *RGB*, and *All Features* as input channels.

B. Visual Assessment

Comparing the predictions in a visual manner (Figure 2), two interesting observation can be made; First, not all noise emitting roads are covered with a prediction. This can most prominently be seen at a horizontal road going west in t502082. Vice versa, not all roads included in *OpenStreetMap* do emit critical noise levels. Exemplified in t89170, there is a road going from north to south, with a junction in the middle of the tile. It appears logical, that the vehicle count will split up into both roads which is why the noise levels decrease in the reference data. Predictions though based on *Roads* only do not consider this context. Whereas the predictions based on 10x10 meter Sentinel-2 *RGB* pixels only do not lead to critical noise levels above 55dB(A) at all.

Second, we want to point out the capabilities of the CNN to consider sound barriers. On a free field, propagating sound pressure is only damped at a logarithmic scale. This effect, also known as geometric attenuation is very well visible at the *Roads* prediction in t171565. The reference though does show significant artefacts in the north east and lower south parts of the image. In comparison with the input data, intensively built-up areas do block the sound here. Although the *RGB* data itself does only produce poor results, in conjunction with *All Features*, at least some artefacts become visible.

V. Conclusions

To the best of our knowledge, this is the first time a semantic segmentation approach has been used for noise mapping. Replacing the machine learning method of conventional LUR models with fully CNNs allow the utilization of ordinal scaled noise maps - most common in Europe [17] - as training reference. Therewith, the existing proof of concept [11] is expanded towards larger data pools. While still more research is required with respect to chosen architecture, hyper parameter tuning and selecting the most appropriate input channels, the first results are promising. Future studies may scale this approach for national noise exposure mappings and beyond.

References

[1] U. N. D. of Economic and S. Affairs, "2018 revision of world urbanization prospects," 2018.

[2] Weltgesundheitsorganisation and Regionalbüro für Europa, *Environmental noise guidelines for the European Region*. 2018. OCLC: 1059293643.

[3] European Environment Agency, "Environmental noise in Europe - 2020," no. 22, p. 104, 2020.

[4] J. Staab, M. Weigand, A. Schady, M. Wurm, T. Lakes, and H. Taubenböck, "Europäische Lärmkarten – Methodik und Bewertung im Kontext überregionaler Umweltgerechtigkeitsstudien," vol. IPP-Schriften of *17*, (Bremen), 2020.

[5] D. J. Briggs, S. Collins, P. Elliott, P. Fischer, S. Kingham, E. Lebret, K. Pryl, H. Van Reeuwijk, K. Smallbone, and A. Van Der Veen, "Mapping urban air pollution using GIS: a regression-based approach," *International Journal of Geographical Information Science*, vol. 11, pp. 699–718, Oct. 1997.

[6] A. Larkin, J. A. Geddes, R. V. Martin, Q. Xiao, Y. Liu, J. D. Marshall, M. Brauer, and P. Hystad, "Global land use regression model for nitrogen dioxide air pollution," *Environmental science & technology*, vol. 51, no. 12, pp. 6957–6964, 2017.

[7] D. Xie, Y. Liu, and J. Chen, "Mapping Urban Environmental Noise: A Land Use Regression Method," *Environmental Science & Technology*, vol. 45, pp. 7358–7364, Sept. 2011.

[8] S. Goudreau, C. Plante, M. Fournier, A. Brand, Y. Roche, and A. Smargiassi, "Estimation of Spatial Variations in Urban Noise Levels with a Land Use Regression Model," *Environment and Pollution*, vol. 3, p. p48, Sept. 2014.

[9] Y. Liu, S. Goudreau, T. Oiamo, D. Rainham, M. Hatzopoulou, H. Chen, H. Davies, M. Tremblay, J. Johnson, A. Bockstael, T. Leroux, and A. Smargiassi, "Comparison of land use regression and random forests models on estimating noise levels in five Canadian cities," *Environmental Pollution*, vol. 256, p. 113367, Jan. 2020.

[10] 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, p. 2365, Feb. 2021.

[11] 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*, pp. 1–12, 2022. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Primary_atype: Research Publisher: Nature Publishing Group.

[12] T. Stark, M. Wurm, X. X. Zhu, and H. Taubenböck, "Satellite-based mapping of urban poverty with transfer-learned slum morphologies," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5251–5263, 2020.

[13] V. Lalitha and B. Latha, "A review on remote sensing imagery augmentation using deep learning," *Materials Today: Proceedings*, vol. 62, pp. 4772–4778, 2022.

[14] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 59–69, 2019.

[15] M. Helleis, M. Wieland, C. Krullikowski, S. Martinis, and S. Plank, "Sentinel-1-based water and flood mapping: Benchmarking convolutional neural networks against an operational rule-based processing chain," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2023–2036, 2022.

[16] N. Garg and S. Maji, "A critical review of principal traffic noise models: Strategies and implications," *Environmental Impact Assessment Review*, vol. 46, pp. 68–81, 2014. 00042.

[17] S. Khomenko, M. Cirach, J. Barrera-Gómez, E. Pereira-Barboza, T. Iungman, N. Mueller, M. Foraster, C. Tonne, M. Thondoo, C. Jephcote, J. Gulliver, J. Woodcock, and M. Nieuwenhuijsen, "Impact of road traffic noise on annoyance and preventable mortality in european cities: A health impact assessment," *Environment International*, vol. 162, p. 107160, Apr 2022.

[18] GDAL/OGR contributors, *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation, 2022.

[19] D. Aylor, "Noise reduction by vegetation and ground," *The Journal of the Acoustical Society of America*, vol. 51, no. 1B, p. 197–205, 1972.

[20] M. Weigand, J. Staab, M. Wurm, and H. Taubenböck, "Spatial and semantic effects of lucas samples on fully automated land use/land cover classification in high-resolution sentinel-2 data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, p. 102065, Jun 2020.

[21] K. Heutschi, "A simple method to evaluate the increase of traffic noise emission level due to buildings, for a long straight street," *Applied Acoustics*, vol. 44, no. 3, p. 259–274, 1995.

[22] J. Van doninck, "Horizon search algorithm," Jul 2018.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[25] P. Iakubovskii, "Segmentation models pytorch." https://github.com/qubvel/segmentation_models.pytorch, 2019.

[26] D. Stiller, T. Stark, M. Wurm, S. Dech, and H. Taubenböck, "Large-scale building extraction in very high-resolution aerial imagery using mask r-CNN," in *2019 Joint Urban Remote Sensing Event (JURSE)*, IEEE, 2019.

[27] WG-AEN, *Good Practice Guide for Strategic Noise Mapping and the Production of Associated Data on Noise Exposure*. 2 ed., Aug. 2007.