ESTIMATION OF POWER GENERATION AND CO_2 EMISSIONS USING SATELLITE IMAGERY

Joëlle Hanna*, Michael Mommert, Damian Borth

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

KEY WORDS: CO_2 estimation, power generation, deep learning, remote sensing

Abstract Burning fossil fuels produces large amounts of carbon dioxide (CO_2), a major Greenhouse Gas (GHG) and a main driver of Climate Change. Quantification of GHG emissions related to power plants is crucial for accurate predictions of climate effects and for achieving a successful energy transition (from fossil-fuel to carbon-free energy). The reporting of such emissions is only required in some countries, resulting in insufficient global coverage. In this work, we propose an end-to-end method to predict power generation rates for fossil fuel power plants from satellite images based on which we estimate GHG emission rates. We present a multitask deep learning approach able to simultaneously predict: (i) the pixel-area covered by plumes from a single satellite image of a power plant, (ii) the type of fired fuel, and (iii) the power generation rate. To ensure physically realistic predictions from our model we account for environmental conditions. We then convert the predicted power generation rate into estimates for the rate at which CO_2 is being emitted, using fuel-dependent conversion factors.

1. INTRODUCTION

Climate change is affecting the entire world, with some regions experiencing extreme weather and rainfall events, and others more prone to severe heat waves and unprecedented droughts. Those environmental episodes are primarily caused by excessive greenhouse gas emissions, which trap heat in the atmosphere and the oceans. Moreover, power plants are usually responsible for most of these emissions, which are primarily produced by burning fossil fuels. Subsequently, a successful energy transition seems to be the only way to save our planet. For this, proper quantification of GHG emissions from individual industrial sites is necessary, but this usually requires the use of dedicated measuring devices that report detailed emissions information, as may be required by environmental protection guidelines. Such data can be used to enforce environmental protection regulations or pollution certificate trading schemes. Unfortunately, reporting requirements differ from country to country, resulting in very heterogeneous coverage of GHG emissions reporting globally.

In this work, we propose a multi-task learning approach to primarily predict the power generation rate from a satellite image of a power plant (treated as a regression problem), together with the type of fired fuel (treated as a classification problem) and plume footprint (treated as a segmentation problem) all from a single satellite image centered on a power plant. Using the predicted power generation rate and the type of fired fuel, we estimate the amount of CO_2 emitted at a given time. Additionally, to account for environmental conditions that may affect the visibility and extent of plumes, we incorporate external weather data into our model, leading to physically realistic values. By using a multitask approach with a shared backbone, we were able to use labeled data, which can be difficult to obtain, more efficiently, reducing the chances of overfitting. Both the predicted power generation rate and the CO_2 emission rate are of interest to various stakeholders and inform efforts to minimize the effects of Climate Change.

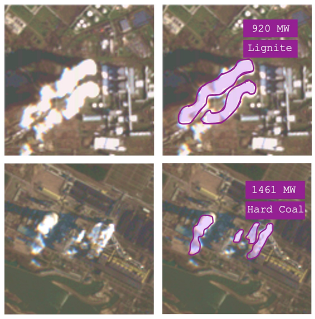

Figure 1: Two examples of our model's outputs.

2. RELATED WORK

To the best of our knowledge, little work has been done concerning the estimation of GHG emissions from fossil fuel power plants based on observational data. The estimation of the power plant generation has mostly been restricted to extracting statistics from available annual data. (1) estimate the deviation of each plant from the average generation of other similar plants based on detailed information on plant-level (such as its size, fuel-type, or installed capacity) and environmental factors using machine learning algorithms such as gradient boosting trees. Other works first predict an observable value which, in a second step, can be used as a proxy for estimating GHGs. For example, (2) infer the amount of Nitrogen Oxides (NO_x) emitted from a small number of US power plants using the Ozone Monitoring Instrument (OMI), a nadir-viewing visual and ultraviolet spectrometer, and use this value as a proxy to calculate the amount of CO_2 emis-

*Corresponding author

sions using a pre-determined relationship between NO_x and CO_2, and measurements provided by the Continuous Emissions Monitoring System (CEMS). In contrast to previous works, our approach allows us to estimate CO_2 emissions using predicted power generation rates as a proxy, based on freely and globally available satellite imagery for hundreds of power stations in Europe

3. METHOD

Dataset We compiled a dataset[1] consisting of remote imaging data of fossil fuel power plants taken by ESA's Sentinel-2 Earth-observing (EO) satellites. We manually annotate them by providing plume masks. We supplement our processed satellite images with power generation data from *ENTSOE* (3) and contemporaneous weather data (temperature, relative humidity and wind speed) from the *ERA-5 data set* (4). The installed capacity of these power plants ranges from 29 MW to 5230 MW with a mean of 1177 MW. Our sample includes power plants that use 4 different types of fuel: hard coal, gas, lignite , and peat. For each observation, our dataset contains a Sentinel-2 satellite image with its segmentation map, the type of fired fuel, the actual power generation rate at the corresponding timestamp, and weather data.

Baseline The proposed approach (5) is a multitask network including a unique shared UNet-like backbone and three distinct task heads: a regression head to predict power generation rates, a segmentation head to semantically segment plumes, and a classification head to predict the type of fuel used. In our case where the amount of labeled data is very limited, the use of a shared backbone has several advantages. The first advantage is to reduce the number of trainable parameters and, therefore, to reduce the risk of overfitting on the small training set. The second advantage is to push the network towards learning a robust representation that generalizes well to different related tasks. Knowing that there is a direct correlation between CO_2 emission rates and power generation rates, our idea is to derive a fuel-dependant conversion factor, which would allow us to estimate the rate of CO_2 emissions by directly using the rate of power generation produced by the power plants and the type of fuel used.

Improvements Although the multitask model gave reasonable results, some improvements were needed (6). Since we rely solely on the visual characteristics of plumes to predict power generation, we naturally have to take into account any environmental conditions that can affect the shape or opacity of the plumes, such as wind or humidity. We suggest using this information as input to the model and as an additional penalty in the loss function. Moreover, in a multitask approach, not all tasks are equal in terms of learning speed and performance. Since treating them in this way would be rather inefficient, we propose to directly follow the work of (7) and therefore to prioritize the difficult tasks during the training process, in a self-paced learning scheme.

[1]The dataset is available at `https://doi.org/10.5281/zenodo.5644746`

4. RESULTS

Our multitask model achieves the best average performance in the multi-spectral setup, and surpasses the segmentation baseline by a relative increase of 8%, and the classification baseline by 3%. It also surpasses the regression baseline by 22% in terms of MAE and 20% for the R^2. Moreover, we notice that taking into account environmental conditions in the loss function improves the performances, especially for the regression task, where we observe relative improvements of 4% for MAE and the R^2.

5. CONCLUSION

In this work we show the possibility to predict power plant emissions from single Sentinel-2 satellite images. Our proposed deep multitask architecture was trained on a combination of three tasks and experiments confirmed that auxiliary tasks can indeed boost the network performance. The free availability of EO data enables the application of such an approach on a global scale and specifically in areas where no emission estimates are available.

REFERENCES

T. Yin, L. Byers, L. Valeri, and J. Friedrich, "Estimating power plant generation in the global power plant database," 2020.

F. Liu, B. N. Duncan, N. A. Krotkov, L. N. Lamsal, S. Beirle, D. Griffin, C. A. McLinden, D. L. Goldberg, and Z. Lu, "A methodology to constrain carbon dioxide emissions from coal-fired power plants using satellite observations of co-emitted nitrogen dioxide," *Atmospheric Chemistry and Physics*, vol. 20, pp. 99–116, 2019.

Entsoe, "Actual generation per generation unit table," 2021, https://transparency.entsoe.eu/generation/r2/.

H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, M. S. J, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thépaut, "Era5 hourly data on pressure levels from 1979 to present. copernicus climate change service (c3s) climate data store (cds)," 2020, 10.24381/cds.bd0915c6.

J. Hanna, M. Mommert, L. M. Scheibenreif, and D. Borth, "Multitask learning for estimating power plant greenhouse gas emissions from satellite imagery," in *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021.

J. Hanna, M. Mommert, and D. Borth, "Physics-guided multitask learning for estimating power generation and co_2 emissions from satellite imagery." submitted, in review.

M. Guo, A. Haque, D.-A. Huang, S. Yeung, and L. Fei-Fei, "Dynamic task prioritization for multitask learning," in *ECCV (16)*, 2018, pp. 282–299.