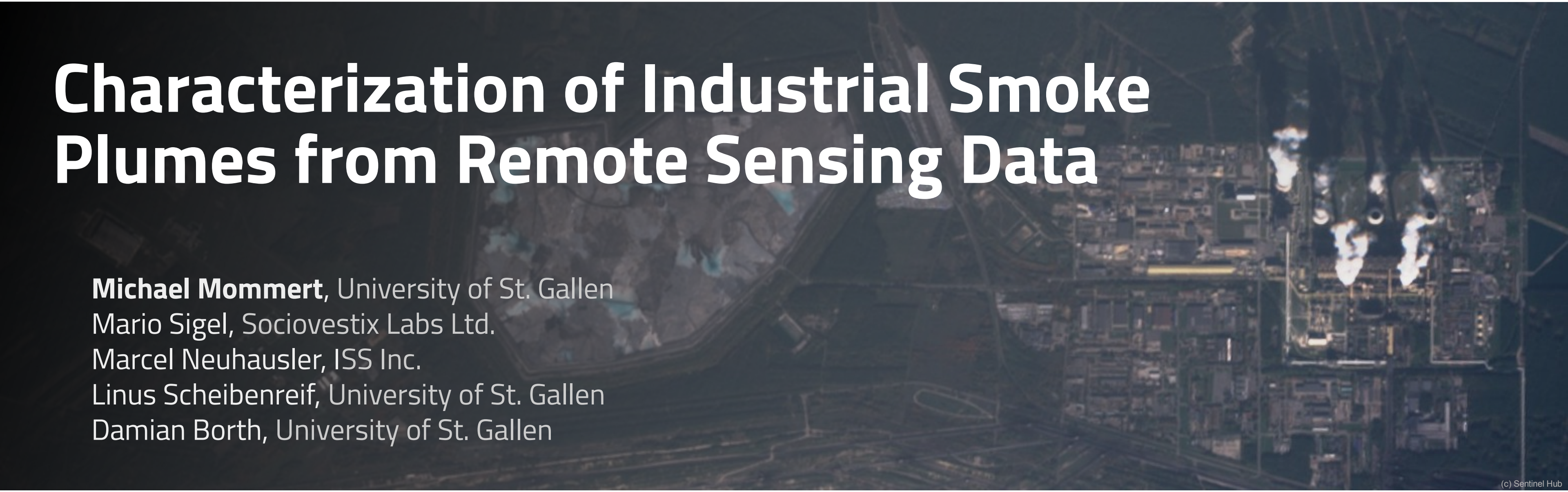
Universität St.Gallen

Characterization of Industrial Smoke Plumes from Remote Sensing Data

Michael Mommert, University of St. Gallen
Mario Sigel, Sociovestix Labs Ltd.
Marcel Neuhausler, ISS Inc.
Linus Scheibenreif, University of St. Gallen
Damian Borth, University of St. Gallen



Tackling Climate Change with Machine Learning workshop at NeurIPS 2020
December 11, 2020

"From insight to impact"

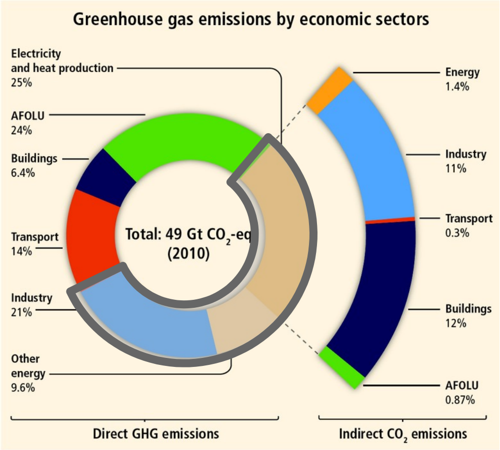
Introduction

The main cause of the currently observed climate change and global warming lies in the excessive emission of greenhouse gases (GHG) through the **combustion of fossil fuels**. The reduction of GHG emissions, e.g., based on emission quotas or emission trading schemes, is mandatory to limit long-term damage to Earth's climate and biosphere.

The majority of GHG emissions are caused by the **power-generating and industrial economic sectors** [1], making them a worthwhile target for emission restriction quotas.

The quantitative measurement of emissions is costly and thus not required in all countries. An independent way to estimate emissions would be useful to

- fully understand the impact of GHG emissions on the climate, and
- to enforce environmental protection emission quotas and emission trading schemes.

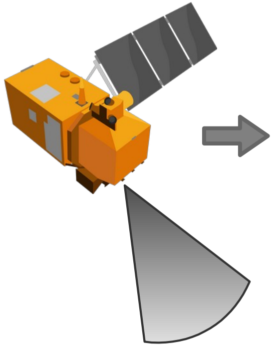


Total anthropogenic greenhouse gas (GHG) emissions (gigatonne of CO2-equivalent per year) from economic sectors in 2010. The energy and industrial sectors (gray highlighting) amount to more than 55% of the total GHG emissions. Source: [1]

References: [1] - IPCC Climate Change 2014 Synthesis Report.

Universität St.Gallen

Introduction II: Remote Sensing



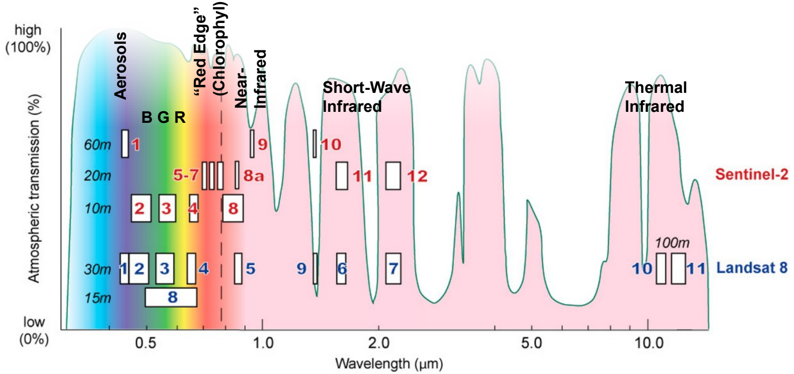
Remote sensing satellites are able to detect smoke plumes as shown in this example for the Kehrichtheizkraftwerk St. Gallen.

Earth-observing satellites are able to detect industrial smoke plumes, pending an unobstructed view (clear skies) and activity of the emitter.

In combination with weather data, satellite imagery of **smoke plumes can serve as a proxy for emissions** from a given emitter.

The detection of smoke plumes is possible through their **spectral characteristics**, which has been exploited in the past for the identification of wild fires (see, e.g., [3]).

Comparison of the spectral coverage of the Sentinel-2 and Landsat 8 Earth-observation satellites. Most bands are placed in wavelength ranges in which the atmosphere is transparent (curve). Important wavelength ranges are labeled. Plot adopted from [2].



References: [2] - Kääb et al. 2016, Remote Sens., 8(7), 598; [3] - Jain et al. 2020, ArXiv abs/2003.00646.

Universität St.Gallen

Objectives

The contributions of this project to climate change monitoring is threefold:

1) we collect a **large scale annotated data set of active industrial sites** with additional segmentation masks for a subset of these smoke plumes,

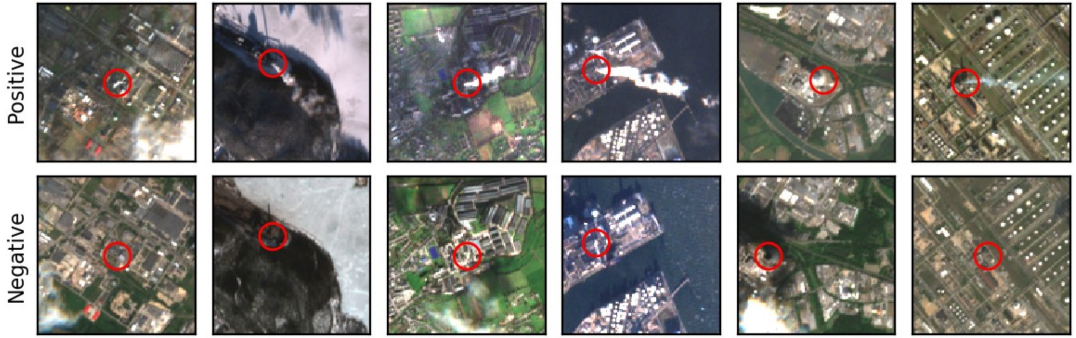
2) we present a modified ResNet-50 approach able to **detect active smoke plumes** with an accuracy of 94.3% and finally,

3) we utilize a U-Net approach to **segment smoke plumes and measure the areal projection** of smoke plumes on average within 5.6% of human manual annotations.

Data

We extract **satellite imagery of industrial sites** obtained by the Sentinel-2 constellation.

Sites were selected based on their annual emissions as reported to the European Pollutant Release and Transfer Register, resulting in a total of 624 locations.

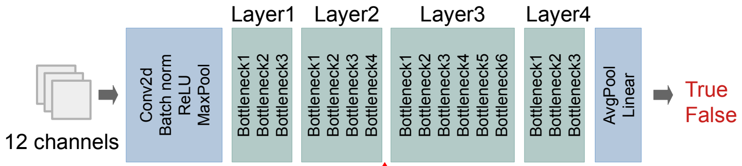


Example images from our image data set. Each column corresponds to a different location; the top row shows locations when a smoke plume is present (positive class), the bottom row shows the same locations during the absence of smoke (negative class). Red circles indicate the approximate origin of the plume.

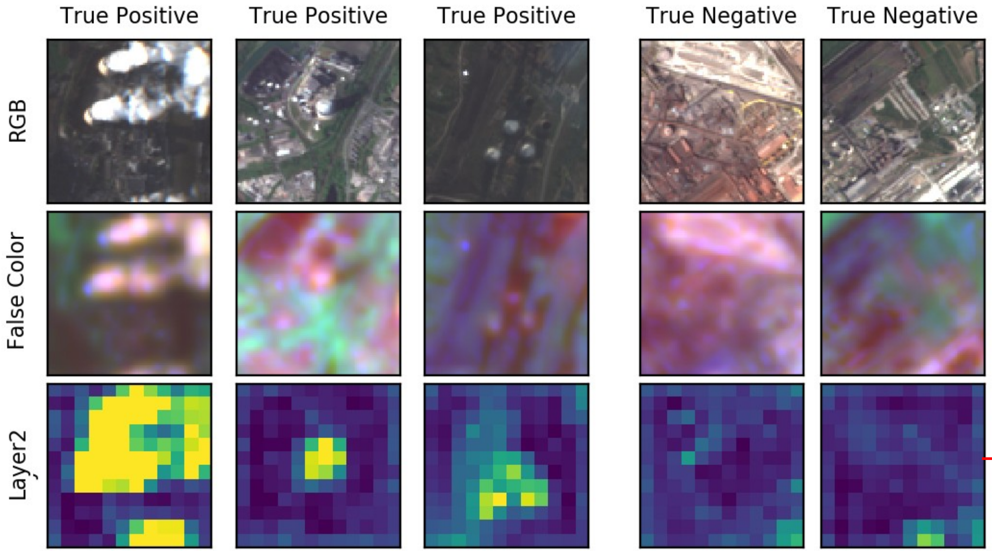
Rendering of one out of two Sentinel-2 satellites run by ESA's Copernicus program. Source: ESA

Level-2A (bottom-of-atmosphere, 12 bands) products for these sites that were acquired during 2019 were downloaded, rescaled, and cropped appropriately to square patches with an edge length of 120 pixels (1.2 km). This data set includes a total of 21,350 images.

We manually labeled all images based on the presence (3,750 *positives*) or absence (17,600 *negatives*) of a smoke plume. We furthermore obtained segmentation labels for 1,437 positive images that mark the location and spread of smoke plumes.



Classification of Smoke Plumes



For different examples (columns), we show the true color RGB image (top row), a false color image (R: aerosols, G: water vapor, B: short-wave infrared), and the activations of a hidden layer ("Layer2") in our ResNet implementation (bottom row, sharing the same scaling across the row).

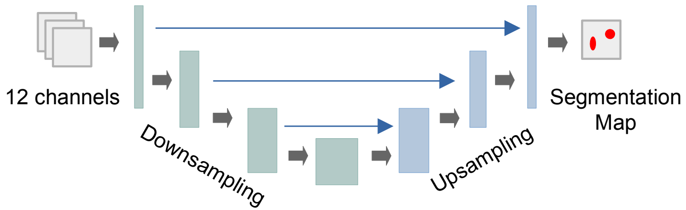
Using a **ResNet-50** convolutional neural network [4], we train a binary classifier based on the presence/absence of smoke in the image.

We identify smoke plumes with an accuracy of **94.3%**. The model successfully ignores natural clouds, but occasionally misclassifies surface features or ignores smoke plumes hidden behind cirrus clouds.

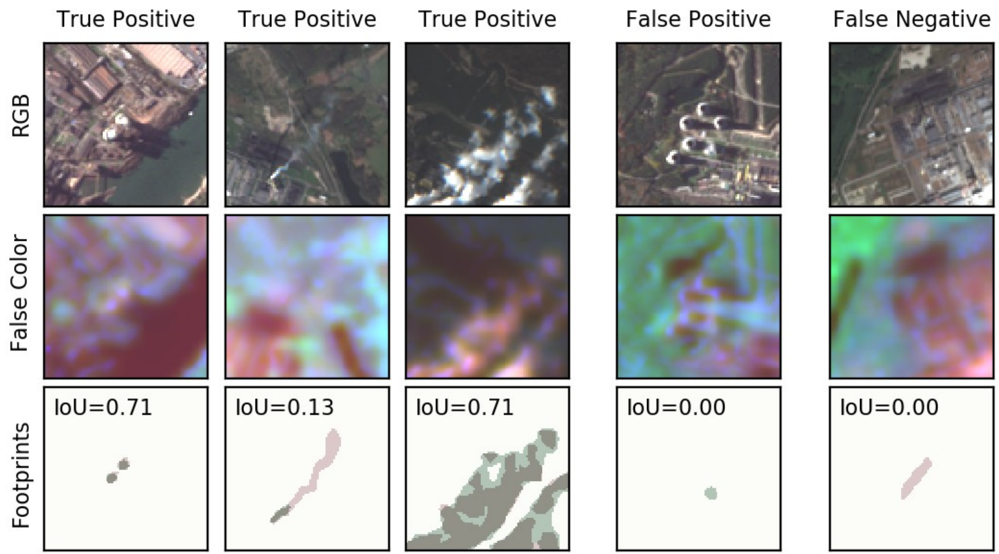
We find that activations in a hidden layer of the network are closely **correlated to smoke features** present in the images, as well as strong signals in imaging bands related to aerosols (Band 01), water vapor (Band 09), and one of the short-wave infrared bands (Band 11).

Based on this localization of smoke, we apply a segmentation model to our data.

References: [4] – He et al. 2016, CVPR, 770-778.



Segmentation of Smoke Plumes



For different examples (columns), we show the true color RGB image (top row), a false color image (R: aerosols, G: water vapor, B: short-wave infrared), and the segmentation ground-truth (red areas) and prediction (green areas); gray areas indicate an overlap between ground-truth and predicted smoke areas.

We train a **U-Net** architecture [5] to perform pixel-wise segmentation of smoke plumes.

The resulting segmentation model achieves a test sample accuracy of **94.0%** and an Intersection-over-Union score of **0.608**. We find that we can replicate manually labeled smoke plume areas within a mean accuracy of **5.6%**.

Due to the smaller size of our positive segmentation data set, the overall accuracy is slightly lower than based on our ResNet approach.

Nevertheless, the results support our approach and show that we can estimate the size of smoke with a high likelihood.

References: [5] – Ronneberger et al. 2015, MICCAI.

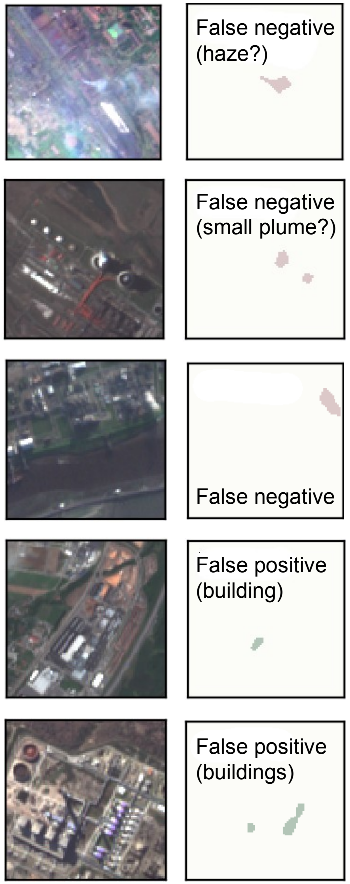
Discussion

Our results show that common convolutional neural networks are **able to identify industrial smoke plumes and their extent** with high confidence.

We find that our segmentation model, which is based on a smaller amount of training data than our classification model, is more likely to be confused by surface objects, cirrus clouds, or ground fog. These shortcomings are likely to be related to human error in the labeling process and might be solved with a more extensive training sample.

In **future work**, we will extend this work by including a regression model that correlates the predicted smoke plume area (as it results from our segmentation model) to ground-truth time-series data of the emission output from industrial sites. This approach will require the combination with time-series weather data to properly account for the appearance of smoke as a function of industrial output and environmental variables.

The goal of that work, once calibrated against a range of industrial sites, will be a framework that will allow for the **estimation of industrial emissions from satellite imaging data on a global scale**.



Segmentation mistakes and their potential causes. (red: ground-truth, green: prediction)

Universität St.Gallen

Conclusions

We investigate the possibility to train a neural network to **identify industrial smoke plumes from remote sensing image data**.

We acquire imaging data taken by the **Sentinel-2** constellation for a select sample of active industrial plants, based on emission reports by the European Union.

We are able to achieve an accuracy of **94.3%** in the binary task of identifying the presence or absence of smoke in Sentinel-2 data. We find strong hints for the **localization of smoke** in the hidden layers of this model.

Based on the localization information from our ResNet approach, we learn a segmentation task on a labeled subset of our data. We reach an accuracy for the detection of any smoke in a random image of **94.0%** and an Intersection-over-Union score of **0.608**. Finally, the segmentation model is able to reproduce the area of smoke within 94.4% of human annotating.

We are able to **identify smoke plumes with high confidence**. We will utilize this ability to establish a regression model that allows for the direct estimation of emission based on these results.

Resources

Data: zenodo

Code: github

Responsible Author:

Michael Mommert

michael.mommert@unisg.ch

mommermi.github.io