ScaDS
DRESDEN LEIPZIG

Deep Learning for Traffic and Air Pollution Prediction

Georges Alkhouri, Moritz Wilke
ScaDS Leipzig

www.scads.de

ScaDS DRESDEN LEIPZIG

Goals

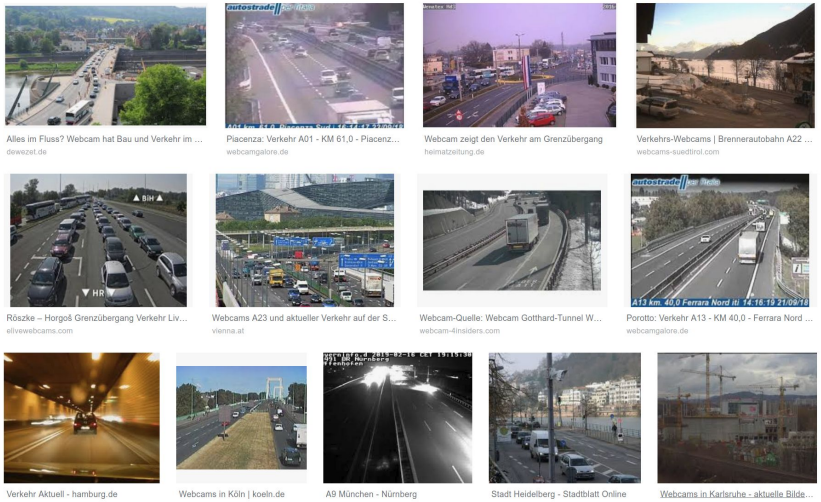
- Predict air pollution
- Investigate the influence of car traffic

Problems

- Are the sensor data reliable?
- Can we build a successful model of the complex system?
- **How can we get traffic data in a similar resolution, time and location?**

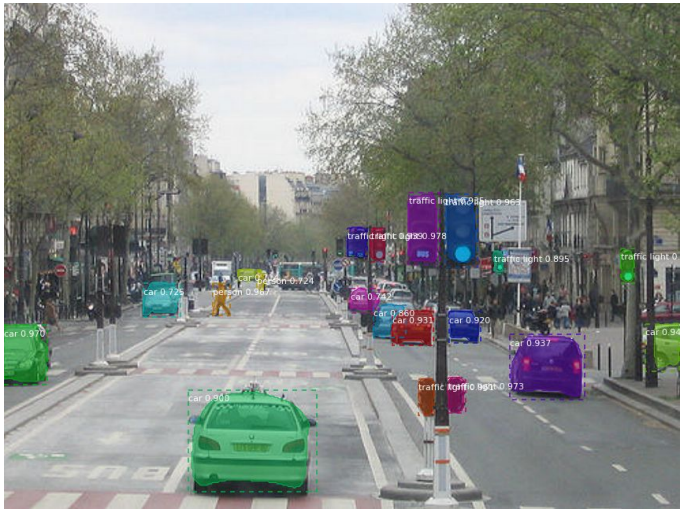
How to get traffic data?

Public webcam



+

Deep Learning for object recognition[1]



= ?

[1] https://github.com/matterport/Mask_RCNN

ScaDS DRESDEN LEIPZIG

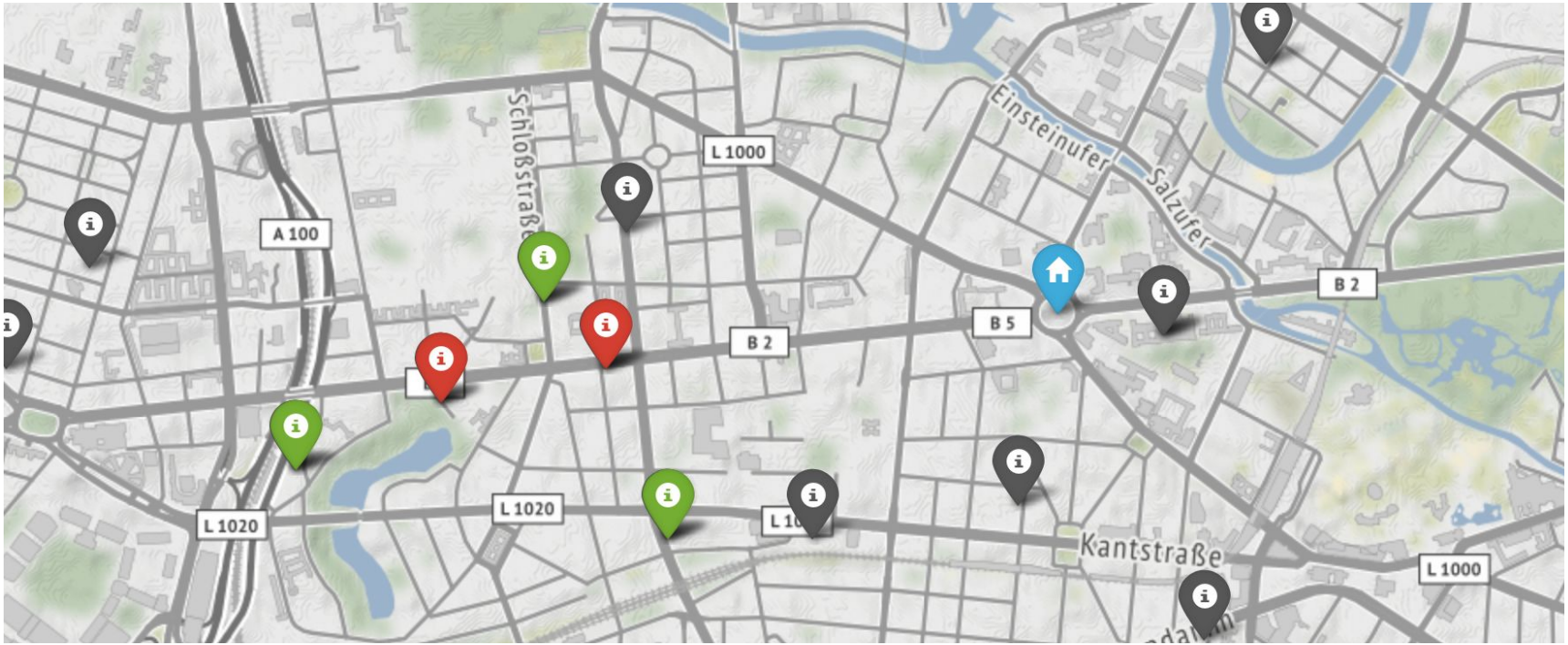
Steps

- Finding the Right Spot
- Preprocessing Sensor Data
- Building a Prediction Model
- Traffic Data
- Transfer Learning
- Results and Outlook

ScaDS
DRESDEN LEIPZIG

Finding the right spot

Finding the right spot

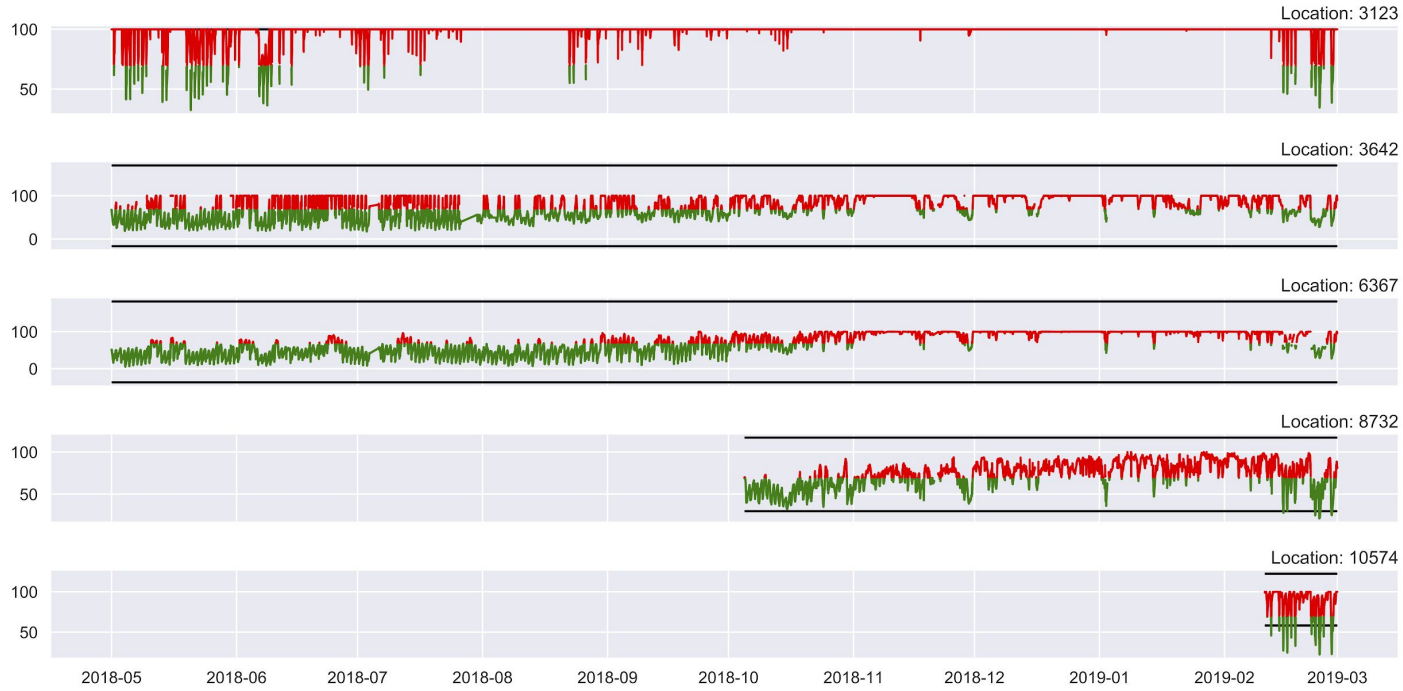


blue: camera, red: too many invalid data points, green: chosen

ScaDS
DRESDEN LEIPZIG

Data Preprocessing

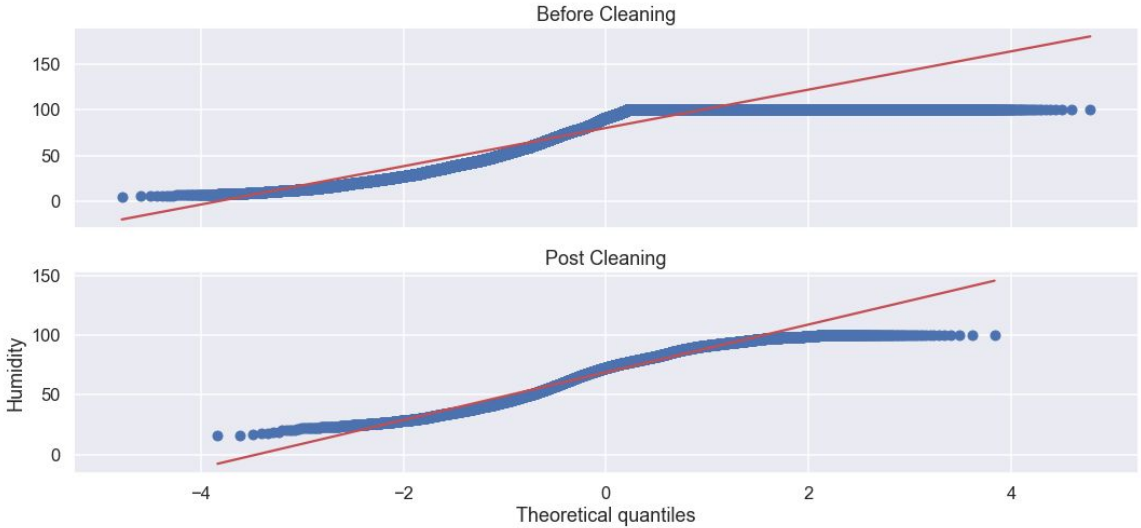
Too much humidity



Red: Not usable data points for PM measurements
Discarded locations 3123, 10574

Data Preprocessing Steps

- remove measurements where humidity > 70%
- outliers removal
- average over all (3) locations

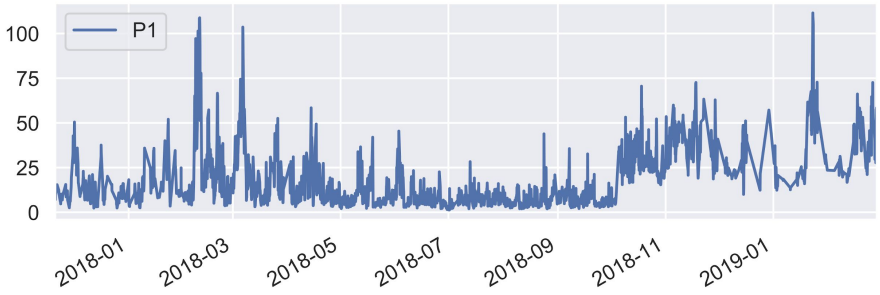


ScaDS
DRESDEN LEIPZIG

Time series prediction model

Model: learning, baseline, results

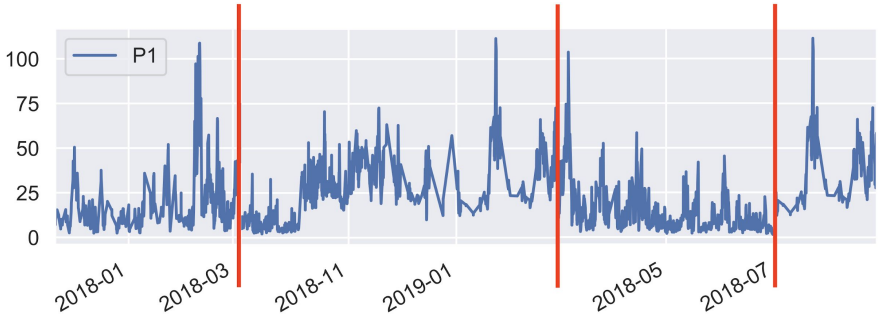
- because of possible time dependency
  → LSTM net usage



Input as time series for Model 2

Models:

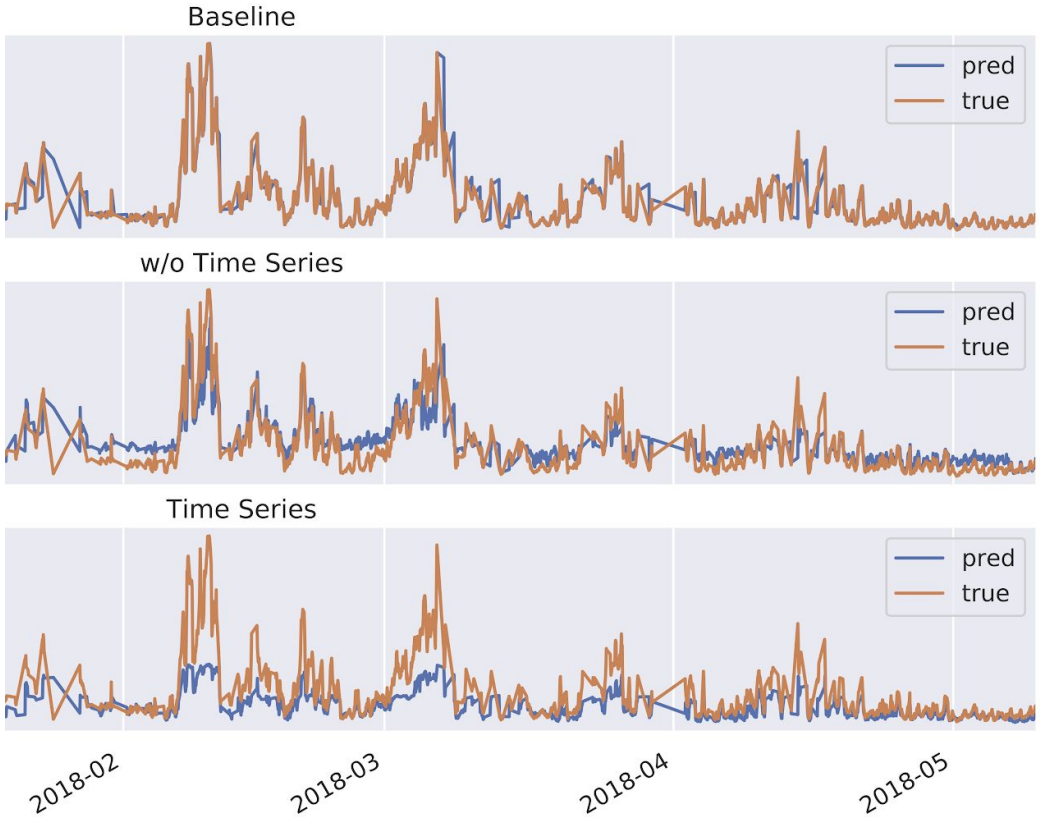
1. Baseline/Persistence Model (PM10t = PM10t-1)
2. TS Model (Stateful LSTM)
3. Normal Model



Shuffled and broken time series for Model 3

Results: Models vs. Baseline

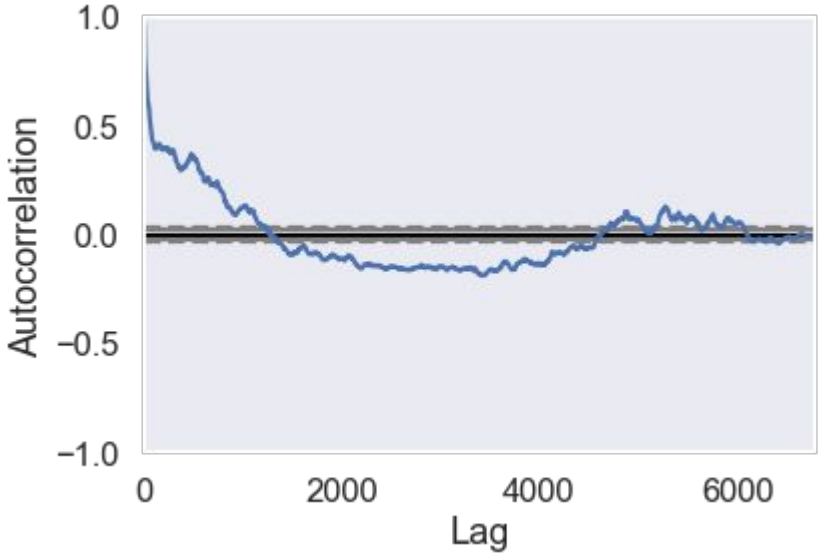
	Persistence Model	Normal Model	TS Model
MSE	**0.00447**	**0.01781**	**0.02953**
std	**0.00236**	**0.00967**	**0.01947**



Data: do we have a Random Walk?

- Persistence Model is still the best
- Temporal dependence from used features is strong

→ add more features to Model



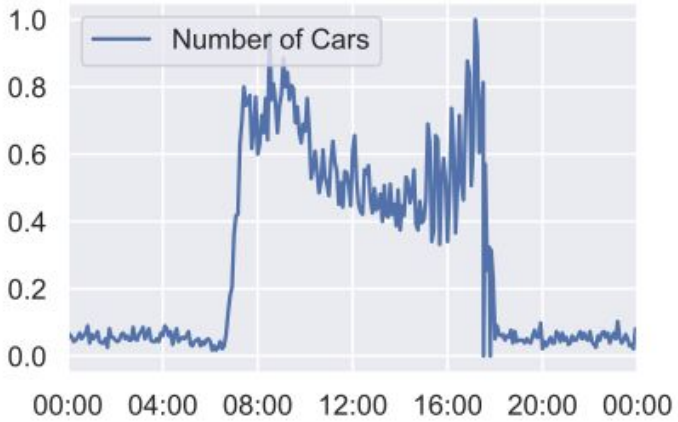
PM10 Autocorrelation

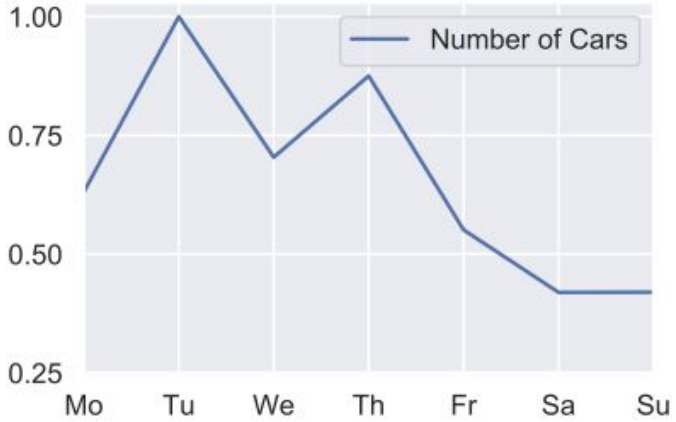
ScaDS
DRESDEN LEIPZIG

Working with webcam data

Webcam Data

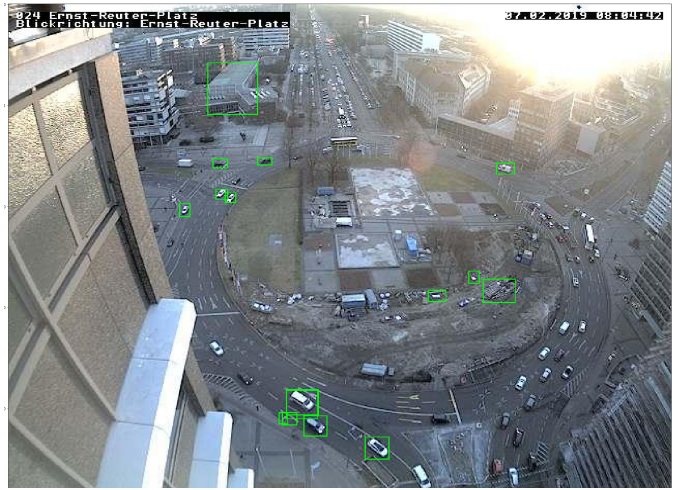
- Store an image every 5 minutes
- Process with pre-trained neural network (Mask R-CNN / COCO)
- Result: car timeseries for february 2019





Webcam Problems

- Only short timespan
- Pre-trained model not optimized on perspective and night
- No exact numbers, only magnitude



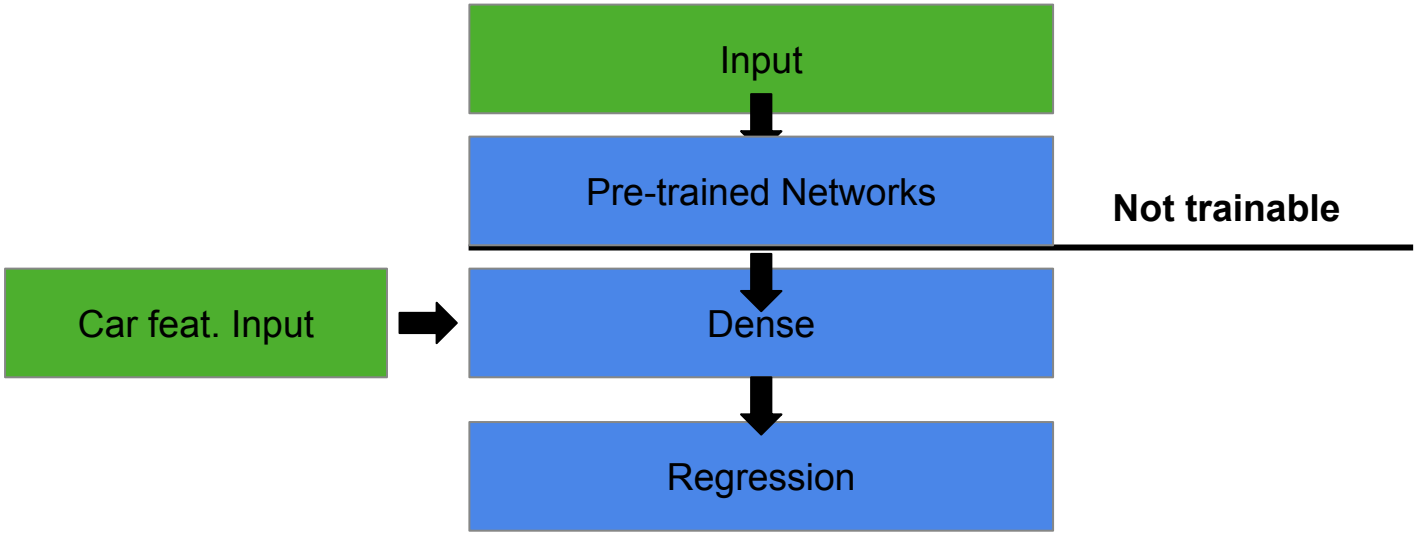
ScaDS
DRESDEN LEIPZIG

Combining both data sources

Transfer Learning

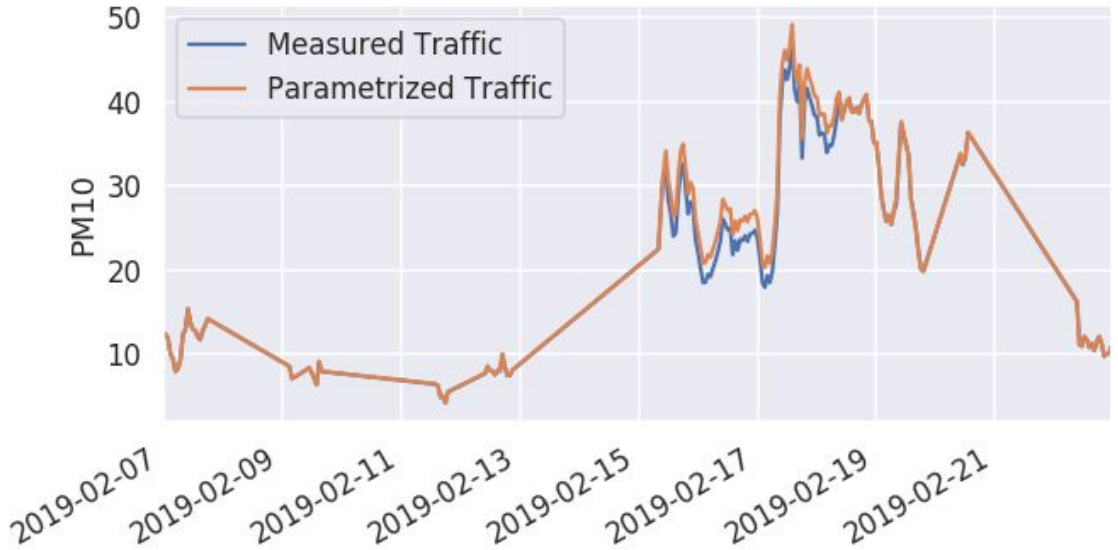
Transfer Learning

- Car Dataset is very small compared to other data
- Approach: train model first and use extracted features with car data



The Influence of Traffic

New models traffic can now be **parametrized manually** to **simulate future traffic** behaviour



Blue: PM measures for actual traffic
Orange: PM measures for simulated traffic
(higher traffic higher traffic density)

ScaDS
DRESDEN LEIPZIG

Conclusion and Outlook

Conclusion and Outlook

- Pollution time series can not easily be forecasted from the given data
- Traffic Data can be obtained from public sources
- To investigate the relationship between particulate matter pollution and traffic, different features and more data are needed

ScaDS
DRESDEN LEIPZIG

On Github

https://github.com/GeorgesAlkhouri/golddust

ScaDS
DRESDEN LEIPZIG

Thank you for your attention.

Questions?