Semi-Automated analysis of aerial images for the detection of photovoltaic solar panel

Curier R. Lyana, Ziemons H., de Jong T., Iren, D., Bromuri S.[1]

Abstract

The study main goal is to assess the effectiveness of deep learning models for the identification of solar panels in aerial images. To this end, various transfer learning approaches are evaluated to identify the most effective strategy. This contribution makes solely use of pre-trained deep learning architectures based on convolutional filters. CNNs can successfully be applied for the classification of solar panels in aerial images. The best performing model was the pre-trained version of the Xception network where all 126 layers were set to trainable. Training and testing accuracy of more than 90%. Nonetheless, we also discovered that generalization is an issue, as it does not seem to be possible to transfer the learning of Xception to other geographical regions. However, the performance of the VGG16 model during the generalization experiments that a successful application of a trained model on images from another area is possible

Key Words: Big Data, Energy Transition, Machine Learning, solar panel

1. Introduction

1.1 Context

This project takes place within the on-going energy transition taking place in Europe. This transition strives to create urban energy systems that emit less carbon and use less energy and is strongly related to Sustainable Development Goals' s 7 and 11. Current statistics on solar energy, in the Netherlands, are based on survey data from solar panels importation and the overall solar production is based on an estimate of installed capacity and production capacity per unit. This methodology solely provides national figures on a yearly basis and has an estimated uncertainty of 20 percent, while the demand is for information at regional level with shorter time scales. High resolution satellite and aerial images provide a wealth of regional statistical information, however, manual analysis of the images is tedious and time consuming. A fast semi-automatic method with a low error rate is therefore highly desirable. In this project, Statistics Netherlands (CBS) together with the Open University and in collaboration with other national statistical offices in Belgium and Germany pool their expertise to provide a complete and detailed overview of the current solar panels installation across Flanders, North Rhine-Westphalia and Limburg. The process of extracting the location of solar panels from images will be automated to produce maps of solar panels spatial distribution along with regional statistics. In practice, several machine learning approaches such as random forest(RF), support vector machine (SVM) and convolutional deep neural networks (i.e. Deep Learning, DL), are being investigated and existing registers such as information on the VAT returns from solar panels owners and information acquired from the energy providers are used to train and validate the algorithm. In this paper, the first findings from some experiments based on the Deep Learning methodology are presented for the state of North Rhine Westfalia (NRW, DE). The results presented in this document were extensively discussed in [6], and the reader is referred to it for more detailed information.

2. Datasets

The results presented in this paper solely focus on the suitability of aerial images. Two datasets were used:

1. Aerial images for *North Rhine Westfalia Data* from OpenGeodata.NRW [4] are available at 20x20 cm. Available images for four different cities with high numbers of registered solar panels and different urban landscapes were

---

[1] Statistics Netherlands and Business Intelligence and Smart Services Institute, The Netherlands (rll.curier@cbs.nl);

chosen Aachen, Münster, Düsseldorf and Hopsten and a balanced datasets of 4.665 annotated images was created. To this end, available images in a 30.000 x 30.000 pixels were downloaded and annotated from each of these cities. From the images 2.332 solar panels were identified and cut out in the size of 75x75 pixels, to reduce computational costs. To acquire negative images, several pictures were randomly tiled into 75x75 pixels cutouts and solar panel-free cutouts were selected.

2. Aerial images for *California Data* [1] It is the most commonly used dataset in solar panel detection research papers [1]). This data served as a benchmark in experiment 2 and 3 to compare the results of the best model found in the first experimentation phase and evaluate its generalizability. This data contains annotations of solar panels from 601 images of four Californian cities, with a resolution of at least 30 cm^2 per pixel. From this, a balanced datasets of 39.722 annotated images of 75x75 pixels was created for training and testing.

3. **Methodology**

3.1 **Deep Learning**

Deep-learning [3] attempts to mimic the activity in layers of neurons in the neocortex, the wrinkly 80 percent of the brain where thinking occurs. The software learns, in a very real sense, to recognize patterns in digital representations of sounds, images, and other data. The basic idea is decades old, and it has led to as many disappointments as breakthroughs. But because of improvements in mathematical formulas and increasingly powerful computers, computer scientists can now model many more layers of virtual neurons than ever before. In this experimentation, pre-existing architectures will be investigated. These architectures are available in the Keras library with pre-trained weights.

3.2 **Transfer Learning**

The training of current state-of-the-art convolutional neural networks for large datasets is both computationally expensive and time consuming. Therefore, in our case, training a CNN from scratch is not an ideal option and transfer learning is preferred. The principal advantage of a transfer learning strategy is to quickly create powerful models. Transfer learning refers to the use of existing pre-trained models to classify images on which they have not been trained and
which may even be labelled with different categories. Transfer Learning can be done via various approaches:

- *New data set is rather small and very different from the original data set*. All but the last, fully-connected layers of the pre-trained model are considered and used as feature extractors as it was done in traditional Computer Vision approaches. A machine-learning classifier (e.g. a SVM) is trained on top of the layers with the labels of the new data set.

- *New data set is very different from the original data set but large*. This approach fine-tunes the weights of the pre-trained model, either for all layers or just for some of the later ones, while keeping the remaining layers frozen.

- Finally, some of the layers of a pre-trained network are used as baseline for a new model, followed by the typical convolutional building blocks which would be trained from scratch.

In general, earlier layers of a CNN pick up more general features, similar even to Gabor filters or colour blobs while the higher-layer neurons are specialized to specific tasks and learn increasingly more complex/composite features. Therefore placing the split between frozen and trainable layers at these locations can lead to a lower performance [5]. In this study, the split between trainable and frozen layers was performed after roughly two thirds of the network layers.

3.3 Model Description

As mentioned in Section 3.2, existing pre-trained models are investigated and further trained on the aerial image datasets.
Here, four models where investigated namely InceptionV3, InceptionResNetV2, DenseNet and Xception. All of these architectures have been created during the last three years, are available in the Keras library with pre-trained weights from the ILSVRC [2] and together they present the most important recent improvements in CNN, as inception modules and residual layers. Still there are substantial differences in their network's architectures, as for example their number of parameters and layers (see table 1 below). For more detailed information, see [6]

3.4 Training Setup

In this study, the InceptionV3, InceptionResNetV2, DenseNet and Xception models with weights from the ILSVRC [2] as baseline will be trained on our new datasets. The pre-trained models were adapted for the specific case of solar panel detection, the output layer of the models are extended by addition of a global average pooling to reduce the dimension, a fully connected layer with 512 nodes and finally a sigmoid layer was appended.

3.5 Evaluation

In this paper, 3 experiments carried out to evaluate the models are presented. The accuracy, F-1-Score are the 2 metrics which were chosen for the evaluation. The *accuracy* measures the overall proportion of correct predictions among the total number of predictions i.e. How many percent of the images are correctly classified (positive or negative). In case of binary classification, the *F1-score* is a measure of a test's accuracy and is useful when you want to seek a balance between precision[2] and recall[3]. The F1 score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

4. First Findings

This paper presents the first findings for the of the application of five CNN models: VGG16, InceptionV3, InceptionResNetV2, DenseNet and Xception. These models are publicly available in the Keras library. The models are trained on aerial photographs of different urban landscapes, using existing registers to train and validate the developed algorithms.

The starting point for the models is the use of weights from the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [2]. In this competition different algorithms are tested on how well they perform in the detection of objects and the classification of different images. Detection of solar panels is not part of this competition.

4.1 Experiment 1- Transfer Learning

In this experiment, we investigate to what extent models that have already been trained for other applications can also be used for the detection of solar panels. To this end, the InceptionV3, InceptionResNetV2, DenseNet and Xception model for different options of transfer learning are considered. The results are presented in Figure 4.1-1, below. The best performing model is a pre-trained version of the Xception network, were all 126 layers are set to trainable, Figure2(a). An accuracy of 81.08% on the training and 82.01% on the test set with a F-score of 83.8%. These results are in agreement with the results observed in the ILSVRC [2] were the Xception model outperformed DenseNet121 and Inception V3. In contrast to the renowned competition, on the NRW data set Xception also outperforms InceptionResNetV2. It should also be noted, that all three versions of the InceptionV3 model perform the least well

---

[2] The *precision* expresses the percent of the correctly classify images from the actual solar panel images i.e. How many of those instances identified as positives are actually positive.

[3] The *recall* states the percent of the actual solar panels are recognized as such i.e. how many of the positive instances in the population were identified correctly by the model.

and that the InceptionResNetV2 model trained with one approach is always performing slightly better than the DenseNet model trained with the same approach.

Figure 4.1-1
Experiment 1 : Results overview for each base model and transfer learning approach.

(a) Xception

(b) Inception V3

(c) DenseNet121

(d) IncResNetV2
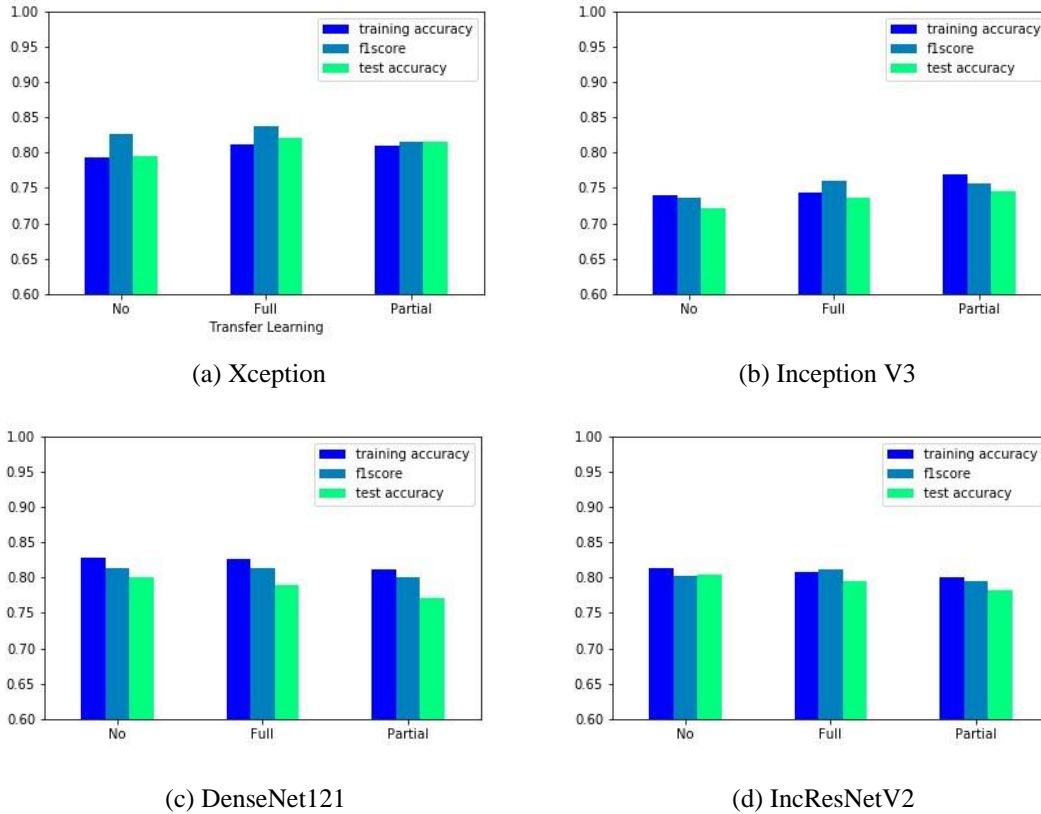
4.2 Experiment 2- Benchmark and Generalization

The aim of this experiment is to investigate how well pre-trained models perform when applied to aerial photographs of other areas. To this end, the best performing model from experiment 1: Xception with full transfer learning trained over North Rhine Westphalia is used as such on aerial photographs of California. Second, the aforementioned best model is used as baseline and further trained further on the California dataset and then evaluated on the NRW and California dataset. Finally, the aforementioned best model is reinitialized with the pre-trained weights from the ILSVRC and then trained on a combined training dataset with images of NRW and California. The performance of the model, presented Figure 4.2-2 clearly decreased when applied to the California dataset - without additional training. By re-training the model completely, the results get better, but surprisingly the best result is obtained if the model is only partly re-trained (a few layers). The same applies to the application at NRW. In the case of partial and, in particular, full retraining, the model performs worse, probably because of overfitting.

Figure 4.2-2
Experiment 2: Results overview for Xception model for the NRW, California and Combined datasets, respectively.


(a) NRW

(b) California

(c) Combined
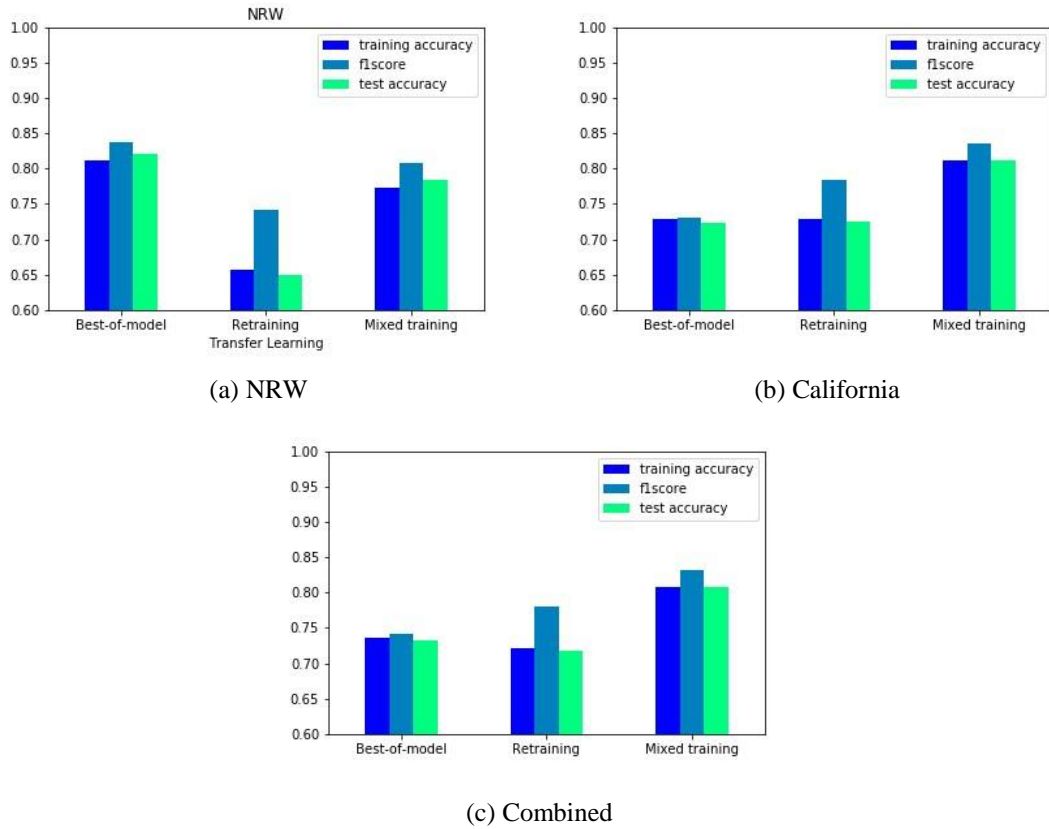
4.3 Experiment 3-Generalization en cross-validation

In the third and last experiment a cross-validation between the aforementioned best model, Xception, and a more traditional architecture, VGG16 is performed. The models are trained at Bradbury (California) and applied to Aachen (North Rhine Westphalia), Fresno (California). The results are presented Figure4.3-3, the Xception model performs slightly better than the VGG16 model over Bradbury and Fresno. When using Xception on the Aachen dataset, the accuracy and precision deteriorate rapidly. The better performance of the VGG16 model over Aachen, on the other hand, indicates that the latter generalizes better than Xception. This decrease in performance could be explained by Xception overfitting to the data.

Figure 4.3-3
Experiment 3 : F-score overview for Xception and VGG16 model over Bradbury, Fresno and Aachen.
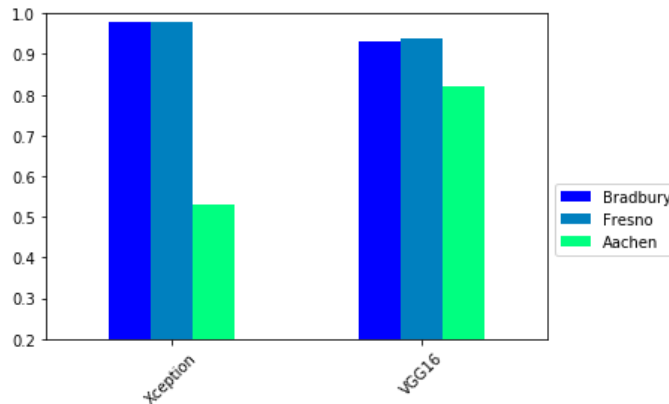
5. Conclusion and Next Steps

First findings in this study show that the existing pre-trained models can be used successfully to classify solar panels from aerial photographs. The best performing models, Xception and VGG16, achieve training and test accuracy and F1-score of more than 90%. Experiment 1, illustrates that the choice of a base model is of prior importance and that this choice cannot always be inferred from state-of-the-art literature. Further, the generalizability to new data sets was investigated, Experiment 2 and 3, which to the authors knowledge has not been done before. The results indicate that the successful application of a model trained on images from a different area is possible, without an extreme loss of information. The performance and generalizability of such a network can be substantially increased, not by retraining the model on the new data, but by taking the model only with its pre-trained weights from the ILSVRC and repeating the training procedure with a mixed data set.

Acknowledgment

References

[1] Kyle Bradbury, Raghav Saboo, Timothy L. Johnson, Jordan M. Malof, Arjun Devarajan, Wuming Zhang, Leslie M. Collins, and Richard G. Newell. Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. *Scientific Data*, 3:160106, December 2016.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[4] Berzirksregierung Köln Geobasis NRW. Digitale orthophotos - paketiert: nach gemeinden [data set]. Retrieved June 14, 2018, from https://www.opengeodata.nrw.de/produkte/geobasis/dop/dop/.

[5] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *CoRR*, abs/1411.1792, 2014.

[6] Hannah. Ziemons. Chasing the sun: A deep learning approach to solar panel classification in aerial images. Master's thesis, Maastricht University School of Business and Economics, 2018.