Poster abstract: predicting heating energy demand by computer vision

Miroslav Despotovic¹ · Muntaha Sakeena² · David Koch¹ · Mario Döller¹ · Matthias Zeppelzauer²

Published online: 30 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract In many countries such as Austria the heating energy demand (HED) is an essential parameter of the energy certification of houses. In this paper, we present an approach in which the HED category for a single family house is—for the first time—determined from a standard photograph directly by means of computer vision and machine learning.

Keywords Heating energy demand · Single family house · Computer vision · Convolutional neural networks

1 Introduction

The detection of energy consumption is usually based on (rather expensive) infrared thermal imaging [1, 3] where thermographic data acquisition and image analysis is applied to identify the energy loss of facades and windows. Based on this, the overall energy consumption of houses is assessed. Further approaches [5] investigate satellite or aerial images in order to identify heat/energy emission. To the best of our knowledge this is the first approach to assess energy efficiency from standard photographs of buildings.

2 Methods

We aim at estimating the HED of a building from an unconstrained image of the outside of the building. Unconstrained

in this context means that, the building can be captured from different perspectives and at different scales and resolutions. The intuition behind our approach is that the visual appearance of buildings (e.g. particular types of windows, roofs, and doors) correlates—to a certain degree—with their HED. To this end, we propose a computer vision-based approach that analyzes local image patches from building pictures. Given an image from a building, in a first step we densely sample differently sized overlapping patches from the image. From these patches we select only a small percentage of patches with the highest intensity contrast. The intuition behind this is to remove patches with low expressiveness, such as patches showing homogeneous regions of the building facade (portions of walls) and patches which lie outside the facade, e.g. in the sky. To model different HED categories we propose an end-to-end learning approach based on convolutional neural networks (CNNs) at patch-level. We employ the AlexNet architecture, pre-trained for the classification of objects in ImageNet images [2]. The network consists of five hierarchical convolutional layers for feature learning and two fully connected layers for classification. We re-define the output with 5 nodes that represent the different HED categories we want to classify. In the test phase, we employ an independent set of building images to objectively evaluate the predictive power of the trained network. Test images are also analyzed at a patch-level (i.e. same patch extraction as for training images). To obtain a robust and final categorization of a building, we perform majority voting on the patch-level HED predictions.

3 Experimental setup and results

Dataset The images used to evaluate our approach were crawled from real estate websites together with their HED

✉ Miroslav Despotovic
miroslav.despotovic@fh-kufstein.ac.at

¹ University of Applied Sciences Kufstein Tirol, Andreas Hofer Strasse 7, 6330 Kufstein, Austria

² St. Pölten University of Applied Sciences, Matthias Corvinus-Strasse 15, 3100 St. Pölten, Austria

categories. The collected dataset comprises 3573 images of different resolutions from a total of 1702 individual family houses spread all over nine Austrian states. We duplicate and flip all the images vertically (data augmentation) to obtain a higher robustness to different perspectives during learning. The resulting dataset thus contains 7146 images. The images are taken as provided. No normalization or segmentation is applied. Each house is assigned to one of the official Austrian HED categories specified by letters “A” to “G” [4], where A represents the lowest HED and G the highest HED. Because of a lack of images in class A (i.e. A++, A+, A) and G in our data, we merged these classes with B and F respectively to balance the dataset, yielding the following five classes: A + B, C, D, E, F + G.

Experimental setup For the training of our classifier we split the (augmented) dataset into three sets: 60% of the images are used for training (4288 images), 15% of the images for the validation of the training progress (validation set, 1058 images) and the remaining 25% (1800 images) represent the independent test set for our evaluation. All three sets are completely disjoint w.r.t. the houses, i.e. houses used for training are not used for validation or testing. The training (fine-tuning) of our network is performed for 40 epochs, i.e. the training data is fed 40 times batch-wise into the network.

Results We employ classification accuracy and the confusion matrix (CM) to assess the performance of our approach. Accuracy is the ratio of correctly predicted images over the total number of images (portion of true positives). The CM (see Table 1) shows the accuracy for each class in its diagonal and the number of falsely classified images in its off-diagonal values. The overall accuracy on the independent test set after 40 epochs of training is 52.50%. This is significantly higher than the random baseline for our dataset (33.44% according to the zero rule) which shows, that our method is able to derive useful visual information related to HED from the photos of the houses. The CM confirms this result. The highest values are located along the diagonal (correct HED predictions). The largest confusions exist between neighboring HED categories (e.g. between A + B and C and between D and C).

Table 1 Confusion matrix for HED classification on the test set

Confusion matrix	A ^p , B ^p	C ^p	D ^p	E ^p	F ^p , G ^p
A ^t , B ^t	234	18	21	6	35
C ^t	106	289	109	38	60
D ^t	59	31	137	16	49
E ^t	46	36	51	96	53
F ^t , G ^t	48	19	38	16	189

Superscript *p* means the predicted class from our approach and superscript *t* refers to the true class

Bold values represent true positives

This is, however, to some degree expected because of the fuzzy transitions between neighboring HED categories.

If we divide the HED into three more generic classes, like “low” (A, B and C), “average” (D) and “high” (E, F and G) the overall accuracy further increases to 63.11%.

4 Conclusion

We have presented first results for the automatic prediction of heating energy demand (HED) from unconstrained photos of houses by computer vision methods and deep learning. Our results confirm a correlation between the appearance of houses and its HED, which can be captured and exploited to predict HED automatically. Our work contributes to real estate image analysis (REIA) which is an emerging interdisciplinary research field with many challenging applications, such as building age and price estimation from images. Although, the current classification accuracy of 52.50% (and 63.11% respectively) is insufficient for a fully automatic appraisal, we expect the features learned from HED classification to be useful input parameters for a more general model to estimate the overall state of a house. Our work represents a first step into this direction. As future work, we further intend to carry out a systematic study of human performance for HED estimation in order to better assess the machine learning performance.

Acknowledgements Open access funding provided by FH Kufstein Tirol - University of Applied Sciences. This work was supported by the Austrian Research Promotion Agency (FFG), Project Nos. 7179681 and 7189193.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Gonzalez-Aguilera D, Hernandez-Lopez D, Lagüela S, Rodriguez-Gonzalez P (2012) Image-based thermographic modeling for assessing energy efficiency of buildings facades. *Energy Build* 65:29–36
2. Krizhevsky A, Hinton EG, Sutskever I (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
3. Möllmann KP, Vollmer M (2010) *Infrared thermal imaging: fundamentals, research and applications*. Wiley, ISBN: 978-3-527-40717-0
4. Richtlinie OIB (2015) 6: Energieeinsparung und Wärmeschutz. Richtlinien des Österreichischen Instituts für Bautechnik, pp 1–17
5. Stathopoulou M, Cartalis C (2007) Use of satellite remote sensing in support of urban heat Island studies. *Adv Build Energy Res* 1:203–212