Chapter 23

Frontiers of Computer Vision Technologies on Real Estate Property Photographs and Floorplans

Yoji Kiyota

Abstract This article describes frontier efforts to apply deep learning technologies, which is the greatest innovation of research on artificial intelligence and computer vision, to image data such as real estate property photographs and floorplans. Specifically, attempts to detect property photographs that violate regulations or were misclassified, or to extract information that can be used as new recommendation features from property photographs, were mentioned. Besides, this article introduces an innovation created by providing data sets for academic communities.

Keywords Deep learning · Computer vision · Real estate property images · Floorplans

23.1 Introduction

According to a recent survey in Japan (The Association of Real Estate Agents in Japan 2016), the proportion of home purchasers who collected real estate information using the Internet has reached 80% or more. Among various information posted on real estate information sites, property photographs are particularly important. In the questionnaire for users of the real estate information sites in Japan (Real Estate Information Site Business Liaison Council 2016), as a result of inquiring about important points (multiple answers allowed) when choosing real estate agents, 80.7% (first place) of the users chose “many photographs are posted”, and 27.5% (fifth place) of the users chose “posted photographs are good” (Fig. 23.1). The result suggests a tendency that property photographs are very important for user experience. The tendency of increasing importance of property photographs can be observed not only in Japan but also in other major countries. Zillow (United States), Rightmove (United Kingdom), SouFun (China), and other major portal sites post a

Y. Kiyota (✉)

AI Strategy Division, LIFULL Co., Ltd, Tokyo, Japan

e-mail: KiyotaYoji@lifull.com

© The Author(s) 2021

Y. Asami et al. (eds.), *Frontiers of Real Estate Science in Japan*, New Frontiers in Regional Science: Asian Perspectives 29,
https://doi.org/10.1007/978-981-15-8848-8_23

325

Fig. 23.1 Points which users of real estate information sites use when choosing agents (cited from p. 7 of the reference Real Estate Information Site Business Liaison Council (2016))

lot of high-quality photographs. In recent years, higher-value image data such as panoramic photographs and movies are also posted. However, there is considerable variation of quality for property photographs posted on real estate information sites, because taking photographs is up to each owner or broker.

A notable feature of real estate property information in Japan is the enhancement of floorplan images. On the real estate information site LIFULL HOME'S, more than 90% of property information is given floorplans. Further utilization of unique contents such as floorplans will also be important in revitalizing the Japanese real estate markets.

As described above, image information such as property photographs and floorplans is very important in the real estate markets, and there are urgent needs for innovation to increase the value of image information. In particular, research and development activities are becoming active all over the world on how to incorporate image processing techniques such as deep learning, which has been rapidly developing in recent years.

This article briefly describes the revolution that deep learning, which is said to be the biggest innovation in recent artificial intelligence research, has brought to the image processing field in Sect. 23.2, and then in Sect. 23.3, research activities on applying deep learning to real estate property photographs, including application to actual services, are introduced. Section 23.4 focuses on an attempt to generate more innovation by providing a large amount of real estate property photographs and floorplan image data to the informatics and computer science research communities. Finally, Sect. 23.5 describes prospects in the future.

23.2 Revolution of Image Processing Technology by Deep Learning

In recent years, there has been an increasing interest in artificial intelligence in society. Today, it is said to be the third artificial intelligence boom following the 1960s and 1980s. “Deep learning” is regarded as a key technology of the third artificial intelligence boom. This section refers to the significant impact that deep learning has had on image processing research.

Deep learning is a type of machine learning and an evolution of neural networks. Studies on neural networks started from imitating the human cranial nerve circuit, and its origin dates back to the 1940s (McCulloch and Pitts 1943).

The first boom on neural network studies began in 1958 with the perceptron (Rosenblatt 1958) published by Frank Rosenblatt. Although this perceptron (simple perceptron) has a simple structure with only two layers, an input layer and an output layer, as shown in Fig. 23.2, it attracted much attention at that time because it can learn and predict. However, Marvin Minsky, a famous artificial intelligence researcher, pointed out in 1969 that a simple problem using an exclusive OR (XOR) operation cannot be solved (Minsky and Papert 1969), and the boom once ended.

Subsequent studies showed that the XOR problem can be solved by inserting a hidden layer into a simple perceptron as shown in Fig. 23.3 to create a multilayer perceptron. Backpropagation, an efficient learning method for multilayered perceptrons, was proposed in 1986 by American cognitive psychologist David Rumelhart and others (Rumelhart et al. 1986), and the boom in neural network research began again. For example, a study in 1998 using the MNIST database,¹ which has been used for handwritten digit recognition tasks for evaluating machine learning algorithms, achieved high performance with an error rate of less than 2.5% with a three-layer perceptron (LeCun et al. 1998).

By the way, is it possible to use neural networks to recognize images that are much more complex than handwritten digits, such as real estate property photographs? It was said that increasing the number of layers could increase the learning ability of the neural network and recognize complex images, but if the number of layers increases, backpropagation will not work well. The result was inferior to other methods using human-designed image features. However, at the ILSVRC² 2012, a competition for image recognition research held in 2012, the University of Toronto’s system SuperVision, which adopted a method developed from a neural network,

¹A data set consisting of a set of handwritten numeric images and correct numeric labels. Provided by the National Institute of Standards and Technology (NIST).

²ImageNet Large Scale Visual Recognition Challenge. A task is required computer to answer what objects (yachts, dogs, cats, flowers, etc.) are in the images. ImageNet is an image database maintained for the purpose of promoting research on image object recognition. More than 14 million image data associated with more than 20,000 synonyms (synsets) of WordNet, a concept dictionary of English.

Fig. 23.2 The structure of the simple perceptron

Fig. 23.3 The structure of the multilayer perceptron

achieved an accuracy that exceeded that of other teams (Krizhevsky et al. 2012). It has had a huge impact on the image processing and artificial intelligence research communities. The method used in SuperVision is the deep learning developed mainly by Professor Jeffrey Hinton at the University of Toronto.

The major point of deep learning is to enable learning of multilayered (deep) neural networks from tens to hundreds of layers by incorporating a kind of “information compressor” called an autoencoder into the neural network. The autoencoder plays the role of “compressing information,” i.e., “extracting only essential features.” The epoch-making point of deep learning is that it acquires high learning ability to capture essential features from images by layering the autoencoders.

When using deep learning actually, it is necessary to learn an enormous number of weighting parameters from an enormous amount of data, and a large amount of

computing power is required. The use of GPGPU (General-Purpose computing on Graphics Processing Units) is practically essential.

Deep learning methods are being applied to various fields such as speech recognition, machine translation, robot control, and automated driving, but the most advanced applications and methods are still in the field of image processing.

23.3 Application of Deep Learning to Real Estate Property Photographs

Almost 5 years have passed since the effectiveness of deep learning became widely known and easy-to-use open-source software libraries have been developed. Research and developments that apply deep learning to real estate property photographs are also increasing. This section introduces some recent examples.

23.3.1 *Photograph Classification for Quality Improvements of Posted Photographs*

As stated in the beginning, quality variation is a major issue in property photographs that are highly valued by users looking for real estate properties. In some cases, photographs that violate the regulations for real estate information are posted. Each company that operates a real estate information site strives to improve information quality through manual checks, etc., but there are limits to manpower in situations where more than millions of photograph data are submitted daily. Since then, efforts are being made to use state-of-the-art image processing technologies such as deep learning.

Kikuta et al. (2016) reported an example of deep learning applied to the task of detecting anomalous photographs that violate regulations at the real estate information site SUUMO. In the task of detecting “photographs with people reflected,” a type of convolutional neural network (Convolutional Neural Network, CNN) that is a deep learning method suitable for image processing is used. They reported that the probability of missing an abnormal photograph is less than 5%.

Ishida and Kiyota (2016) used the LIFULL HOME’S data set (described later) to evaluate the accuracy of automatic discrimination by deep learning of 13 types³ of photographs. It is reported that the error rate of 14.3% was achieved by learning from 130,000 photograph data (10,000 samples randomly selected for each type) using CNN. As shown on the left of Fig. 23.4, although the accuracy is low for classifications such as “living,” where the judgment by humans also tends to fluctuate, the

³13 types of photographs, including floorplan, map, entrance, living room, kitchen, bath, restroom, washbasin, storage, equipment, balcony, entrance hall, and parking.

Fig. 23.4 Classification of property photograph types using deep learning

“kitchen” and “bath” achieve extremely high accuracy. Even in the error example, there are not a few examples that are considered to be classified into multiple types. On the right side of Fig. 23.4, there are subtle examples of errors such as classifying “bathroom washbasin” (the correct answer is “washbasin”) as “bath.”

As mentioned above, real estate property photograph classification by deep learning has achieved the same level of accuracy as human beings at present, so application examples in the business are being reported. The author’s company has been operating a system for detecting inconsistencies in the category of real estate property photographs submitted by real estate companies since December 2016 (LIFULL Co. Ltd. 2016). LIFULL HOME’S has a system that gives priority to displaying properties with more room photographs registered in the search results from the viewpoint of providing more useful information to users. As in the case of photographs, there is a problem that inconsistencies occur such as that photographs other than indoor photographs are registered by indoor type. Therefore, by using deep learning, the consistency rate is automatically calculated as shown in Fig. 23.5, and for the photographs that are inconsistent with the registration type, the registration real estate company is encouraged to correct it.

23.3.2 *Photograph Analyses for Promoting Values of Property Information*

In response to the diversification of users’ needs for finding real estate properties, the real estate information site also supports adding various search conditions such as “counter kitchen,” “broadband connection,” and “convenience store nearby.”

Posted photo	Specified type	Score	Result
Kitchen	Kitchen 70.3% Entrance 23.7% Equipments 2.804%

Kitchen	Kitchen 97.3%

Kitchen	Kitchen 0% Entrance 0%

Fig. 23.5 Inconsistency detection of property photograph classification using deep learning

However, since there are so many factors related to the ease of living of the property, the maintenance of the database has not kept up with the diversification of needs.

In response to the diversification of users' needs for finding real estate properties, the real estate information site also supports adding various search conditions such as "counter kitchen," "broadband connection," and "convenience store nearby." However, since there are so many factors related to the ease of living of the property, the maintenance of the database has not kept up with the diversification of needs. Therefore, attempts have been made to improve the value of real estate property information by extracting indices related to comfortability of living from property photographs. Ishida and Kiyota (2016) focused on "comfortability of use of the kitchen," which greatly affects the ease of living and conducts an experiment to distinguish two types of indicators, "Kitchen type" and "Workspace," using deep learning. For the former, create a data set (consisting of 1000 photographs of each type, a total of 5000 photographs) classified into five types: "system kitchen," "simplified system kitchen," "non-system kitchen," "kitchen part," and "others." And by learning with CNN, a high accuracy of 11.6% error rate has been achieved. For the latter, we created a data set (categorized into 5500 photographs in Fig. 23.6, consisting of a total of 5500 photographs) that was categorized into 6 types including "very narrow" to "very wide" plus "others." Although the error rate of category discrimination is not so good at 36.2%, it can be seen from the mixing matrix (lower left of Fig. 23.6) that the size can be identified to some extent. When the correlation coefficient is calculated by assigning a breadth score to each category, it is 0.717

Fig. 23.6 Detection of workspace width of kitchens

(lower right in Fig. 23.6), and it can be expected that practical accuracy will be achieved by expanding the data set.

23.4 Promotion of Open Innovations in the Real Estate Industries Through Provision of Data Sets for Academic Communities

As mentioned above, applications of deep learning to real estate property photographs become active in business situations. However, there is an overwhelming shortage of human resources to implement deep learning in order to further draw out the potential of advanced image processing technologies such as deep learning and create new innovations. In particular, human resources who are familiar with deep learning are rare, and it is not realistic to create innovation with just one company.

Therefore, our company began to activate studies related to real estate by providing a data set including image data such as property photographs and floorplans held by our company for academic research purposes. With the cooperation of the National Institute of Informatics of Japan (NII), we started providing “LIFULL HOME’S Data set” (National Institute of Informatics 2015) (Fig. 23.7) in

The screenshot shows the website of the Informatics Research Data Repository. The header includes the NII 20th anniversary logo, the IDIR logo, and the text 'Informatics Research Data Repository'. A navigation bar contains links for HOME, Data List, Paper List, About Us, Related Links, and Contact. Below the navigation bar is a breadcrumb trail: > HOME > Data List > LIFULL HOME'S Dataset.

On the left side, there is a sidebar with a list of datasets: Yahoo! Dataset, Rakuten Dataset, LIFULL HOME'S Dataset (selected), NTCIR Test Collection, and Speech Corpus.

The main content area is titled 'LIFULL HOME'S Dataset'. It contains the following text: 'The National Institute of Informatics provides LIFULL HOME'S Dataset to researchers, which was offered by [LIFULL Co., Ltd.](#) for promoting research in informatics and the related fields.' The update date is listed as 'update: 2020-09-07'.

Below this, there is a section titled 'Outline of the Data'. It states: 'The dataset contains the data of [LIFULL HOME'S](#), a Real Estate Information Service in Japan.'

The outline lists three main categories of data:

- 1. Snapshot Data of Rentals** (snapshot of 2015-09)
 - Rental data (5.33 million all over Japan): rental fee, area, location, age, floor plan, structure, facilities, etc.; approx. 1.6GB .tsv format files.
 - Image data (83 million files): floor plan image, room view, etc. of all the above items; approx. 210GB .jpg format files, max size: 120x120.
- 2. High Resolution Floor Plan Image Data**
 - High resolution version data of floor plan image (5.31 million files) included in Snapshot Data of Rentals; approx. 140GB .jpg format files. Additional application required to use this data (see Application section below).
- 3. Monthly Data of Rentals and Sales** (2015-07 - 2017-06, 24 months)
 - Property data of rentals and sales (5.33 million all over Japan): rental fee/price, area, location, age, floor plan, structure, facilities, etc.; approx. 1.7 - 4.5GB .tsv format files for respective months.

Fig. 23.7 LIFULL HOME'S data set

November 2015. The LIFULL HOME'S data set includes information on all properties for rent (approximately 5.33 million) that were listed on LIFULL HOME'S as of September 2015, property photographs (approximately 83 million items) associated with it, and floorplan images (approximately 5.15 million items). It is currently provided to more than 80 university laboratories and research institutions in Japan and overseas. More than 3 years have passed since the launch, and very interesting research is being announced.

I would like to briefly introduce one of the very interesting research cases using the property photographs and floorplan image data included in the LIFULL HOME'S data set. A study group at Simon Fraser University in Canada has shown that it is possible to create new applications by solving the task of correlating floorplans with indoor photographs using deep learning (Liu et al. 2016). Consider the “quiz for selecting the correct bathroom photograph corresponding to the floorplan” as shown in Fig. 23.8 (the correct answer is (A)). This quiz is a very difficult task for humans, and even in an experiment by a crowdsourcing service (Amazon Mechanical Turk) worker, the correct answer rate is only 43%, and it takes 30 seconds or more on average to solve one task. However, by using a deep neural

Fig. 23.8 Which of the four photographs corresponds to the left floorplan?

Fig. 23.9 The architecture of the deep neural network for answering the questionnaire in Fig. 23.8

network as shown in Fig. 23.9, a correct answer rate of 72% far exceeding that of human beings has been achieved, and more than 20 problems can be solved in one second.

If the deep neural network learned as described above is used, there is a possibility that the position on the floorplan corresponding to the indoor photograph can be estimated. When the visualization method is used, as shown in Fig. 23.10, it can be seen that the position on the floorplan corresponding to the indoor photograph (center) is correctly pointed by the red spot on the right side of the figure. This result

Fig. 23.10 Visualization for the receptive field of the prediction by the neural network

seems to suggest the possibility of realizing new navigation based on floorplans on the real estate information site.

23.5 Conclusion

In this article, we introduced the outline and application examples of image processing technologies, especially deep learning, to further enhance the quality and value of property photographs and floorplans that are very important in real estate property information. Image processing technology is still developing rapidly, and it is expected that even greater innovations will be generated one after another.

On the other hand, researchers and engineers who are familiar with image processing technologies such as deep learning are extremely rare even in the world, and the competition for human resources is not only between companies but also between industrial fields. With research and development in various industrial fields such as advertising, finance, automobiles, and robots, to create new innovations in the real estate field, it is important to develop a mechanism that encourages people familiar with such technologies to engage in the real estate field. In order to attract such people, it is indispensable to present challenging tasks and to develop a data set and research community as infrastructure for research and development. I would like to make further contributions to the creation of such R&D infrastructure in the real estate field.

References

- Ishida Y, Kiyota Y (2016) Trial of information extraction by deep learning from real estate image for the purpose of housing selection support. In Proceedings of the special interest groups for web intelligence and interaction (ARG WI2 No. 8), pp 29–30. https://www.sigwi2.org/wp-content/uploads/2017/08/WI2_2016_8.pdf (in Japanese)
- Kikuta Y, Nomura S, Li S, Kobayashi S, Kozu T (2016) Inappropriate image detection based on deep learning. In Proceedings of the 30th annual conference of the Japanese Society for Artificial Intelligence, 1A4-OS-27b-1. <https://www.ai-gakkai.or.jp/jsai2016/webprogram/2016/pdf/664.pdf> (in Japanese)
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In Proceedings of advances in neural information processing systems 25 (NIPS 2012), pp. 1097–1105
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- LIFULL Co. Ltd. (2016) Press release: started inconsistent image detection of properties using AI. <http://lifull.com/news/7529/>. Accessed 31 Aug 2019
- Liu C, Wu J, Kohli P, Furukawa Y (2016) Deep multi-modal image correspondence learning. <http://arxiv.org/abs/1612.01225>. 30 Apr 2017
- McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5(4):115–133

- Minsky M, Papert S (1969) *Perceptrons: an introduction to computational geometry*. MIT Press, Cambridge
- National Institute of Informatics (2015) Informatics research data repository: LIFULL home's data set. <http://www.nii.ac.jp/dsc/idr/lifull/homes.html>. Accessed 31 Aug 2019
- Real Estate Information Site Business Liaison Council (2016) Real estate information site user awareness questionnaire survey results. <https://www.rsc-web.jp/pre/img/161027.pdf>. Accessed 31 Aug 2019 (in Japanese)
- Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 65(6):386–408
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(6088):533–536
- The Association of Real Estate Agents in Japan. (2016) The 21st consumer trend survey on real estate distribution businesses: survey results report (summary version). https://www.frk.or.jp/information/2016shouhisha_doukou.pdf. Accessed 31 Aug 2019 (in Japanese)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.