

**Transfer learning for hostel image classification**

Journal:	<i>Data Technologies and Applications</i>
Manuscript ID	DTA-02-2021-0042.R2
Manuscript Type:	Article
Keywords:	Transfer learning, CNN, Image classification, Hostel image, Artificial neural network, digital image

SCHOLARONE™  
Manuscripts

**Title**

# Transfer learning for hostel image classification

**Abstract**

**Purpose:** Due to the fast-growing digital image collections on online platforms and the transfer learning ability of deep learning technology, image classification could be improved and implemented for the hostel domain which has complex clusters of image contents. This paper aims to test the potential of 11 pretrained Convolutional Neural Network (CNN) with transfer learning for hostel image classification on the first hostel image database in order to advance the knowledge and fill the gap academically, as well as to suggest an alternative solution in optimal image classification with less labour cost and human errors to those who manage hostel image collections.

**Design/methodology/approach:** The hostel image database is first created with data pre-processing steps, data selection and data augmentation. Then, the systematic and comprehensive investigation is divided into seven experiments to test 11 pretrained CNNs which transfer learning was applied and parameters were fine-tuned to match this newly created hostel image dataset. All experiments were conducted in Google Colaboratory environment using Pytorch.

**Findings:** The 7,350-hostel image database is created and labelled into seven classes. Furthermore, its experiment results highlight that DenseNet 121 and DenseNet201 have the greatest potential for hostel image classification as they outperform other CNNs in terms of accuracy and training time.

**Originality/value:** The fact that no existing academic work dedicating to test pretrained CNNs with transfer learning for hostel image classification and no existing hostel image-only database have made this paper a novel contribution.

**Keywords:** Transfer learning, hostel image classification, CNN

**Article Classification:** Research paper

## 1. Introduction

1  
2 Due to the economic and societal impacts of current coronavirus pandemic on the hostel sector, and other  
3 sectors in the global economy, as well as the fast-growing digital image collections on online platforms, it is  
4 crucial to utilize the state-of-the-art technology available to organize/classify these images and improve the  
5 sector's services/products in order to bounce back to a stronger position and accommodate the changing  
6 behaviours of potential customers. Therefore, a brief review of three key areas is addressed as follow.  
7  
8

### 9 *1.1 Tourism industry and hostel sector in coronavirus pandemic*

10  
11 Tourism has been a significant sector of many national economies around the globe and, unfortunately, it has  
12 also been one of the most directly affected industry in the current global pandemic of coronavirus (COVID-  
13 19). According to the Organization for Economic Co-operation and Development (OECD, 2020) report, the  
14 COVID-19 impact results to around 80% decline in international tourism in 2020. Several governments have  
15 introduced tourism specific measures to support the industry which some measures are specifically for hostel  
16 market. For example, the Belgian government subsidies six million euros to the tourism industry, dedicatedly  
17 for the Flemish region with one million euros for youth hostels (Bruegel Datasets, 2020), and the Mayor of  
18 London makes 40 million pounds fund available for COVID-19 safe hostels (Greater London Authority, 2020).  
19 Despite the economic downturn and tight budget, it can be seen that the governments see the opportunities for  
20 economic recovery through the hostel sector. Furthermore, the Hostels Global Market Report 2020-2030  
21 mentions that the market could grow from 6.08 billion US dollars in 2019 to 7.5 billion US dollars in 2023  
22 (Businesswire, 2020).  
23  
24  
25  
26  
27

### 28 *1.2 The fast-growing digital image collections*

29  
30 In spite of a number of major hostel chains such as A&O, Generator, Meininger, and St. Christopher's Inns,  
31 the hostel market consists of countless independent hostels which heavily rely on online travel agencies for  
32 their customers as these e-intermediaries play major roles as a fare aggregator and a reservation portal. Some  
33 of the key players are [www.hostelz.com](http://www.hostelz.com), providing a list of almost 50,000 hostels in over 7,000 cities,  
34 [www.hostelworld.com](http://www.hostelworld.com), containing a database of over 36,000 properties, [www.hostels.com](http://www.hostels.com) and  
35 [www.booking.com](http://www.booking.com), listing over 30,000 hostels on their sites. Furthermore, the fact that the hostel market  
36 considers its images as a critical factor as images portrayed its facilities and unique atmosphere which highly  
37 influence the decision-making process of potential guests when choosing an accommodation (Bunda, 2014)  
38 and the fact that digital image sharing on social media via smartphone has become a phenomenon in today's  
39 society, 350 million images are uploaded every day on Facebook, more than 50 billion images so far have been  
40 uploaded to Instagram, and over 1.2 billion images and videos are uploaded on Google Photos (Aslam, 2021a;  
41 Aslam, 2021b; Sabharwal, 2017; Stadlen, 2019), the vast and growing collections of digital images have  
42 become a challenge for those who manage these image data including e-intermediaries' portals such as  
43 [booking.com](http://booking.com) where travellers upload more than 150 million images onto the system in order to share their  
44 experiences during their stay (Tasli, 2017a). Therefore, to manually or automatically classify this enormous  
45 amount of images in their database and systematically organize for future use such as image retrieval or image  
46 analysis could be costly and time-consuming, let alone correctly classify these images. In addition, by  
47 considering a hostel image, particularly the interior image, it tends to have more complex clusters of image  
48 contents when compared to a hotel image. This leads to more difficulty in image classification/organization.  
49  
50  
51  
52  
53  
54  
55

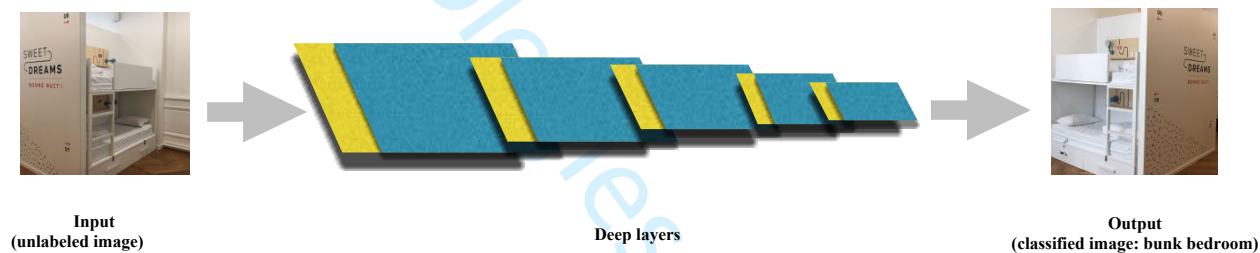
### 56 *1.3 Deep learning and transfer learning for image classification*

57  
58 To organize the huge image collections and fasten the image indexing task, various attempts on image  
59 classification have been found in the tourism domain. For example, Amato *et al.* (2015) conduct a monument  
60 recognition experiment using the k-Nearest Neighbour technique, Zheng *et al.* (2014) apply Scale-Invariant

Feature Transform (SIFT) and binary features in image matching on Holidays database, Zhu *et al.* (2015a) use multimodal hypergraph to categorize landmarks. Despite the image content could be represented by classical features such as colour, texture, shape, SIFT or spatial positions, artificial Intelligent-based features have become in focus among academia and the industries for image classification task as an Artificial Neural Network (ANN) algorithm could simulate human cognition process and acquire training to classify an image through deep learning with less labour cost and human errors. Among various types of Artificial Neural Network (ANN), Convolutional Neural Network (CNN) has been primarily applied on renowned image classification works including AlexNet, ZF Net, GoogLeNet, and ResNet, the winners of ImageNet Large Scale Visual Recognition Competition (ImageNet, 2020).

A CNN model consists of a set of components such as convolutional layers, pooling layers, and fully connected layers. The convolutional operations traverse input matrices with convolutional kernels, which act as filters in order to automatically learn parameters such as kernel size, strides, channels, padding, and activation via deep learning method, in the low-level feature extraction process. Then, the output is subsampled by pooling layers to learn high-level features representing the original input. The fully connected layers are placed lastly to generate numerical output or act as the classifier. Despite CNN models vary in a number of deep layers and layer type, a typical CNN architecture can be illustrated in figure 1 below.

Figure 1: CNN model architecture



However, the accuracy of image classification using CNN highly depends on large scale database as the more information it receives, the more accurate performance results it is likely to provide, high-end computational unit to execute this compute-intensive task in a short period of time which is costly to operate and procure, and the network depth as the deeper network could create more detailed representatives of image input and become more efficient in computation and number of parameters (Shaha and Pawar, 2018a). Therefore, the transfer learning technique is further applied as a solution for image classification task on small scale dataset with less computational cost as its concept is to use a CNN model which is pretrained on a large dataset and transfer some of its learned parameters to the adapted version of the pretrained CNN in order to perform classification task on a much smaller and different image dataset. Despite the fact that image classification on hostel images is more challenging than hotel images due to more complex clusters of image contents, to the best of our knowledge, there is no academic study on image classification with transfer learning for the hostel sector. Nor an existing database specifically dedicated to the hostel image collection.

As a result, the first hostel image database is created and tested for the potential application of transfer learning on pretrained CNNs for the image classification task. The rest of this paper organized as follows. Section two addresses the relevant literature. Section three provides chosen methodology including the sources and pre-processing steps of the newly created hostel image database, investigated CNN models, and approach for experiments. Results with critical analysis are shown in section four. Lastly, the conclusion and future directions are demonstrated in section five.

## 2. Literature review

Similar to human judgement, the more information we receive, the better decision we make, it is essential for CNN to have sufficient information in order to achieve a good result. Therefore, large databases are created to train CNN such as ImageNet which contains over 14 million images and Place365 which contains over 1.8 million images. Several CNN models were developed and extensively trained on these large databases, especially ImageNet which is a large-scale database used in Large-Scale Visual Recognition Challenge, such as AlexNet, Visual Geometry Group (VGG), GoogLeNet, and Residual Network (ResNet) and many of these state-of-the-art neural networks are widely used in image classification tasks. For example, Dawud *et al.* (2019a) and Zhao *et al.* (2017a) applied AlexNet in their experiment, Li *et al.* (2020a), Mateen *et al.* (2020a), Shaha and Pawar (2018b) and Stylianou *et al.* (2019b) used VGG, Szegedy *et al.* (2015) preferred GoogLeNet, and Mateen *et al.* (2020b), Stylianou *et al.* (2019c), and Xiao *et al.* (2018a) chose ResNet.

However, for image classification task or other tasks on a much smaller database, even with data augmentation applied by executing rotation, horizontal flipping, vertical flipping, or resizing images in the database in order to provide more information to CNN model via these artificial increase of training images, it could be difficult for the model-CNN to learn millions of its network parameters and provide good performance as it is likely to overfit the training data with low bias which lead to false performance. As a result, transfer learning technique is introduced to small scale database tasks as an alternative to deep learning the CNN from scratch. The process of transfer learning can be explained in two steps, importing the fully pretrained CNN construction and fine-tuning its parameters, as illustrated in figure 2 below.

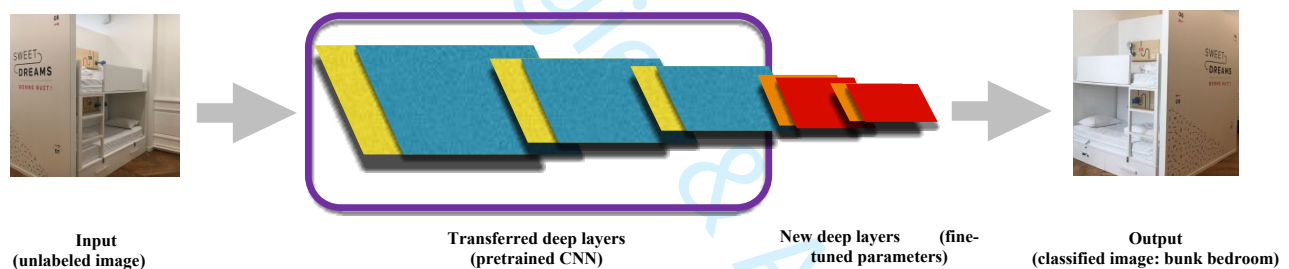


Figure 2: CNN transfer learning for image classification

To reduce computational cost and achieve high accuracy, transfer learning takes advantage of pretrained CNN by using the learned parameters/weights from its earlier layers which extract low-level features such as edges, patterns, corners, and gradients which are general features in any image. Despite different database used on the pretrained CNN, these low-level features are similar and can be transferred to the new task resulting in increasing the feature extraction capability of the adapted CNN by requiring with less computational time in the training processes. The next step is to fine-tune parameters by making small adjustments on parameters/weights of the pretrained CNN on the latter layers which extract high-level features such as objects and events which are specific features in a given image in order to match new dataset and then, add a new classifier and train this adapted CNN for the new task. This fine-tuning step can be repeated until the desired result is achieved. Transfer learning does not only help to fasten the overall process of image classification but also reduces computational resources used such as large storage space, high-speed network connection, and multi-core processors. Furthermore, better performance could be achieved by using a lot fewer data during the training of adapted CNN. ~~Despite different database used on the pretrained CNN, the low-level features are similar and can be transferred to the new task resulting in increasing the feature extraction capability of the adapted CNN with a less computational cost.~~

Several previous works apply transfer learning by directly employing the pretrained CNN and fine-tuning it for their databases such as Shaha and Pawar (2018c), Stylianou *et al.* (2019d), Xiao *et al.* (2018b) and Zhao *et al.* (2017b) and some experiments additionally modify the pretrained CNN by changing its layers (Dawud *et al.*, 2019b; Li *et al.*, 2020b; Mateen *et al.*, 2020c; Zhao *et al.*, 2017c). Moreover, there is evidence of transfer learning application on CNN features on hotel images such as hotel recognition for human trafficking investigation (Stylianou *et al.*, 2019a), hotel image classification using deep learning (Kanchinadam, 2016), hotel photo content analysis (Ren *et al.*, 2020), Kayak's hotel image categorization with deep learning (Rath, 2018), and automated hotel image tagging (Tasli, 2017b). Furthermore, Kaggle (2016) launched a competition on hotel image classification to encourage students applying deep learning to a real-world problem. Nonetheless, there is no relevant evidence on hostel image experiment and it is not necessary that the high accuracy of the same pretrained CNN which used ImageNet database is transferable to other tasks which use different datasets (MathWorks, 2021). As a result, the hypothesis of this study is constructed below.

Hypothesis: Transfer learning of pretrained CNNs provides good performance in hostel image classification

### 3. Research methodology

As the methodology is a system of ways used to conduct research which is linked to ontology and epistemology, two different types of research, primary research and secondary research, are described by Mason (2014), according to its source of information. Therefore, this study is considered primary research as the testing potential of 11 pretrained CNNs with transfer learning for hostel image classification has not been investigated before. Additionally, modelling/simulation method is chosen to test the algorithms resulting in quantitatively outcomes.

#### 3.1 Hostel image database

##### A. Data sources

According to the relevant literature, the majority of relevant experiments used an existing database(s) (Abdullahzadeh and Mohanna, 2013; Wengert *et al.*, 2011; Zhu *et al.*, 2015b), whereas some experiments used images specifically photographed for their research projects (Premchaiswad *et al.*, 2010; Raisi *et al.*, 2011), and some experiments combined both data collection methods to create their image database (Raisi *et al.*, 2014). Because of this, data collection methods are subjective. Therefore, in the case of this experiment, hostel images are collected from five freely available sources which are online travel agencies' websites such as [www.booking.com](http://www.booking.com) and [www.hostelworld.com](http://www.hostelworld.com), Google, Flickr, Places365 database, and by taking photographs at the premises, where possible, to create the first hostel image-only database which has hostel images taken by both professional and non-professional photographers. Consequently, as various data sources are included, images in this database could represent different hostel designs which varying from traditional to modern, simple to eclectic, and/or budget to boutique. Moreover, images of hostels across the world are considered in order to reduce bias. Lastly, a variety of image quality from the professional and the non-professional and time that these images are available and collected, as of December 2020, could be well-demonstrated the real-world digital image collections on current online platforms.

##### B. Data pre-processing



## Data selection

Similar to the data collection method, it seems there is no academic scholar suggests a number of reliable data size or how to calculate the appropriate data size for research in relation to image classification. Nonetheless, the key advantage of transfer learning is to improve classification accuracy when using deep features on small scale datasets and it is evidenced in Rath's experiment (2020a) that for the 8,677-image dataset the classification performance can reach 95% accuracy. Therefore, in this work, **7,350** hostel images are collected which consist of images from chain hostels, hostels which are corporately owned and usually have a similar design, and independent hostels, hostels which are usually privately owned and have unique designs. These images are chosen from hostels in different locations such as the United States in North America, the United Kingdom in Europe, Thailand in Asia, and Ghana in Africa. Moreover, only interior images which contain complex clusters are considered in order to demonstrate the greater challenge of hostel images when compared to hotel images as it is more likely that there are more than two beds in the same room in a hostel bedroom, there are more than one washing machine/dryer in a hostel laundry room, there are more than one shower/toilet in a hostel bathroom, and there are more objects in a hostel kitchen/living room.- Therefore, it is possible that hostel image features consist of vaster feature types and levels such as many colours, many textures, and many shapes which account for the higher number of features to consider when performing classification task. Then, The dataset is divided into two sets, training/validation set and test set. The training/validation set consists of the training set which is used to train the algorithm/model and the validation set which is used to unbiasedly evaluate the algorithm/model while fine-tuning parameters, containing **7,000** images. The test set, containing **350** images, is used to unbiasedly evaluate the final algorithm/model. Images in training/validation set are labelled and grouped into seven classes, normal bedroom, bunk bedroom, other bedroom, bathroom, kitchen, living room, and laundry room. Since the bedroom is the key feature for hostel accommodation, three out of seven classes are allocated to three types of bedroom. Furthermore, each class contains **1,000** images equally in the training/validation set and **50** images equally in the test set in order to control this variable which might affect the test performance in this study. Additionally, despite these images are in different sizes, all of them are; in three colour channels and stored in JPEG format as suggested by Wallace *et al.* (1991) as they illustrated that JPEG is an appropriate still image format for the use of image retrieval /classification purpose. For non-commercial research and educational purposes, this dataset is made available on Google Drive (2021) and hostel image examples of each class are demonstrated in figure 3 below.

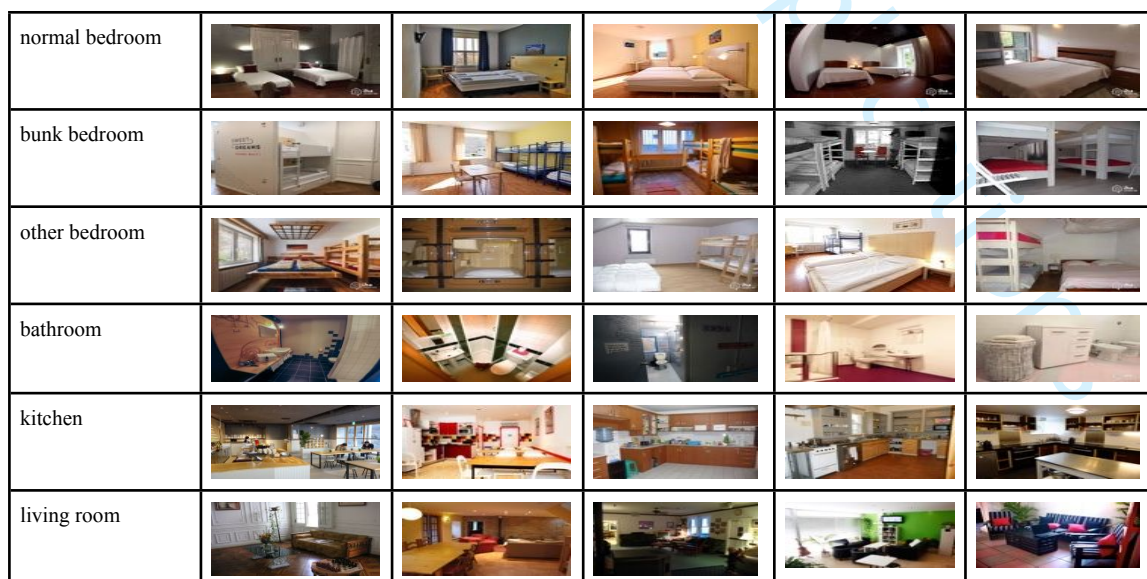




Figure 3: Hostel image examples of each class

### Data augmentation

In order to help increasing accuracy of this deep learning experiment, data augmentation is applied to provide multiplied data to the system which could also help to prevent the overfitting during training. Image data can be augmented by rotation, horizontal flipping, vertical flipping, shifting, and/or resizing.

Furthermore, some of the regularization techniques are also applied in order to facilitate CNN models. These techniques include weight sharing, dropout, early stopping, batch normalization, noise injection, and parameter norm penalties.

### 3.2 Pretrained CNN models

Although more complex models might lead to more accurate results, to be compatible with digital image sharing on mobile devices, in this paper 11 small-size CNN models are explored which are SqueezeNet models (SqueezeNet1\_0 and SqueezeNet1\_1), ResNet models (ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152), and DenseNet models (DenseNet121, DenseNet161, DenseNet169, and DenseNet201).

SqueeneNet models are the most compact networks available which take advantage of fire module using only 1 x 1 filters in squeezed convolutional layers and a mix of 1 x 1 and 3 x 3 filters in expanded convolutional layers. ResNet models provide optimized solutions to the degradation problem, which tends to cause by an increase of deep layers, by letting the stacked layers fit a residual mapping instead of the underlying mapping. Similar to ResNet models, the DenseNet models add shortcuts in deep layers but also concatenate the outputs from previous layers in deep dimensions. Therefore, DenseNet models use fewer parameters than ResNet to represent image features. Moreover, models with different numbers of deep layers are included for a comprehensive approach and their accuracies and relative prediction time using GPU of these models are also feasible for this study. The information about the chosen CNN models is demonstrated as follow.

CNN model	Model size	No. of deep layer	No. of parameter
SqueezeNet1_0 (Iandola <i>et al.</i> , 2017)	5.2MB	18	1.2 million
SqueezeNet1_1 (Iandola <i>et al.</i> , 2016)	5.2MB	18	1.2 million
ResNet18 (He <i>et al.</i> , 2015a)	44MB	18	11.7 million
ResNet34 (He <i>et al.</i> , 2015b)	87MB	34	21.2 million
ResNet50 (He <i>et al.</i> , 2015c)	96MB	50	25.6 million



ResNet101 (He <i>et al.</i> , 2015d)	167MB	101	44.6 million
ResNet152 (He <i>et al.</i> , 2015e)	230MB	152	58.2 million
DenseNet121 (Huang <i>et al.</i> , 2018a)	33MB	121	7.2 million
DenseNet161 (Huang <i>et al.</i> , 2018b)	107MB	161	12.3 million
DenseNet169 (Huang <i>et al.</i> , 2018c)	57MB	169	12.8 million
DenseNet201 (Huang <i>et al.</i> , 2018d)	77MB	201	20 million

Table I: CNN models information

### 3.3 Experiments

This study is conducted in the Google Colaboratory environment, a free cloud service with free graphics processing unit (GPU), using Pytorch on Intel(R) Core (TM) i7-6700 CPU @3.40GHz machine with 16GB RAM and NVIDIA Tesla K80 GPU with 12GB memory. The implemented Pytorch code is an adapted version of Rath's algorithm (2020b). The approach is to directly employ 11 pretrained CNNs. Then, transfer learning is implemented with fine-tuning some parameters on the latter layers to match the hostel image database collected. Furthermore, all images, 7,000 images, in training/validation set are randomly rotated between -30 and 30 degrees, centre cropped and resized to 224x224 format as an input format required by the selected pretrained CNNs, and horizontally flipped for data augmentation process. On top of this, batch normalization which is a supervised learning technique that standardizes the inputs in a neural network layer, weight sharing which is a technique that uses the same numbers within each neuron/filter in a particular layer, and dropout which is a technique that ignores randomly selected neurons during training are chosen as the regularization techniques to facilitate these CNN models.

SqueezeNet1\_0 model is selected as a CNN baseline in this study which is divided into seven experiments as follow:

**1st experiment** is to test the baseline performance for five epochs on various proportions of data split in training/validation set (90:10, 80:20, 70:30, 60:40, and 50:50) and its effect on the test set performance;

**2nd experiment** is to test the baseline performance for five epochs on various dropout rate in training/validation set (0.1, 0.2, 0.3, 0.4, and 0.5) and its effect on the test set performance;

**3rd experiment** is to test 11 pretrained CNNs performances at the best-performed data split and dropout rate;

**4th experiment** is to test 11 pretrained CNNs performance for various epochs at the best-result data split and dropout rate;

**5th experiment** is to test 11 pretrained CNNs performance for 25 epochs at the best-result data split and dropout rate and its effect on the test set performance;

**6th experiment** is to test 11 pretrained CNNs performance for 25 epochs at the best-result data split and dropout rate and its effect on the test set performance on the database with original image sizes and the database with resized images (300x300);

**7th experiment** is to analyze the accuracy of the best-performed CNN on the training set and validation set.

### 3.4 Performance measurement

To evaluate the transfer learning performance of pretrained CNN models on hostel image classification task, two measures are implemented as follow.

A. Accuracy is a ratio of the number of correctly classified images to the total number of all images.

$$\text{Accuracy} = \frac{\text{The number of correctly classified images}}{\text{The total number of all images}} \quad (1)$$

B. Training time (minutes)

## 4. Results with critical analysis

### 4.1 1st experiment

Data split (Training set: Validation set)	Accuracy (%)	
	Validation set	Test set
90:10	<b>74.228</b>	70.857
80:20	72.626	69.714
70:30	72.960	70.286
60:40	69.472	67.143
50:50	66.924	<b>71.714</b>

Table II: An accuracy comparison of SqueezeNet1\_0 on validation set & test set at different data split. Bold values indicate the highest accuracy and the optimal option is highlighted.

Looking at an accuracy of SqueezeNet1\_0 (baseline CNN) on the validation set, it can be seen that the data split between the training set and the validation set at 90:10 provides the highest accuracy, 74.228%, whereas the accuracy on test set demonstrates the highest at 50:50 proportion, 71.714%. Having said this, as one proportion can only be selected to apply in further experiments, the accuracies on validation set and test set need to be considered together. Despite the data split at 90:10 achieves the best results, the 70:30 proportion could also be considered as it provides the second-best results. Therefore, due to the benefit of validation set which is to provide an unbiased performance of the training model, in the 2nd experiment data split at 90:10

and 70:30 are tested to finalize the most appropriate data proportions for the rest of the experiments. Additionally, it is worth noting that even though a low-cost model like SqueezeNet1\_0 is implemented, the classification performances are reasonable for mobile application.

#### 4.2 2nd experiment

Dropout	Accuracy (%) at 90:10 data split		Accuracy (%) at 70:30 data split	
	Validation set	Test set	Validation set	Test set
0.1	<u>71.769</u>	<u>69.376</u>	71.632	70.000
0.2	<u>67.395</u>	<u>61.958</u>	68.966	62.571
0.3	<b><u>72.140</u></b>	<u>69.965</u>	<b>72.960</b>	70.286
0.4	<u>69.264</u>	<u>70.011</u>	69.748	70.571
0.5	<u>64.752</u>	<b><u>71.378</u></b>	63.378	<b>72.571</b>

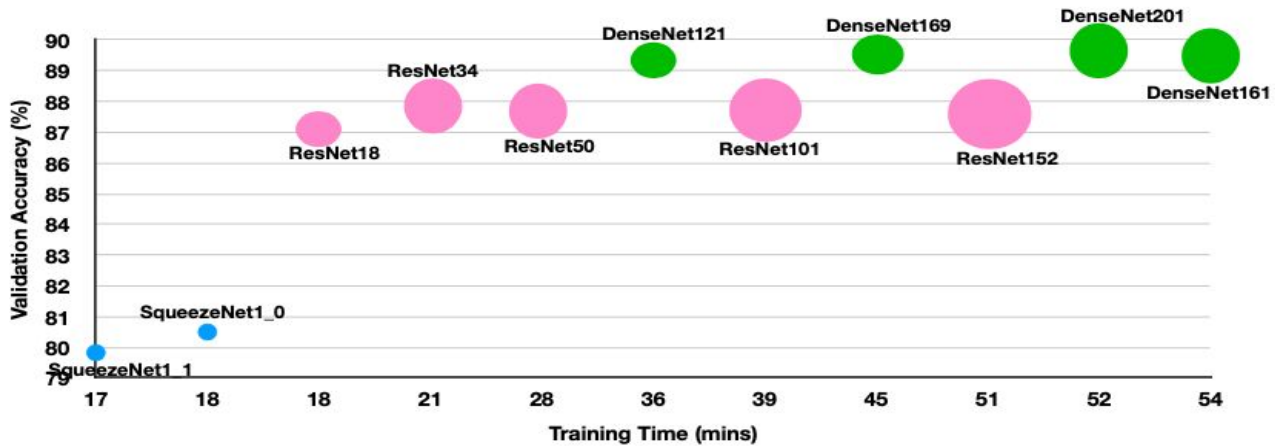
Table III: An accuracy comparison of SqueezeNet1\_0 on validation set & test set at different dropout at 90:10 and 70:30 data split. Bold values indicate the highest accuracy and the optimal option is highlighted.

Considering an accuracy of the baseline CNN on the validation set, at 0.3 dropout rate results in the highest accuracy, 72.140% at 90:10 data split and 72.96% at 70:30 data split, whereas on the test set at 0.5 dropout achieves the highest result, 71.378% at 90:10 data split and 72.571% at 70:30 data split. Similar to the 1st experiment on different data splits, it is essential to consider the accuracies on the validation set and test set all together at one dropout rate. Consequently, as shown in table III, the dropout rate at 0.3 contributes to the highest accuracies of SqueezeNet1\_0 after transfer learning and therefore, the remaining experiments also apply the 0.3 dropout rate. As for the most appropriate data proportion, despite 90:10 data split provides the highest accuracy in the 1st experiment, when considering different dropout 70:30 data split performs better in many cases. Therefore, for the rest of the experiments, 70:30 proportion is applied.

#### 4.3 3rd experiment

CNN model	Accuracy (%) on validation set	Training time (mins)
<i>SqueezeNet1_0</i>	80.501	18.021
SqueezeNet1_1	79.827	<b>17.005</b>
ResNet18	87.164	17.672
ResNet34	87.725	20.593
ResNet50	87.640	28.181
ResNet101	87.718	39.399
ResNet152	87.576	50.891
DenseNet121	89.391	35.951
DenseNet161	89.557	53.737
DenseNet169	89.443	44.912
DenseNet201	<b>89.722</b>	51.691

Table IV: An accuracy and training time comparison of 11 pretrained CNNs after transfer learning. Bold values indicate the best



performance and the optimal options are highlighted.

Figure 4: An accuracy and training time comparison of 11 pretrained CNNs after transfer learning

As demonstrated in figure 4, it can be seen that DenseNet models, especially DenseNet201, outperform the other CNN models after applying the transfer learning technique. Nonetheless, if considering only training time, SqueezeNet models spend less computational time due to their smaller model sizes which are illustrated in smaller circles. Having said this, when taking both accuracy and training time into account, it seems that DenseNet121 is one of the most potential models for hostel image database. Therefore, in further experiments, DenseNet121 and DenseNet201 are in focus.

#### 4.4 4th experiment

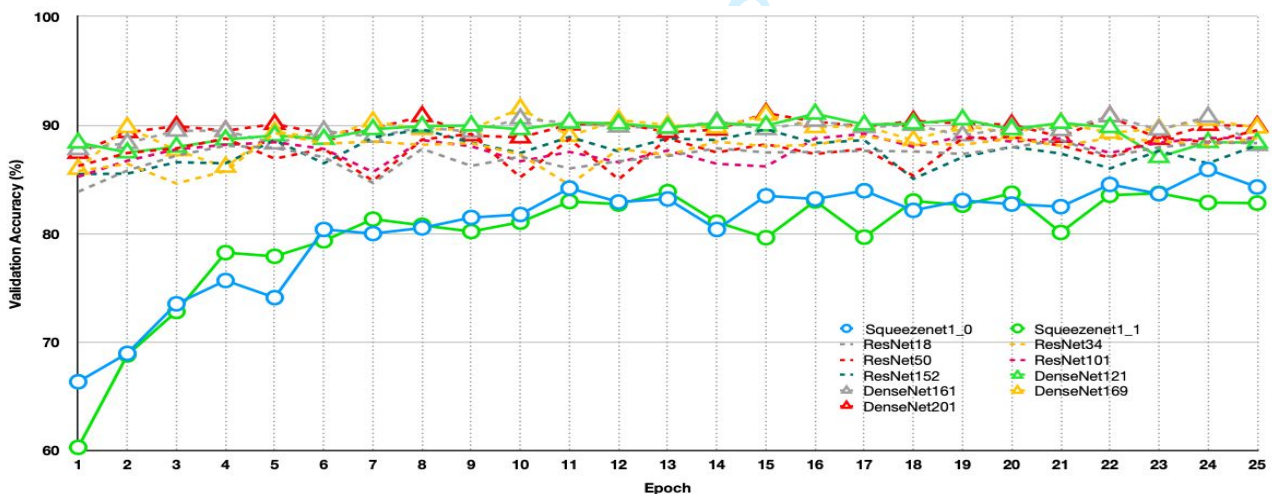


Figure 5: An accuracy comparison of 11 pretrained CNNs after transfer learning at different epoch

When testing pretrained CNNs at different epoch, figure 5 shows that DenseNet models and ResNet models provide similar and stable performance across each of these models, whereas SqueezeNet models start to perform better at six epochs. After this, their accuracy improves slightly and become more stable. Therefore, it can be seen that the CNN models could achieve high accuracies with only a few iterations. Having said this, in order to continue investigating SqueezeNet models at their best performances, including SqueezeNet1\_0 as a baseline CNN, the rest of the experiments are conducted at 25 epochs.

#### 4.5 5th experiment

CNN model	Accuracy (%)	
	Validation set	Test set
<i>SqueezeNet1_0</i>	80.501	81.429
SqueezeNet1_1	79.827	82.000
ResNet18	87.164	<b>88.571</b>
ResNet34	87.725	88.286
ResNet50	87.540	83.429
ResNet101	87.718	86.857
ResNet152	87.576	83.143
DenseNet121	89.391	83.714
DenseNet161	89.557	83.429
DenseNet169	89.433	85.714
DenseNet201	<b>89.722</b>	86.857

Table V: An accuracy comparison of 11 pretrained CNNs after transfer learning on validation set & test set. Bold values indicate the highest accuracy and the optimal options are highlighted.

It can be seen in table V that in most CNNs simulations the accuracies on validation set are higher than the accuracies on the test set. However, four models, SqueezeNet1\_0, SqueezeNet1\_1, ResNet18, and ResNet34, have the opposite results. It could be assumed that the lesser deep layers, the more it becomes overfitting. Nonetheless, the accuracies of DenseNet121 and DenseNet201 on both validation set and test set are still considerably high, 89.391% and 83.714% for DenseNet121 and 89.722% and 86.857% for DenseNet201, which reassure their potential for hostel image classification with transfer learning.

#### 4.6 6th experiment

CNN model	Original size database			Resized database		
	Accuracy (%)		Training time (mins)	Accuracy (%)		Training time (mins)
	Validation set	Test set		Validation set	Test set	
<i>SqueezeNet1_0</i>	76.260	81.429	7.208	71.200	77.143	3.999
SqueezeNet1_1	76.059	82.000	<b>6.802</b>	75.353	77.143	<b>3.726</b>
ResNet18	86.564	<b>88.571</b>	7.069	87.173	86.571	4.007
ResNet34	87.237	88.286	8.237	87.728	87.143	4.858
ResNet50	87.284	83.429	11.272	87.338	86.857	8.209
ResNet101	87.307	86.857	15.760	87.071	81.143	12.118
ResNet152	87.337	83.143	20.356	87.484	84.857	16.140
DenseNet121	88.932	83.714	14.380	88.557	<b>88.571</b>	9.954
DenseNet161	89.145	83.429	21.495	<b>90.048</b>	87.429	17.143
DenseNet169	88.876	85.714	17.965	89.419	85.714	12.505
DenseNet201	<b>89.376</b>	86.857	20.676	89.296	86.286	15.180

Table VI: An accuracy comparison of 11 pretrained CNNs on original size database & resized database at 10 epochs. Bold values indicate the best performance and the optimal options are highlighted.

Due to different image data sources, gathered hostel images come in various image sizes. This 6th experiment tests another hostel image database, which is the same database in the previous experiments but reduced its image sizes to one image size (300x300) prior and in addition to data augmentation in data pre-processing



stage. Table VI shows that accuracies on validation set are not affected much by resizing the database and remain similar in most CNN cases. Having said this, accuracies on the test set improve slightly in some CNN models, including DenseNet121 which is increased from 83.714% to 88.571%. Furthermore, the computational time is significantly decreased on the resized database for all CNN models.

#### 4.7 7th experiment

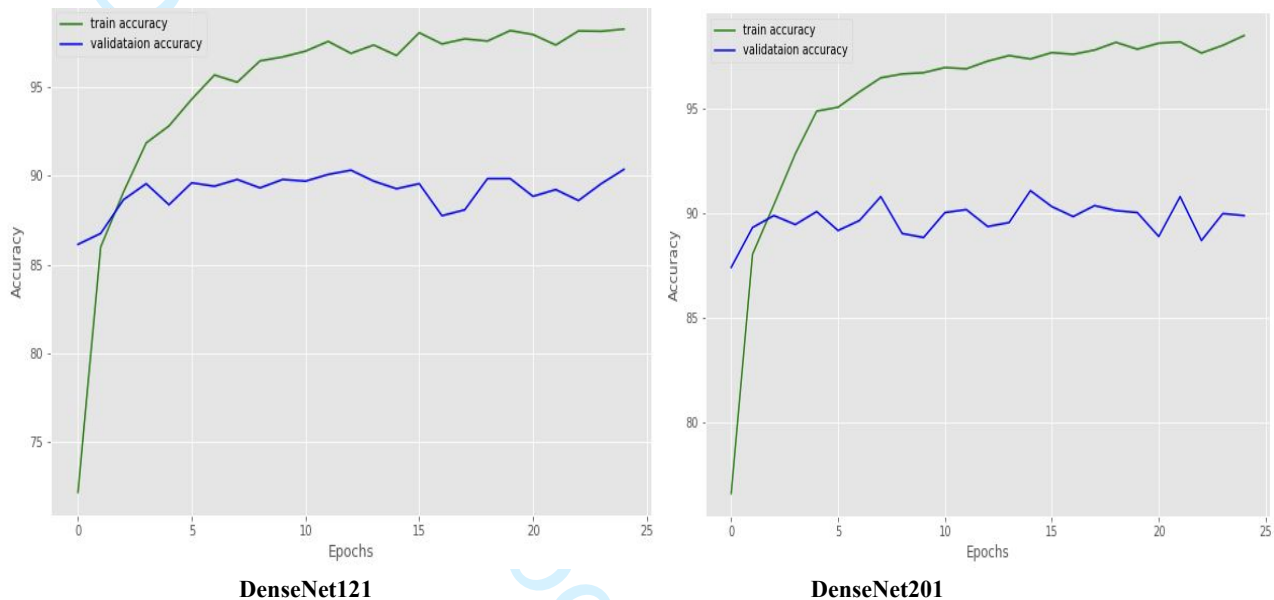


Figure 6: An accuracy comparison of DenseNet121 & DenseNet201 on training set & validation set

Figure 6 illustrates the plots of the accuracy of DenseNet121 and DenseNet201 on both the training set, green colour, and validation set, blue colour. It is clear that validation accuracies remain stable at around 90% for both models. That means the models perform well and are not overfitting on the current database size. Therefore, the database size could be increased in order to potentially increase the accuracy of hostel image classification.

## 5. Conclusion and future directions

### 5.1 Conclusion

In response to the fast-growing digital image collections on online platforms and the emerging transfer learning technology for image classification, as well as the impacts of coronavirus pandemic on hostel market, this study tested the potential of 11 pretrained CNNs with transfer learning technique for hostel image classification on the first hostel image database which is created for this project. The results from seven experiments highlight that, with the optimization of parameter fine-tuning, various pretrained CNNs with transfer learning can provide good performance. Particularly, DenseNet121 and DenseNet201 outperform the other CNNs, in terms of accuracy rate and training time, and achieve impressive outcomes for hostel image classification. Different elements have been explored in order to find the optimal combination and achieve the highest classification performance. SqueezeNet1\_0 is selected as a baseline CNN due to its most compact model size, the smallest number of deep layer, and less number of parameter. In terms of the data split during model training, it shows that the training set and the validation set should be split into 70% and 30%, respectively. The dropout rate at 0.3 contributes to the highest accuracy during training and relatively high during testing.

When all 11 pretrained CNNs are considered, the classification accuracy can reach its peak with only a few iterations and become stable for DenseNet models and ResNet models, whereas SqueezeNet models reach 80% accuracy at six epochs and continue to improve slightly. DenseNet201 best-performs during training and relatively similar during testing, while DenseNet121 demonstrates competitive results. Furthermore, when the size of the hostel image database is reduced, in addition to the data augmentation, the computational time is significantly decreased, even though the classification accuracies only slightly increase. Based on these results, it can be seen that, by applying this state-of-the-art transfer learning technique in pretrained CNNs for image classification, hostel image managers could automatically organize their growing collections and fasten the image indexing task with high accuracy, less labour cost, and less human errors on this accommodation type which has complex clutters of content on an image. Additionally, this optimal solution could contribute to the improved services/products of the hostel sector by utilizing the correctly classified images in database and offering more targeted solutions to their current or potential customers which could lead to a better position to recover from the current global economic downturn.

## 5.2 Future directions

Despite the promising results from this study, it is undeniable that there is plenty of room to further investigate. Suggestions for future directions are stated as follow.

### A. CNN layer modification

Due to the operation of the latter layers of pretrained CNNs which is to extract high-level image features, in addition to fine-tuning parameters to match new image classification task, further modification on some of these layers could be explored in order to compare its results with the fine-tuned only version of the pretrained CNNs. Consequently, the classification accuracy could be improved, as well as the computational cost.

### B. PlacesCNN models comparison

Even though available pretrained CNN models were originally trained on ImageNet database which contains object-centric image data, Zhou *et al.* (2017) gathered Places365 database, which contains a large-scale scene-centric image data, and trained various CNNs from scratch with this database without changing network architectures. Therefore, by comparing the results from pretrained Places365-CNNs with the original pretrained CNNs could benefit future experiments on hostel image classification task.

### C. Building small size CNN model

Since smartphone users are expected to increase from 2.5 billion in 2016 to 3.8 billion in 2021 (O'Dea, 2021<sup>9</sup>), small size CNN model with good performance like SqueezeNet models could be created in order to facilitate mobile applications on image classification task and other computer vision tasks in the future.

### D. Video classification

Apart from a vast collection of digital images on online platforms, it is undeniable that video is another data type that is enormously created in today's society. TankovskaClement (2021<sup>9</sup>) reports that, as of May 2019, more than 500 hours of video were uploaded to YouTube in every minute and this number of video content

hours grew by around 40% between 2014 and 2019. Therefore, video classification could be further researched to accommodate this continuously increasing demand.

#### E. Image retrieval application

As the image classification process is a part of image retrieval task, it is crucial to progress to the image retrieval step in order to gain greater benefits. In particular for hostel search by image, can also be called content-based image retrieval, the hostel image managers could practically take advantage of the results from this study and further apply the state-of-the-art knowledge and technology to their real-life business operation.

#### References

- Abdullahzadeh, A. and Mohanna, F. (2013), "Content-based image retrieval based on affine noisy invariant colour region", *International research journal of applied and basic sciences*, Vol. 6 No. 5, pp.598-606.
- Amato, G., Falchi, F. and Gennaro, C. (2015), "Fast image classification for monument recognition", available at: <https://dl.acm.org/doi/pdf/10.1145/2724727> (accessed ~~31 June~~ ~~May~~ 2021~~0~~)
- Aslam, S. (2021a), "Facebook by the numbers: stats, demographics & fun facts", available at: <https://www.omnicoreagency.com/facebook-statistics/> (accessed ~~31 June~~ ~~February~~ 2021)
- Aslam, S. (2021b), "Instagram by the numbers: stats, demographics & fun facts", available at: <https://www.omnicoreagency.com/instagram-statistics/> (accessed ~~31 June~~ ~~February~~ 2021)
- Bruegel Datasets (2020), "The fiscal response to the economic fallout from the coronavirus", available at: <https://www.bruegel.org/publications/datasets/covid-national-dataset/#belgium> (accessed ~~31 June~~ ~~January~~ 2021)
- Bunda, R. (2014), "The business of beds: an exploration hotel and hostel business strategy", available at: [https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=1358&context=srhonors\\_theses](https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=1358&context=srhonors_theses) (accessed ~~31 June~~ ~~May~~ 2021~~18~~)
- Businesswire (2020), "Impact of COVID-19 on the global hostels market 2020-2030", available at: <https://www.businesswire.com/news/home/20200615005443/en/Impact-of-COVID-19-on-the-Global-Hostels-Market-2020-2030---ResearchAndMarkets.com> (accessed ~~31 June~~ ~~January~~ 2021)
- ~~Clement, J. (2020), "Hours of video uploaded to YouTube every minute 2007-2019", available at: <https://www-statista-com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/> (accessed 12 October 2020)~~
- Dawud, A., Yurtkan, K. and Oztoprak, H. (2019), "Application of deep learning in neuroradiology: brain haemorrhage classification using transfer learning", *Computational intelligence and neuroscience*, Vol. 2019, pp.1-12.
- ~~Google Drive (2021), "Hostel Image Dataset 7.350.zip", available at: <https://drive.google.com/file/d/1byxXEIA1wvReSjIRW4k5Y1JFyuLnLrHy/view?usp=sharing> (accessed 3 June 2021)~~
- Greater London Authority (2020), "Mayor offers funding to make homeless hostels Covid19 safe", available at: <https://www.london.gov.uk/press-releases/mayoral/mayor-offers-funding-to-make-hostels-covid-safe> (accessed ~~31 June~~ ~~October~~ 2021~~0~~)
- He, K., Zhang, X., Ren, S. and Sun, J. (2015), "Deep residual learning for image recognition", available at: <https://arxiv.org/pdf/1512.03385.pdf> (accessed ~~31~~ ~~2~~ ~~June~~ ~~October~~ 2021~~0~~)
- Huang, G., Liu, Z. and Maaten, L. (2018), "Densely connected convolutional networks", available at: <https://arxiv.org/pdf/1608.06993.pdf> (accessed ~~31~~ ~~2~~ ~~June~~ ~~October~~ 2021~~0~~)

- 1  
2 | ImageNet (2020), "Large scale visual recognition challenge", available at: <http://image-net.org/challenges/LSVRC/> (accessed ~~3+~~  
3 | ~~June~~~~October~~ 2021~~0~~)  
4  
5  
6 | Kanchinadam, T. (2016), "Hotel image classification using deep learning", available at:  
7 | [https://tkanchin.github.io/images/ml\\_report.pdf](https://tkanchin.github.io/images/ml_report.pdf) (accessed 12 October 2020)  
8  
9 | Kaggle (2016), "Hotel image classification", available at: <https://www.kaggle.com/c/hotel-image-classification/overview>(accessed  
10 | ~~3+2~~ ~~June~~~~October~~ 2021~~0~~)  
11  
12 | Landola, F., Han, S., Moskewicz, M., Ashraf, K., Dally, W. And Keutzer, K. (2016), "SqueezeNet: AlexNet-level accuracy with 50x  
13 | fewer parameters and <0.5MB model size", available at: <https://arxiv.org/pdf/1602.07360.pdf> (accessed ~~3+2~~ ~~June~~~~October~~ 2021~~0~~)  
14  
15 | Li, F., Kant, S., Araki, S., Bangerla, S. and Shukla, S. (2020), "Neural networks for fashion image classification and visual  
16 | search", available at: <https://arxiv.org/abs/2005.08170> (accessed ~~3+5~~ ~~June~~~~ly~~ 2021~~0~~)  
17  
18 | Mason, P. (2014), *Researching tourism, leisure and hospitality for your dissertation*, Goodfellow, Oxford, UK.  
19  
20 | Mateen, M., Wen, J., Nasrullah, N., Sun, S. and Hayat, S. (2020), "Exudate detection for diabetic retinopathy using pretrained  
21 | convolutional neural networks", *Complexity*, Vol. 2020, pp.1-11.  
22  
23 | MathWorks (2021~~0~~), "Pretrained deep neural networks", available at:  
24 | <https://uk.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html> (accessed ~~3+~~ ~~June~~~~October~~ 2021~~0~~)  
25  
26  
27 | O'Dea, S. (2021~~0~~), "Number of smartphone users worldwide from 2016-2026~~+~~", available at:  
28 | <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> (accessed ~~3+2~~ ~~June~~~~October~~ 2021~~0~~)  
29  
30 | OECD (2020), "Rebuilding tourism for the future: COVID-19 policy responses and recovery", available at:  
31 | [http://www.oecd.org/coronavirus/policy-responses/rebuilding-tourism-for-the-future-covid-19-policy-responses-and-recovery-](http://www.oecd.org/coronavirus/policy-responses/rebuilding-tourism-for-the-future-covid-19-policy-responses-and-recovery-bced9859/)  
32 | [bced9859/](http://www.oecd.org/coronavirus/policy-responses/rebuilding-tourism-for-the-future-covid-19-policy-responses-and-recovery-bced9859/) (accessed ~~3+~~ ~~June~~~~February~~ 2021)  
33  
34 | Premchaiswadi, W. et al, Tungkatsathan, A. and Premchaiswadi, N. (2010), "Mobile image search for tourist information using  
35 | ACC algorithm", available at:  
36 | [https://ieeexplore.ieee.org/abstract/document/5671757?casa\\_token=9U9V4B9AdmcAAAAA:NIkmA12MV50anmxH-](https://ieeexplore.ieee.org/abstract/document/5671757?casa_token=9U9V4B9AdmcAAAAA:NIkmA12MV50anmxH-RIMcrkAFti80mFyXILgf_-NIXQMeN27De1vG31_cWDLfHCQ16XqMLg)  
37 | [RIMcrkAFti80mFyXILgf\\_-NIXQMeN27De1vG31\\_cWDLfHCQ16XqMLg](https://ieeexplore.ieee.org/abstract/document/5671757?casa_token=9U9V4B9AdmcAAAAA:NIkmA12MV50anmxH-RIMcrkAFti80mFyXILgf_-NIXQMeN27De1vG31_cWDLfHCQ16XqMLg) (accessed ~~3+~~ June 2021~~0~~)  
38  
39  
40 | Raisi, Z., Mohanna, F. and Rezaei, M. et al. (2011), "Content-based image retrieval for tourism application", available  
41 | at:[https://ieeexplore.ieee.org/abstract/document/6121857?casa\\_token=MF2x6Fvh3J4AAAAA:bjgdsbZ4jshw6PiK54QChqQWK](https://ieeexplore.ieee.org/abstract/document/6121857?casa_token=MF2x6Fvh3J4AAAAA:bjgdsbZ4jshw6PiK54QChqQWK)  
42 | [KsusDC7vAUG5cJIZTxUUAA5diCORIfknpwqLgd9WeRpFEM](https://ieeexplore.ieee.org/abstract/document/6121857?casa_token=MF2x6Fvh3J4AAAAA:bjgdsbZ4jshw6PiK54QChqQWK) (accessed ~~3+~~ June 2021~~0~~)  
43  
44 | Raisi, Z., Mohanna, F. and Rezaei, M. et al. (2014), "Applying content-based image retrieval techniques to provide new service  
45 | for tourism industry", *International journal advanced networking and applications*, Vol. 6 No. 2, pp.2222-2232.  
46  
47  
48 | Rath, S. (2020), "Getting 95% accuracy on the caltech101 dataset using deep learning", available at:  
49 | <https://debuggercafe.com/getting-95-accuracy-on-the-caltech101-dataset-using-deep-learning/> (accessed ~~3+~~ June 2021~~0~~)  
50  
51 | Rath, T. (2018), "Hotel image categorization with deep learning", available at [https://medium.com/kayak-tech/hotel-image-](https://medium.com/kayak-tech/hotel-image-categorization-with-deep-learning-ffa8429e55b5)  
52 | [categorization-with-deep-learning-ffa8429e55b5](https://medium.com/kayak-tech/hotel-image-categorization-with-deep-learning-ffa8429e55b5) (accessed ~~3+2~~ ~~June~~~~October~~ 2021~~0~~)  
53  
54 | Ren, M., Vu, H., Li, G. and Law, R. (2020), "Large-scale comparative analyses of hotel photo content posted by managers and  
55 | customers to review platforms based on deep learning: implications for hospitality marketers", available at  
56 | <https://www.tandfonline.com/doi/pdf/10.1080/19368623.2020.1765226?needAccess=true&> (accessed ~~3+2~~ ~~June~~~~October~~ 2021~~0~~)  
57  
58 | Sabharwal, A. (2017), "500 million people using Google Photos and three new ways to share", available at:  
59 | <https://blog.google/products/photos/google-photos-500-million-new-sharing/>(accessed ~~3+~~ ~~June~~~~May~~ 2021~~0~~)  
60

1  
2 Shaha, M. and Pawar, M. (2018), "Transfer learning for image classification", *the 2nd international conference on electronics,*  
3 *communication and aerospace technology*, IEE conference record, pp.656-660.

4  
5 Stadlen, A. (2019), "Update on creative commons licenses and in memoriam accounts", available at:  
6 <https://blog.flickr.net/2019/03/08/update-on-creative-commons-licenses-and-in-memoriam-accounts/> (accessed 1 May 2020)  
7  
8

9  
10  
11 Stylianou, A., Xuan, H., Shende, M., Brandt, J., Souvenir, R. and Pless, R. (2019), "Hotels-50K: a global hotel recognition  
12 dataset", available at: <https://www.aaai.org/ojs/index.php/AAAI/article/view/3863> (accessed ~~3+5 June~~ 2021)  
13

14  
15 Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. and Rabinovich, A. (2015), "Going deeper with convolutions",  
16 paper presented at the IEEE conference on computer vision and pattern recognition, available  
17 at: [https://www.cvfoundation.org/openaccess/content\\_cvpr\\_2015/html/Szegedy\\_Going\\_Deeper\\_With\\_2015\\_CVPR\\_paper.html](https://www.cvfoundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html) (acce  
18 ssed 1 October 2020).  
19

20  
21 Tasli, E. (2017), "Automated image tagging at Booking.com", available at: [https://booking.ai/automated-image-tagging-at-booking-](https://booking.ai/automated-image-tagging-at-booking-com-7704f27dcc8b)  
22 [com-7704f27dcc8b](https://booking.ai/automated-image-tagging-at-booking-com-7704f27dcc8b) (accessed ~~3+2 June~~ October 2021)

23 [TankovskaClement, HJ. \(20210\), "Hours of video uploaded to YouTube every minute 2007-2019", available at:](https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/)  
24 <https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/> (accessed ~~3+2 June~~ October 2021)  
25

26 Wallace, G. K. (1991), "The JPEG still picture compression standard", *Communications of the ACM*, Vol. 34 No. 4, pp.31-45.  
27

28 Wengert, C., Douze, M. and Jegou, H. (2011), "Bag-of-colours for improved image search", available at:  
29 <https://dl.acm.org/doi/pdf/10.1145/2072298.2072034> (accessed ~~3+ June~~ May 2021)  
30

31  
32 Xiao, T., Liu, L., Li, K., Qin, W., Yu, S. and Li, Z. (2018), "Comparison of transferred deep neural networks in ultrasonic breast  
33 masses discrimination", *Biomed research international*, Vol. 2018, pp.1-9.  
34

35  
36 Zhao, B. Huang, B. and Zhong, Y. (2017), "Transfer learning with fully pretrained deep convolution networks for land-use  
37 classification", *IEEE geoscience and remote sensing letters*, Vol. 14 No. 9, pp.1436-1440.  
38

39 Zheng, L., Wang, S. and Tian, Q. (2014), "Coupled binary embedding for large-scale image retrieval", *IEEE transactions on image*  
40 *processing*, Vol. 23 No. 8, pp.3368-3380.  
41

42  
43 Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A. (2017), "Places: a 10 million image database for scene recognition",  
44 working paper, Computer science and artificial intelligence laboratory, Massachusetts Institute of Technology, USA.  
45

46  
47 Zhu, L., Shen, J., Jin, H., Zheng, R. and Xie, L. (2015), "Content-based visual landmark search via multimodal hyper graph  
48 learning", *IEEE transactions on cybernetics*, Vol. 45 No. 12, pp.2756-2769.  
49

50 .  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60