Intelligence as an add-on to mobiles through image assistance

Project guide

Dr. Easwara Kumar K S

Team members

2014103572 Ananda Kumar P

2014103567 Vijay S P

2014103556 Roshni Rajan

Signature of the guide

TABLE OF CONTENTS

Abstract

Since the rise of mobile age, the whole world has invested immense time and research value to use mobiles as assistant for people. As new hardware technologies emerged, it paved way for assistance via text, voice etc. Current trend in mobile assistance are voice support for searching, information retrieval and system actions. We have a strong belief that images will also can be used for providing assistance and such a technology is not far from reality as tech giants are currently focussing on it. Our project focuses on applying Convolutional Neural Networks in recognising objects and providing useful information regarding them from a picture taken by mobile camera in real time. This project also focuses on building a model that withstands the time and new objects.

Base papers

1. C. McCool, T. Perez and B. Upcroft, "Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics," in *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1344-1351, July 2017.doi: 10.1109/LRA.2017.2667039 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7849167&isnumber=7875382

2. H. J. Jeong, M. J. Lee and Y. G. Ha, "Integrated Learning System for Object Recognition from Images Based on Convolutional Neural Network," *2016 International Conference on Computational Science and Computational Intelligence*, Las Vegas, NV, 2016.
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7881453&isnumber=7881293

3. Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai and E. I. C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence,2014.
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6853873&isnumber=6853544

Introduction

Image processing is the analysis and manipulation of a digitized image, especially in order to improve its quality for further processing or extracting features from the image.

Object detection and recognition from the extracted features using Convolutional neural networks.

In machine learning, a **convolutional neural network** (**CNN**, or **ConvNet**) is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery.

Problem Statement

Object identification and recognition by categorization or name. Providing the user some useful information about all those objects. Should be compatible with mobile applications and work in real time.

Overall Objective

1. Identifying all useful sections of the picture. (identifying all objects in the image)
2. Retrieving useful information from them. (recognising each of them and finding some useful information from sources like google)

Issues

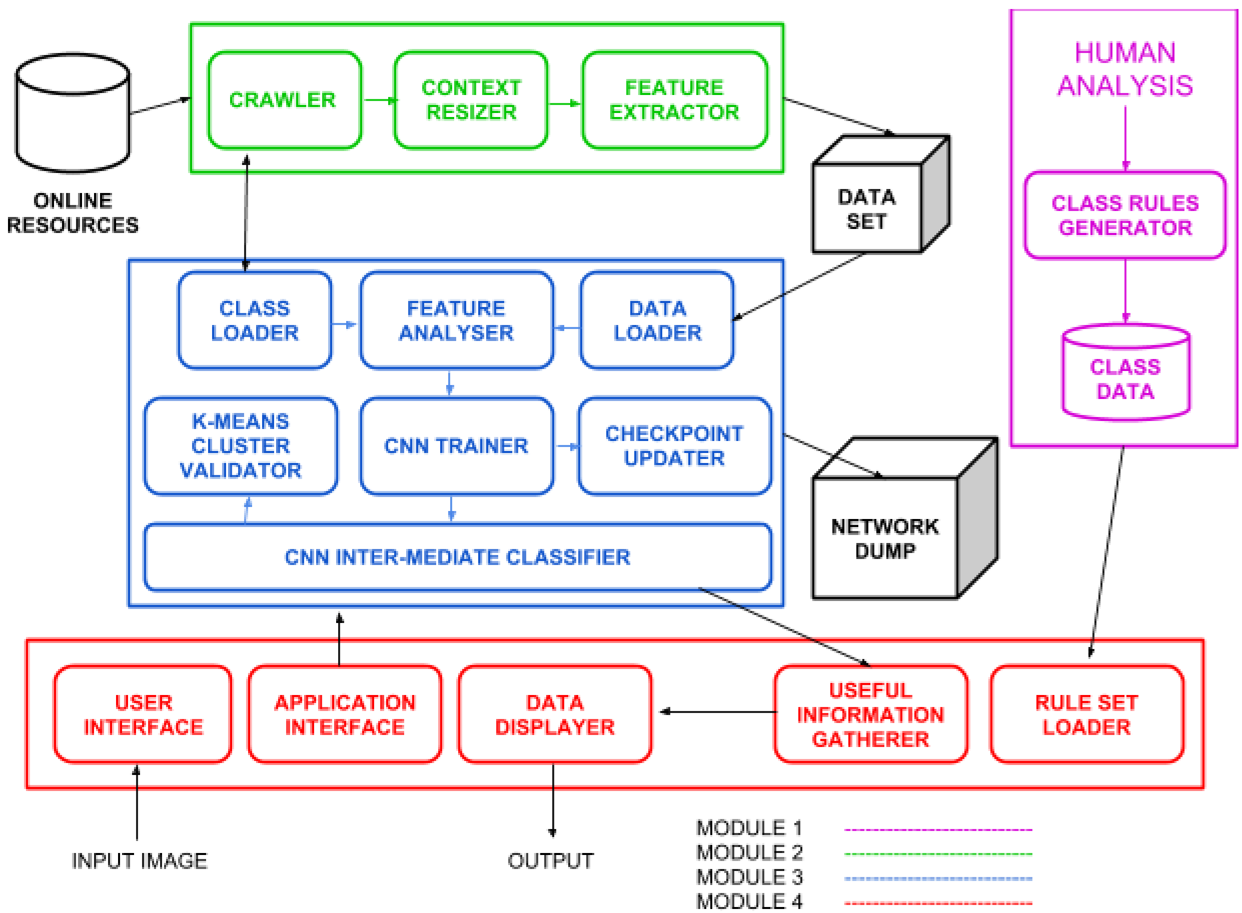
1. Handling large number of classes for categorizing is difficult.
2. The success of the project involves extensive training for each category. This will take high computational power and more time.
3. The algorithms must be fast enough to run real time.

Related work

There has been many significant studies that have aimed to achieve object recognition. Several recent papers pave proposed way of using depth information[11,12]. Paper [11] suggested depth kernel descriptor and paper [12] proposed 3D SURF algorithm to makes object tridimensional. Both are good design but these algorithms cause incorrect result because simply processed image data has latent erroneous information. To solve this problem, many researches proposed deep learning for object recognition. Research [3] proposed image analysis system in multiple instance using deep learning.

From the point of view of performance, like an accuracy and recognizing time, many designs are proposed. Paper [8] designed Content-based Image Retrieval (CBIR) system with deep learning. It trains deep CNN models using ImageNet and adopt model in new domain for CBIR. In the CNN algorithm, there are many methods in the pooling process. Related work [9] shows max-pooling is a fastest method. Research[10] proposed three-dimensional image recognition system with RGB-D deep learning that trains image with RGB, Depth and Grayscale. It shows a great performance. Through these studies, it is fact that CNN deep learning makes possible accurate and fast image recognition.

Block diagram



Modules split-up

1. Module 1 (Analysis module)

    1.1. Classes identification
    1.2. Information gathering
    1.3. Generating rulesets

2. Module 2 (Preprocessing module)

    2.1. Crawler
    2.2. Context resizer
    2.3. Feature extractor

3. Module 3 (Classifier module)

    3.1. Class loader
    3.2. Data loader
    3.3. Feature analyser
    3.4. Custom K-NN algorithm
    3.5. CNN trainer
    3.6. Checkpoint updater
    3.7. CNN intermediate classifier

4. Module 4 (User view module)

    4.1. User interface
    4.2. Application interface
    4.3. Ruleset loader
    4.4. Useful information gatherer
    4.5. Data displayer

Proposed system

1. Module 1 (Analysis module)

    1.1. Classes identification
    This part is for confirming the environment in which we will focus and dataset confirmation.
    1.2. Information gathering

Identify the various objects possible and make them as classes.

1.3. Generating rulesets
Input: object name
Output: class with rulesets for input object

2. Module 2 (Preprocessing module)

2.1. Crawler
Input: class name or object name
Output: images crawled from google for such class or object

2.2. Context resizer
Input: images of different resolutions and background
Output: normalised images of same size and clipped background

2.3. Feature extractor
Input: normalised images
Output: a vector of extracted features for each image with class label

3. Module 3 (Classifier module)

3.1. Class loader
Input: class name
Output: class with rulesets, next_step algo, subclass list and other necessary details

3.2. Data loader
Input: class
Output: dataset for the specified algorithm and subclass

3.3. Feature analyser
It analyses the feature set and finds the best subset for classifying and training

3.4. Custom K-NN algorithm
Input: input feature set and dataset from data loader
Output: final classified output class

3.5. CNN trainer
Input: class and feature sets
Output: trained model

3.6. Checkpoint updater

Input: existing model and small trained model
Output: a single integrated model

3.7. CNN intermediate classifier
Input: trained model and input feature set
Output: output class name after classification

4. Module 4 (User view module)

4.1. User interface
Input: a picture taken from mobile camera
Output: request to server or application

4.2. Application interface
This handles the data transfer between the mobile and external application server

4.3. Ruleset loader
From the intermediate result from application, we load the class rules for individual objects

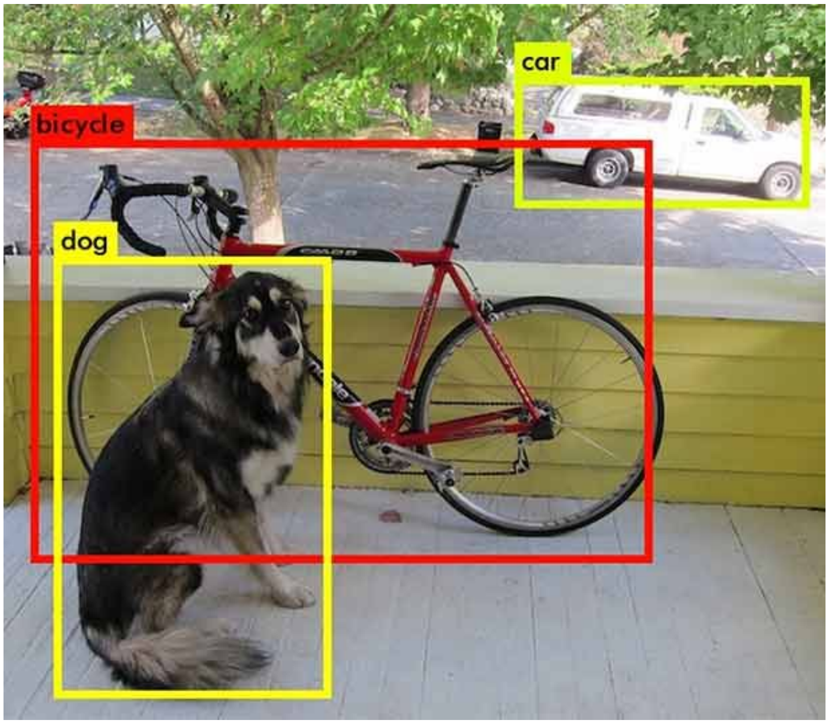
4.4. Useful information gatherer
Input: class name and rule sets
Output: information for each ruleset.

4.5. Data displayer
This displays the final output. Future versions amy also send accuracy details, user suggestions, feature for additional training.

Input output description

Input - pictures from mobile camera.
Output - list of all objects with additional information for each object.

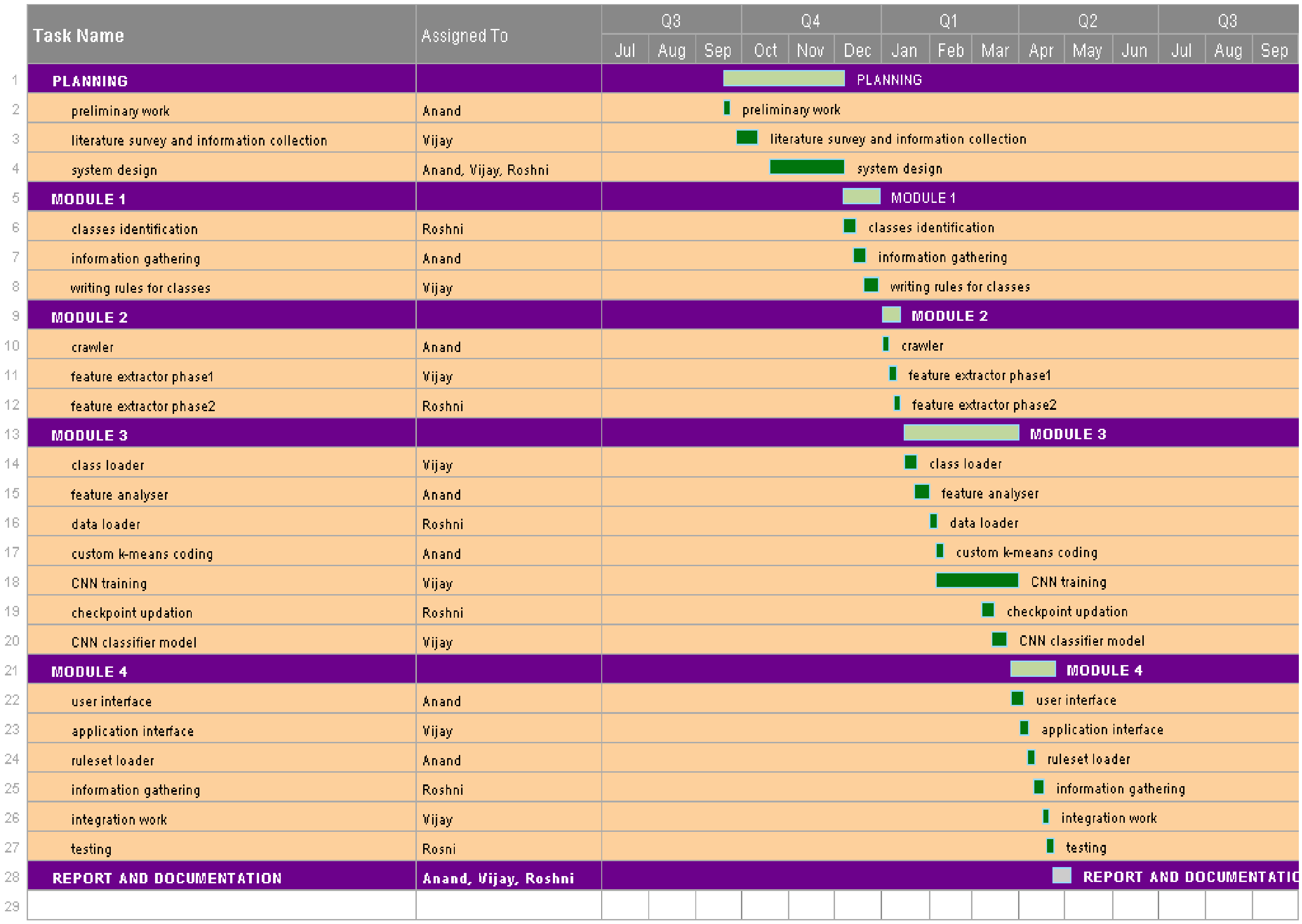
Pictorial representation



Output:
---dog---
Breed: german shepard
Age: 2 yrs
---bicycle---
Color: red
Condition: New
---car---
Color: white
Condition: Undecidable
Position: Stationary

Data sets: Only the sample data needs to be shown.

Gantt chart


Task Name
Assigned To
Q3
Q4
Q1
Q2
Q3
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
1 PLANNING
PLANNING
2 preliminary work Anand
preliminary work
3 literature survey and information collection Vijay
literature survey and information collection
4 system design Anand, Vijay, Roshni
system design
5 MODULE 1
MODULE 1
6 classes identification Roshni
classes identification
7 information gathering Anand
information gathering
8 writing rules for classes Vijay
writing rules for classes
9 MODULE 2
MODULE 2
10 crawler Anand
crawler
11 feature extractor phase1 Vijay
feature extractor phase1
12 feature extractor phase2 Roshni
feature extractor phase2
13 MODULE 3
MODULE 3
14 class loader Vijay
class loader
15 feature analyser Anand
feature analyser
16 data loader Roshni
data loader
17 custom k-means coding Anand
custom k-means coding
18 CNN training Vijay
CNN training
19 checkpoint updation Roshni
checkpoint updation
20 CNN classifier model Vijay
CNN classifier model
21 MODULE 4
MODULE 4
22 user interface Anand
user interface
23 application interface Vijay
application interface
24 ruleset loader Anand
ruleset loader
25 information gathering Roshni
information gathering
26 integration work Vijay
integration work
27 testing Rosni
testing
28 REPORT AND DOCUMENTATION Anand, Vijay, Roshni
REPORT AND DOCUMENTATIO
29

Literature Survey

PAPER TITLE	CONCEPTS/ ALGORITHMS	PROBLEMS DEALT	ADVANTAGES	DISADVANTAGES
C. McCool, T. Perez and B. Upcroft, "Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics," in *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1344-1351, July 2017.	1)Deep convolutional neural networks (DCNNs) 2)Model compression 3)Distillation techniques.	1)Trained deep convolutional neural networks (DCNNs) to tradeoff complexity (e.g., memory size and speed) with high accuracy. 2)Using adapted model (Adapted-IV3) to train a much lightweight DCNN. 3)Combining a set of K lightweight models as a mixture model to enhance the performance.	Weed segmentation is made easier by using model compression and distillation techniques in K-lightweight DCNN.	1)Difficulty in training a DCNN from limited data. 2)Pre-trained model (Inception-v3) object classification rate is slow.
H. J. Jeong, M. J. Lee and Y. G. Ha, "Integrated Learning System for Object Recognition from Images Based on Convolutional Neural Network," *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, 2016, pp. 824-828.	1)Deep Learning 2)Image Recognition 3)Convolutional Neural Network 4)Ontology 5)Crawler	1)Designing a proper CNN layer that perform fast and accurate learning. 2)Feature extraction from small area using CNN in loop. 3)Using Dropout concept for increasing accuracy on the test dataset	The Integrated object recognition system classifies the learned images with high accuracy and greater performance.	Increase of recognition rate and lack of memory space occurs when large sized images are given as input.
Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai and E. I. C. Chang, "Deep learning of feature	1)Deep learning 2)Feature learning 3)Supervised and unsupervised	1)Automatic extraction of feature representation and detailed annotation of medical objects is studied through deep	Compared accuracies of different feature representations on the dataset consisting of colon	Performance of unsupervised feature is slightly worse than supervised on unlabelled data.

representation with multiple instance learning for medical image analysis," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 1626-1630.	learning 4)Multiple instance learning (MIL) 5)K-means clustering	learning (DNN). 2)Comparison of different feature extraction methods like Manual Feature(MF),K-mean s,DNN1-F and DNN2-F	cancer histopathology images.	
X. Jia, "Image recognition method based on deep learning," *2017 29th Chinese Control And Decision Conference (CCDC)*, Chongqing, 2017, pp. 4730-4735.	1)Deep Learning 2)Convolutional Neural Network 3)Spatial Pyramid Pooling 4)Restricted Boltzman Machines 5)Autoencoder Parse Coding	1)Presenting a comprehensive review of deep learning. 2)Developing a categorization scheme to analyze the existing deep learning literature.	Comparison of various deep learning methods for image recognition is done.	Lack of training data and training of large networks effectively.
L. Liu, C. Shen and A. v. d. Hengel, "Cross-Convolution al-Layer Pooling for Image Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2305-2313, Nov. 1 2017.	1)Convolutional Neural Network 2)deep learning 3)Pooling 4)fine-grained object recognition	1)Cross-layer pooling is performed on the extracted densely sampled regions to create image representations from the activations of two consecutive convolutional layers of a pre-trained CNN. 2)Conducted experiments on popular image classification datasets and image retrieval datasets	The cross-layer pooling concept is applied to eliminate the computational cost of extracting features through image recognition.	The retrieval performance started to drop When 'k' value increases.
Y. Tsuzuki, K. Sawada, K. Hashimoto, Y. Nankaku and K.	1)Image recognition 2)Hidden Markov model	1)Image recognition is achieved based on log linear models (LLMs) using features extracted	Recognition rate is achieved higher for larger geometric variations	Image recognition takes more time for smaller geometric variations compared to other CNN based

Tokuda, "Image recognition based on discriminative models using features generated from separable lattice HMMS," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 2607-2611.	3)Separable lattice HMM(SL-HMMs) 4)Log linear model	from SL-HMMs. 2)Using SL-HMM this approach have solved the problem of degradation of recognition performance by geometric variations such as that in position and size of the object to be recognized.	compared to other CNN based methods.	methods.
H. M. Bui, M. Lech, E. Cheng, K. Neville and I. S. Burnett, "Object Recognition Using Deep Convolutional Features Transformed by a Recursive Network Structure," in *IEEE Access*, vol. 4, pp. 10059-10066, 2016.	1)Image Recognition 2)Deep learning 3)Recursive Neural Network 4)Feature extraction 5)Support Vector Machine (SVM)	Improving the AlexNet feature extraction technique by using a recursive neural network structure on features extracted by a deep convolutional neural network pre-trained on a large data set.	1)Higher recognition accuracy is achieved with lower computational cost and structural simplicity. 2)This approach requires no training during the feature extraction stage, and can be performed very efficiently.	The accuracy gets slightly decreased when softmax classifier is used.

References

[1] C. McCool, T. Perez and B. Upcroft, "Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics," in *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1344-1351, July 2017.

[2] H. J. Jeong, M. J. Lee and Y. G. Ha, "Integrated Learning System for Object Recognition from Images Based on Convolutional Neural Network," *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, 2016, pp. 824-828.

[3] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai and E. I. C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 1626-1630.

[4] X. Jia, "Image recognition method based on deep learning," *2017 29th Chinese Control And Decision Conference (CCDC)*, Chongqing, 2017, pp. 4730-4735.

[5] L. Liu, C. Shen and A. v. d. Hengel, "Cross-Convolutional-Layer Pooling for Image Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2305-2313, Nov. 1 2017.

[6] Y. Tsuzuki, K. Sawada, K. Hashimoto, Y. Nankaku and K. Tokuda, "Image recognition based on discriminative models using features generated from separable lattice HMMS," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 2607-2611.

[7] H. M. Bui, M. Lech, E. Cheng, K. Neville and I. S. Burnett, "Object Recognition Using Deep Convolutional Features Transformed by a Recursive Network Structure," in *IEEE Access*, vol. 4, pp. 10059-10066, 2016.

[8] Wan, Ji, et al. "Deep learning for content-based image retrieval: A comprehensive study." Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.

[9] Giusti, Alessandro, et al. "Fast image scanning with deep max-pooling convolutional neural networks." arXiv preprint arXiv:1302.1700 (2013).

[10] Socher, Richard, et al. "Convolutional-recursive deep learning for 3d object classification." Advances in Neural Information Processing Systems. 2012.

[11] Bo, Liefeng, Xiaofeng Ren, and Dieter Fox. "Depth kernel descriptors for object recognition." 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2011.

[12] Redondo-Cabrera, Carolina, et al. "Surfing the point clouds: Selective 3d spatial pyramids for category-level object recognition." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.