Automated Catalog Management and Image Quality Assessment using Convolution Neural Networks and Transfer Learning

Mani Garlapati and Souradip Chakraborty

I. Abstract

Catalogue management is a very important aspect in the field of ecommerce as it helps the visitors in efficiently selecting the necessary interest items. In every retail website, all the items in the catalogue are in a particular order of different categories. In this work, we have developed an entire pipeline where the first task to automatically classify the various orientations (front view, side view, top view etc.) of the images sent by the vendor using Transfer learning. In the second part of our pipeline, we have eased the process of catalogue management with the image quality assessment of the vendor images using Image processing and Transfer Learning. Finally, the automatic ordering of items is done as per requirements.

Keywords: Convolution Neural Networks, Transfer Learning, Image quality assessment, Structural similarity index

II. Introduction

Efficient Catalogue management is very important and vital for ecommerce retailers since it helps online visitors in selecting the necessary items and if the catalogues are well organized it serves as a great aid for the customers which help them in turning to loyal customers. Many research works have been done in the field of image classification using convolution neural network [1] and Transfer learning [2], but very few works have been done using a combination of both in classification of various orientations (different views like side view, front view etc..) of images of items sent by vendors which is being done as a part of catalogue management in this work and parameter tuning has been done using Bayesian optimization [3] and results have been compared with a baseline model. Since manual ordering of the images sent by vendors is being done in majority of industries currently which is extremely time-consuming and hence it can be improved vastly by the above way. Secondly, quality of the image sent by the vendors plays a crucial part since improper image quality is a major part of customer dissatisfaction [4]. Structural similarity index [5] has been considered as an index in this case for assessment of the quality of images of items sent by vendor. In this case, one of the challenges have been blurring and its various types since in many cases vendors send blurred images which is one primary cause for customer dissatisfaction and hence by the image quality assessment it can automatically be detected which of the images are below a certain quality level and further actions will be taken on that [6]. Another challenge was that in many of the items, the available amount of data is not very high and hence the methodology has been developed keeping that constraint in mind.

The next part of the paper explains the dataset considered and the detailed methodology of Automated catalogue management and Image quality assessment with the results. Phase 2 describes the dataset that has been used for training. Phase 3 describes the baseline Histogram of oriented gradients feature based model and the Convolution Neural Network and Transfer Learning (with pre-trained Convolution Neural Network) based model taken up in this work for image orientation classification (Front view, Side view and Back view).In the same phase the model hyper parameters have been optimized using Bayesian Optimization. Phase 4 describes the image quality assessment of the images sent by the vendors using Structural similarity index and Transfer Learning.

III. Dataset Description

The dataset that has been used for image orientation classification consists of 3 classes- Front view, Side view and Top view and the size has been kept low to meet the constraints mentioned earlier. The dataset consists of 312 images in total out of which 95 of back view, 108 of front view and 109 of side view images have been used to train. The challenge was to show good accuracy even with small datasets.

IV. Image Orientation Classification using Convolution Neural Network and Transfer Learning

A. Histogram of Oriented Gradients as Baseline Model:

For implementation of the task of classification of image into one of the 3 categories, the baseline model that has been used is with the histogram of oriented gradients features as it has been used in many places where image orientation classification is the prime objective [7]. Since the primary concern is to classify different orientation so it makes sense to use Histogram of Oriented gradients features as shown in Figure 2.

Using Histogram of oriented gradient features as predictors, 5 different classification models were fitted to the training data and for each of the models, the ideal hyper parameters were computed using Bayesian Optimization of hyper parameters

Mani Garlapati (e-mail: mani.garlapati@walmart.com)
Souradip Chakraborty (e-mail: Souradip.Chakraborty@walmartlabs.com)

Fig. 1. Back, Front and Side view of the images trained.

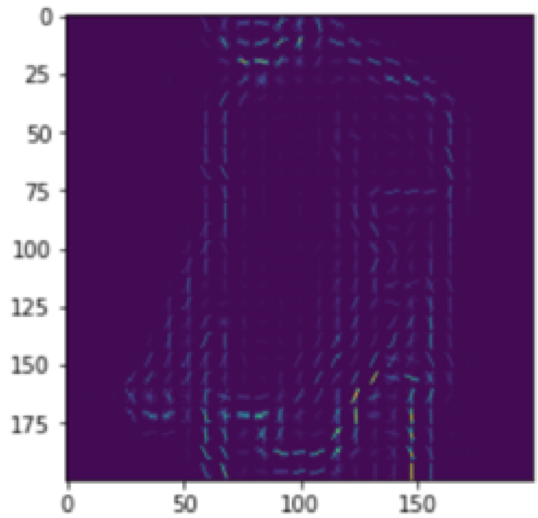

Fig. 2. Histogram of oriented gradient features of Image side view

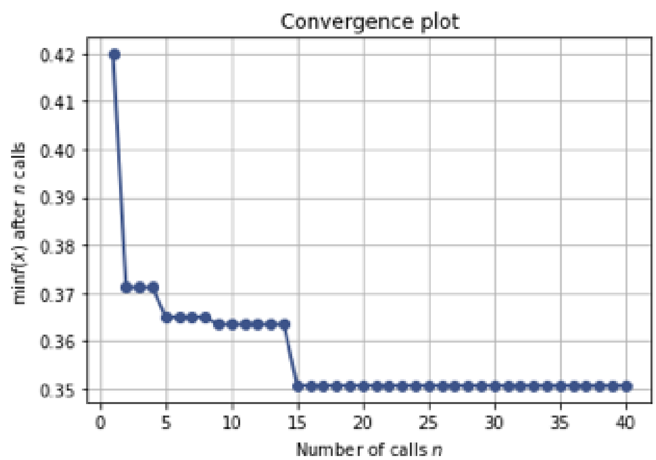

Fig. 3. Convergence plot of model hyper parameters in Bayesian Optimization

[3], the convergence plot of the same (sample) is shown in Figure 3.

The cross-validation accuracy of each of the models thus computed is shown below in Table 1.

TABLE I
CROSS VALIDATION ACCURACY OF VARIOUS CLASSIFICATION MODELS WITH HISTOGRAM OF ORIENTED GRADIENT FEATURES

Classifiers	Cross-Validation Accuracy
SVM	62.22%
Multinomial Logistic	71.23%
Naïve Bayes	62.12%
Decision Tree	55%
Random Forest	70%

As shown in Table 1, the cross validation accuracy is quite low from all of the classification models and hence we modify it using our methodology.

B. Convolution Neural Networks and Transfer Learning features based model:

Recently image classification task using Convolution Neural Networks (CNN) and Transfer Learning has gained huge success [1],[2]. So to solve the image orientation classification problem (front, side and back view) three pre-trained Convolution Neural Network model features have been extracted. The three models are Mobile net [8], VGG16 [9] and Inception [10] from which the last layer features have been extracted which consists of the most important and specific features for the classification task. Each of the pre-trained features has been finally trained on our dataset. The pre-trained features act as the predictors and all the 5 models mentioned previously which consists of SVM, Multinomial Logistic, Naive Bayes, Decision Tree and Random Forest with the response variable having 3 classes' i.e Front view, Side view and Back view.

The cross-validation accuracy for each of the pre-trained features and each model has been shown in Table 2.

TABLE II
CROSS VALIDATION ACCURACY OF PRE-TRAINED CNN FEATURES FOR EACH OF THE CLASSIFICATION MODELS

	Cross Validation Accuracy		
Classifiers	Mobile net	VGG16	Inception
SVM	94.2%	83%	82.9%
Multinomial Logistic	94.69%	89%	90%
Naïve Bayes	85.6%	80.2%	69%
Decision Tree	90%	76%	65%
Random Forest	93%	88%	81%

As shown in Table 2, the Mobilenet features clearly outperform each of the other pre-trained CNN features even for a relatively small dataset and hence the same has been chosen for the process of image orientation classification.

The major advantage of Mobile net split the convolution into a 3x3 depth wise convolution and a 1x1 point wise convolution as described in [8] which makes it accurate and fast. The resource/accuracy optimization has been done in the most efficient manner in Mobile net.

It can be clearly seen from both Table 1 and 2 that amongst all the classifiers used, Multinomial logistic regression clearly outperforms all the other classifiers for both pre-trained CNN features as well as Histogram of oriented gradient based features. Hence, Multinomial Logistic Regression model with Mobile net features has been selected as the final model of classification of image orientation which gives an accuracy of approximately 95%.

C. Comparison of final model with baseline model based on Cross validation accuracy:

A 10 fold cross validation was performed for both the Mobilenet feature based multinomial logistic regression model and Histogram of gradient feature based multinomial logistic model and a Student's t-test [11] was performed to show that the accuracy in the former is significantly better than the later as shown in Table 3.

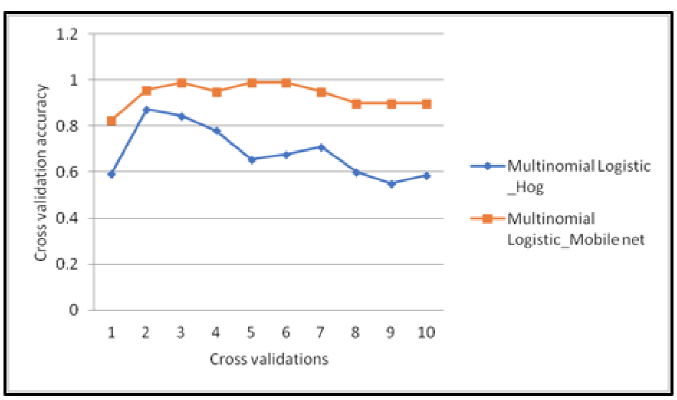

Fig. 4. Cross-validation accuracy for both the models

TABLE III
STUDENT'S T-TEST FOR COMPARISON OF CROSS-VALIDATION ACCURACY OF MODELS

t-Test: Two-Sample Assuming Equal Variances		
	Multinomial logistic _Hog	Multinomial Logistic_Mobile net
Mean	0.68747	0.93525
Variance	0.012784722	0.002838069
Observations	10	10
Pooled Variance	0.007811396	
Hypothesized Mean Difference	0	
df	18	
t Stat	-6.26883626	
P(T<=t) one-tail	3.26456E-06	
t Critical one-tail	1.734063607	
P(T<=t) two-tail	6.52912E-06	
t Critical two-tail	2.10092204	

As it can be seen from both Table 3 and Figure 4, Mobilenet CNN model features with Multinomial Logistic Regression classifier trained on our dataset outperform significantly our baseline model and hence that has been selected for the image orientation classification. This constitutes the first part of our pipeline.

V. IMAGE QUALITY ASSESSMENT USING STRUCTURAL SIMILARITY INDEX AND TRANSFER LEARNING

For the task of quality assessment of images sent by vendor automatically, structural similarity has been used as the desired index as mentioned in [5]. The conventional metrics such as the peak signal-to-noise ratio (PSNR) and the mean squared error (MSE) which operate directly on the intensity of the image don't qualify as human visual system based quality metric. But in our case, it is very important to use a quality index which is very similar to human perception and hence Structural similarity index which takes into account the impact of changes in luminance, contrast and structure in an image has been considered.

A. Introduction of Noise to the images of our dataset (mentioned in 2.0):

The main challenge in the field of image quality assessment is we won't have the perfect image of an item every time with the imperfect/poor quality ones so that we can assess the quality of the images. Hence, we need a methodology where quality of the image can be assessed without reference image and which can work for small datasets as well [6].

The first step is to add distortion to the reference images of the datasets with different noise signals and artificially create our own datasets of good images and distorted images. For our case the noise signals considered are different types of blurring since that is one major area of concern for the images sent by vendor which is shown in Table 4.(Here reference image is only for the training set, for test set there won't be any)

The operation has been done for all the 312 images and each type of distortion has been considered as a separate class/category which makes a total of 13 categories including the reference good images.

TABLE IV
DIFFERENT DISTORTION TYPES ADDED TO REFERENCE IMAGES

Type of Noise added	Kernels and Parameters
Mean Blur	(5,5),(25,25),(55,55),(75,75)
Gaussian Blur	(5,5),(25,25),(55,55),(95,95)
Bilateral Blur	(9,50,50),(9,125,125)
Median Blur	5,27

B. Image quality based classification using Mobile net CNN features and Deep Learning Classification algorithm:

In the second step of the process of image quality assessment, the pre-trained Mobilenet[8] last layer features have been extracted for all the images of 13 different classes mentioned above which includes the good/reference class images, Mean blurred images (4 different classes), Gaussian Blurred images (4 different classes), Bilateral Blurred images(2 different classes) and Median Blur(2 different classes) . The Mobilenet [8] final layer features of the images contain all the important features and information about them.

Then a deep learning classification model has been trained on those features and the model specification is shown below in Table 5.

TABLE V
DEEP LEARNING MODEL DESCRIPTION

Layers	No. of Neurons	Activations Function
Input Layer	1024	-
Hidden Layer1	512	Relu
Hidden Layer 2	256	Relu
Hidden Layer 3	112	Relu
Output Layer	13	Softmax

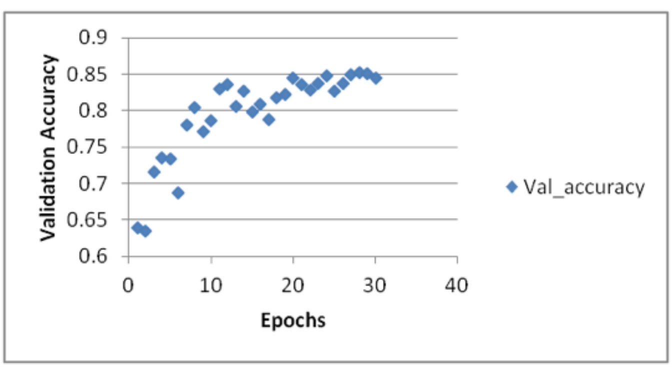

Fig. 5. Validation accuracy of the Deep Classification model

As shown in figure 5, the cross validation accuracy obtained by the model was 84.5% which is quite high considering the amount of data used. The final layer of the deep model is extracted as these features are the quality-related features for these images. The logic behind the statement is that in these image classes (13) the only difference is the image quality and all other things are same for all the classes and hence if a model is differentiating between these images it clearly indicates the features will be those features which are related to quality characteristics of the images.

C. Computation of Structural Similarity Scores and prediction using Ridge Regression model

Once the quality related features have been extracted for the images, the Structural similarity scores for all the distorted images are computed from the reference images as in [5]. Then the quality related features of the images have been taken as predictor variables and the Structural similarity scores for the same images as the response variables and a Ridge regression is fitted with an 80-20 validation and an accuracy of 83% is achieved by this methodology. So, now whenever a new image is there, Mobilenet features will be extracted from the image which will be considered as a test data point for our Ridge regression model and the Structural similarity score for that image will be predicted using the model and based on which business will decide on a cutoff value of the same below which it might not be acceptable and necessary actions can be taken.

Finally the ordering is done as per business requirements which complete the pipeline of our process.

VI. EVALUATION:

As shown in Table 1 and Table 2, our MobileNet CNN features with Multinomial logistic regression performs much better than the baseline model and other pre-trained CNN features.

TABLE VI
CROSS VALIDATION ACCURACY OF VARIOUS MODELS

	Cross Validation Accuracy			
Classifiers	Hog	MobileNet	VGG16	Inception
SVM	0.6222	0.942	0.83	0.829
Multinomial Logistic	0.7123	0.9469	0.89	0.9
Naïve Bayes	0.6212	0.856	0.802	0.69
Decision Tree	0.55	0.9	0.76	0.65
Random Forest	0.7	0.93	0.88	0.81

As it can be seen that MobileNet CNN features with Multinomial logistic regression performance is much superior and that has been tested in Table 3 via t-test.

TABLE VII
VALIDATION ACCRACY OF DEEP LEARNING MODEL FOR IMAGE QUALITY CLASSIFICATION

Epochs	Validation Accuracy
1	0.6394
7	0.78
15	0.7982
20	0.8445
25	0.8263
28	0.8528
30	0.8453

An accuracy of 85% was achieved by the deep learning quality classification model and finally the Ridge regression model had an accuracy of 83%.

VII. CONCLUSION:

In this work, we have successfully built a pipeline where in the first step we have classified the image orientations, in this case Front-view, Side-view & Back-view with a cross validation accuracy of 94% with pre-trained Mobile Net features and

Multinomial Logistic Regression approach and that too with small dataset which was one of the challenge for our work. This process actually reduces the manual labor and helps in easing the process of catalogue management.

The next most important part of our pipeline of automated catalogue management was to successfully implement image quality assessment with no-reference image. This is a very important area since many of the images of items sent by the vendor are not as per required quality which causes the customer to move to different industries. Moreover, this is a reasonably challenging task to assess image quality when the reference image is not present.

In the methodology developed to solve this problem, the first step is to add distortions/noise to our reference images and then extract MobileNet features and finally a deep learning model is trained in such a way that it can uniquely identify the different classes of images. The last layer features from this deep learning model has been extracted since it consists of the quality characteristics of the images.

The structural similarity index has been used as the index to measure the structural similarity between the reference and distorted images as it is almost similar to the way human perceives image quality. Using the structural similarity scores as the response and the features of the deep model as predictor, a Ridge regression model is being fitted with an accuracy of 83% which is quite good considering the complexity of the problem. So, now whenever a new image comes, first the MobileNet features will be extracted from it and its structural similarity score will be predicted from the Ridge regression model.

Finally the ordering is done as per Business requirements and this wraps up the pipeline built for automated catalogue management.

Further scope of research is there to classify more orientations of images for image orientation classification. In image quality classification task, ensemble models can be used to make the accuracy better and many different types of noise signals can also be added to make the model much better.

VIII. References

1) A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.
2) Sri Winiarti, Adhi Prahara, Murinto and Dewi Pramudi Ismi, “Pre-Trained Convolutional Neural Network for Classification of Tanning Leather Image”, in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 1, 2018.
3) Jasper Snoek, Hugo Larochelle and Ryan P. Adams, “Practical Bayesian Optimization of Machine Learning Algorithms”, in Advances in Neural Information Processing Systems 25 (NIPS 2012).
4) Z. Wang, A. C. Bovik, and L. Lu, “Why is image quality assessment so difficult”, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 4, Orlando, FL, May 2002, pp. 3313–3316.
5) Zhou Wang, Alan C. Bovik, Hamid R. Sheikh and Eero P. Simoncelli , “Image Quality Assessment: From Error Visibility to Structural Similarity” in IEEE Transactions on Image Processing, VOL. 13, NO. 4, APRIL 2004.
6) Sebastian Bossey, Dominique Maniryy, Klaus-Robert M¨uller, Thomas Wiegand, and Wojciech Samek, “Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment” in arXiv:1612.01697v2 [cs.CV] 7 Dec 2017.
7) Paul E. Rybski and Daniel Huber and Daniel D. Morris and Regis Hoffman, “Visual Classification of Coarse Vehicle Orientation using Histogram of Oriented Gradients Features” in 2010 IEEE Intelligent Vehicles Symposium.
8) Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto and Hartwig Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications” in arXiv:1704.04861v1 [cs.CV] 17 Apr 2017.
9) Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition” in arXiv:1409.1556v6 [cs.CV] 10 Apr 2015.
10) Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, “Going Deeper with Convolutions” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
11) Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for engineers. 6th edition, Wiley India.