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

**Souradip Chakraborty**[*], **Mani Kanteswara Rao Garlapati**[*]

[*] *Data & Analytics, Walmart Labs, Bangalore*

***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 an online store, customers are unable to touch the product before buying it and this can only be compensated by providing a good sensory experience through image catalogue and efficient management of the same. 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 is to automatically classify the various orientations (front view, side view, top view etc.) of the images sent by the vendor using CNN and 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 No reference image quality assessment and finally the automatic ordering of items are done as per thresholding. Good quality images from all orientations plays a critical role in making a customer-friendly online store leading to customer satisfaction.

***Index Terms-*** Convolution Neural Networks, Transfer Learning, Image quality assessment, Structural similarity index, Quality Embedding.

## I. 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 hyperparameter tuning has been done using Bayesian optimization [3] which gave much superior results when compared to the baseline model. Since manual/decision rule based ordering of the images sent by vendors are being done in majority of industries currently which is extremely time-consuming and hence it can be improved vastly by the our methodology . Secondly, quality of the images sent by the vendor plays a crucial part as improper image quality in an online platform might directly lead to customer dissatisfaction. The way human perceives image quality is very unique and to make the machine understand and learn that way makes image quality assessment a very difficult task to perform[4]. Hence, Structural similarity index [5] has been considered as a metric in this case for assessment of the quality of images of items sent by vendor which gives the human-perceived notion of quality . The major challenge faced with respect to quality is the blurring effect in images sent by vendors which also is one of the primary cause of customer dissatisfaction as understood from various customer feedback and surveys .Hence the second phase of the pipeline deals with assessing the quality of the image automatically once it falls below a certain quality threshold. The main contribution of the paper lies in the development of quality embeddings which projects each and every image in some latent dimensions which represents various quality attributes and using the same ,the human perceived quality metric has been predicted for every image. The concept of quality embeddings have not been used before and it helps in no-reference image quality assessment task efficiently. Another contribution of the paper lies in the synthetically generated noisy datasets which eliminated the manual annotation process very effectively and helps in the no-referencing quality assessment. The key idea lies in the concept that human beings while detecting if an image is of poor quality or not doesn't need the true reference superior quality version of the image. If an image is a lit blurred, human beings are well adept in detecting the same and hence for machines to reflect the same intuition , the above methodology has been implemented. Bayesian optimization has been leveraged in the process of hyperparameter tuning which reduces the time complexity of the pipeline significantly and provides an intelligent approach to search the best hyperparameters in the given space.

## II. IDENTIFY,RESEARCH AND COLLECT IDEA

There is lot of research work that has happened over the years in the field of image classification and orientation detection, but in majority of the models developed there is a requirement that the dataset size should be large enough since deep convolution based models will have a lot of parameters. The complexity lies in this case since there will be multiple new items for which dataset size won't be large and the model thus developed should be robust enough to work in such scenarios as well. It also includes other constraints such as time complexity, simplicity and light weight models for the pipeline to work optimally. The creation of robust features using light weight MobileNet CNN helps in achieving the objective of orientation classification.

There has been work done in the field of image quality assessment but in many of the cases manually annotation of datasets have been used. In our case , we have synthetically generated noisy datasets which reduces the manual efforts of annotating. The quality embeddings developed in our architecture has never been used/developed till date for image quality assessment.

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.

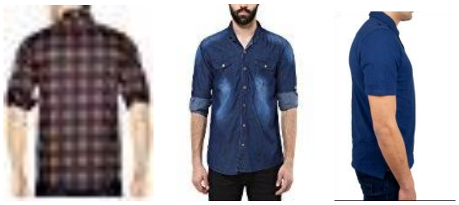


Figure 1: Back, Front and Side view of the images trained

## III. STUDIES AND FINDINGS

## Image Orientation Classification using Convolution Neural Network and Transfer Learning

### Histogram of Oriented Gradients(HOG) as Baseline Model:

The HOG features are mainly used in image processing object detection tasks The key idea behind the histogram of oriented gradients descriptor is that local object appearance within an image can be described by the distribution of edge directions. The image is divided into small connected regions and for the pixels within the regions, a histogram of gradient directions are computed. The final feature vector is the combination of all these histogram features.

For implementation of the task of classification of image orientation 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 widely used in many places where image orientation classification is the prime objective [7].

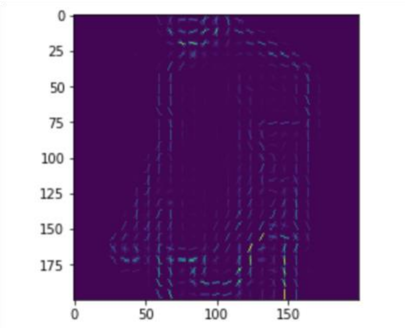


Figure 2: Histogram of oriented gradient features of Image side view

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

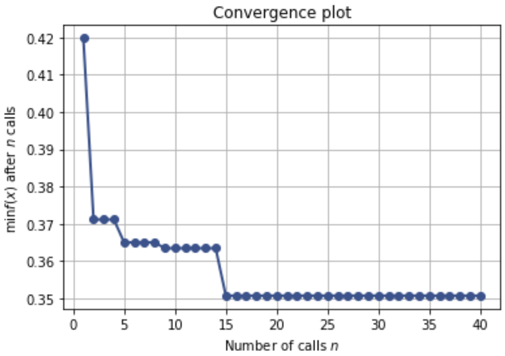


Figure 3: Convergence plot of model hyper parameters in Bayesian Optimization

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

Table1: 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% |

Bayesian Optimization helps in reducing the time complexity associated with grid search for the hyperparameters significantly as it implements an intelligent way of searching the space using Gaussian process. So, at each iteration it implements a trade-off between exploration and exploitation and thus forms an utility function and optimizing the same it chooses the next best hyperparameter.

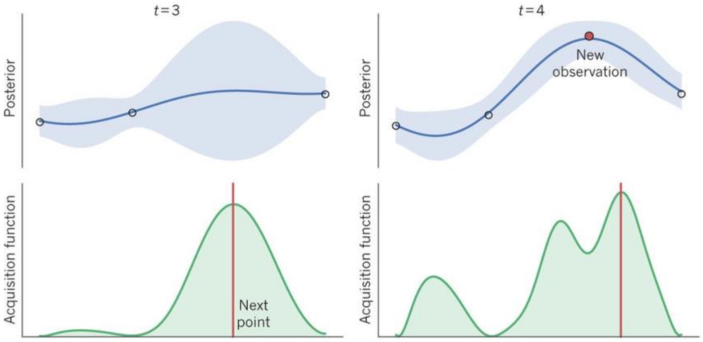


Figure 4: Bayesian Optimization and Gaussian Process

But as can be seen in Table1, HOG features fail to give a good accuracy in orientation classification problem and thus we use our CNN and Transfer Learning based approach to implement the same.

### Convolution Neural Networks and Transfer Learning features based model:

Recently image classification task using Convolution Neural Networks (pre-trained on ImageNet dataset) 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 2: 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 MobileNet is that it uses depth wise separable convolutions to build light weight deep neural networks[8]. Another advantage being it has only two global hyperparameters which can be tuned very easily for the trade-off between latency and accuracy.

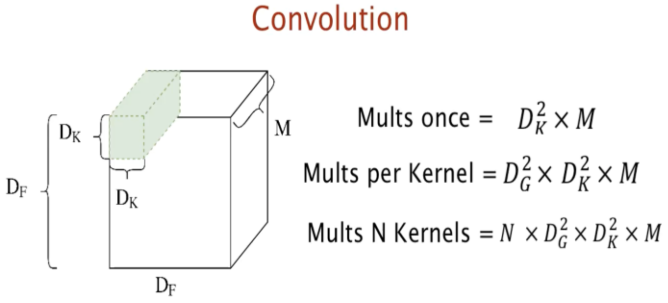

Convolution
$D_K$
$D_K$
$M$
$D_F$
$D_F$
Mults once = $D_K^2 \times M$
Mults per Kernel = $D_G^2 \times D_K^2 \times M$
Mults N Kernels = $N \times D_G^2 \times D_K^2 \times M$

Figure 5: Number of Computation in Vanilla Convolution

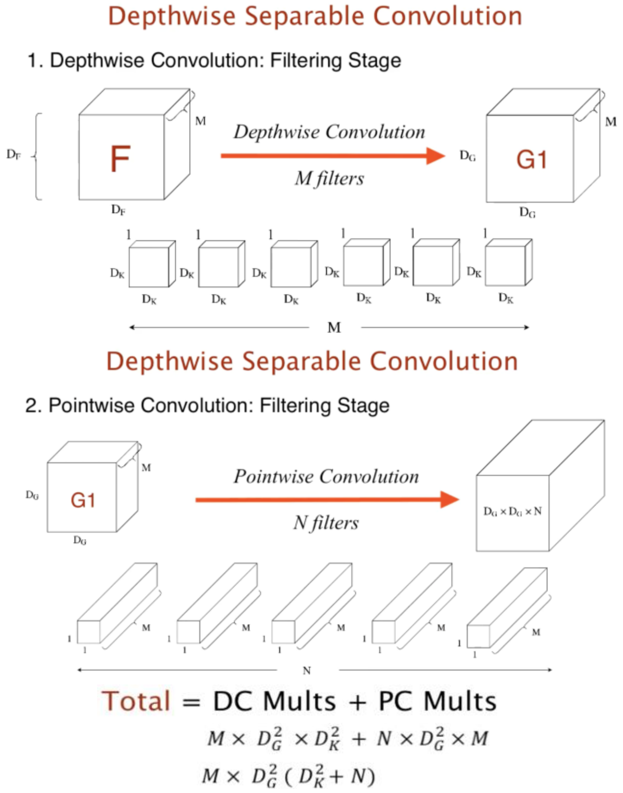

Depthwise Separable Convolution
1. Depthwise Convolution: Filtering Stage
$D_F$
F
M
$D_F$
Depthwise Convolution
M filters
$D_G$
G1
M
$D_G$
1
$D_K$
$D_K$
M
Depthwise Separable Convolution
2. Pointwise Convolution: Filtering Stage
$D_G$
G1
M
$D_G$
Pointwise Convolution
N filters
$D_G \times D_G \times N$
1
1
M
N
Total = DC Mults + PC Mults
$M \times D_G^2 \times D_K^2 + N \times D_G^2 \times M$
$M \times D_G^2 ( D_K^2 + N)$

Figure 6: Number of Computations in Depth wise Convolution

As discussed, the number of computations has been significantly reduced in case of depth wise convolution and thereby decreasing the time complexity and making the model light.

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%.

**Comparison of final model with baseline model based on Cross validation accuracy: Statistical Significance**

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. The p-value<0.05 which indicates the statistical significance.

Table 3: Student's t-test for comparison of Cross-validation accuracy of models

<table>
<tr><td colspan="3">t-Test: Two-Sample Assuming Equal Variances</td></tr>
<tr><td></td><td><i>Multinomial logistic _Hog</i></td><td><i>Multinomial Logistic Mobilenet CNN</i></td></tr>
<tr><td>Mean</td><td>0.68747</td><td>0.93525</td></tr>
<tr><td>Variance</td><td>0.012784722</td><td>0.002838069</td></tr>
<tr><td>Observations</td><td>10</td><td>10</td></tr>
<tr><td>Pooled Variance</td><td>0.007811396</td><td></td></tr>
<tr><td>Hypothesized Mean Difference</td><td>0</td><td></td></tr>
<tr><td>df</td><td>18</td><td></td></tr>
<tr><td>t Stat</td><td>-6.26883626</td><td></td></tr>
<tr><td>P(T<=t) one-tail</td><td>3.26456E-06</td><td></td></tr>
<tr><td>t Critical one-tail</td><td>1.734063607</td><td></td></tr>
<tr><td>P(T<=t) two-tail</td><td>6.52912E-06</td><td></td></tr>
<tr><td>t Critical two-tail</td><td>2.10092204</td><td></td></tr>
</table>

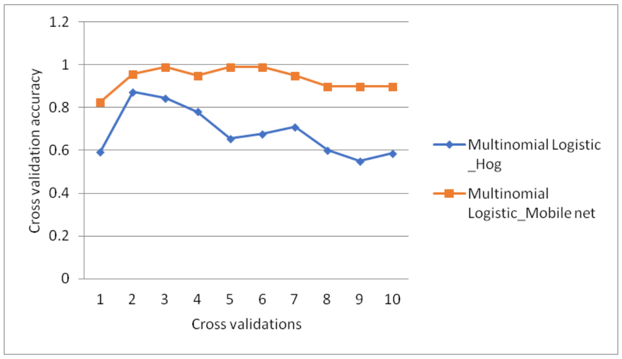


Figure 7: Cross-validation accuracy for both the models

As it can be seen from both Table 3 and Figure 7, 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.

## 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 considers the impact of changes in luminance, contrast and structure in an image has been considered as shown in Figure 8.

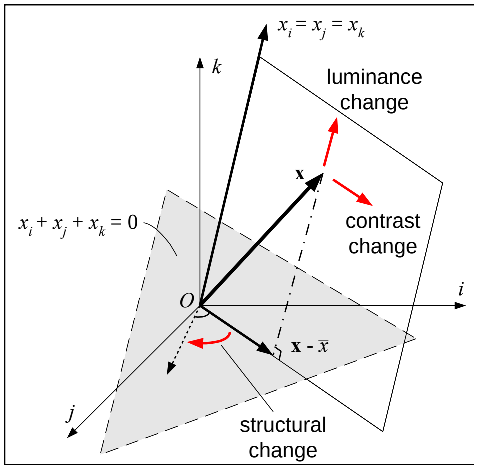


Figure 8: Structural Similarity Index

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \qquad (1)$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \qquad (2)$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \qquad (3)$$

$$SSIM(\boldsymbol{x},\boldsymbol{y}) = l(\boldsymbol{x},\boldsymbol{y}) \cdot c(\boldsymbol{x},\boldsymbol{y}) \cdot s(\boldsymbol{x},\boldsymbol{y}) \qquad (4)$$

As shown in Figure8 and Equation 4, the structural similarity metric incorporates the illuminance, contrast and structural components of an image and hence is likely to capture the human perception whereas other metrics like MSE and PSNR etc. only captures the pixel wise difference between the two images which is not the way human perceives quality.

### Introduction of Noise to the images of our dataset: Synthetic Data Generation

The way human perceives quality is very unique and every time some image of poor quality comes, it is a very easy task for human to detect that the quality is not adequate may be some blurring, other noise factors are there in the image. Humans won’t need any reference image of superior quality for that poor quality image to tell that which motivates us to the concept of no-reference image quality assessment.

The main challenge in the field of image quality assessment is that we won’t have the perfect image of an item every time with its corresponding imperfect/poor quality version for assessing the quality of the images. Hence, we need a methodology where quality of the image can be assessed without reference image [6]and which can work for small datasets as well. The idea is to make the machine learn the way human perceives quality in such cases.

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. There can be various types of noise signals which can be given to the image but for our case we have considered blurring as the noise factor with various factors and kernels of the same. 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).

Table 4: 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 |

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. Since each parameter induces blurring of different types and each type has been considered as a separate class for the supervised framework that we have created.

**Image quality based classification using Mobile net CNN features and Deep Learning Classification algorithm**: **Quality Embeddings**

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. As discussed earlier as well the benefits of having a light weight model with depth wise convolution, MobileNet CNN captures the most relevant features from the image . Then we have built deep layers on top of it which basically projects the features into different dimensions. Finally using a SoftMax layer, we have classified them into the 13 different classes and the model is trained on the same. The deep learning architecture after extraction of the MobileNet embeddings have been shown below.

Table 5: Deep Learning Model Description

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

The key idea and innovation of the work lies in the concept of creating quality embeddings for each of the images. The last but one layer before the SoftMax layer of the model described above projects the images into quality dimensions. The main intuition behind that is if the model is generating such features in the final layer such that it is being able to classify images which are similar otherwise and the only difference lies in quality of the images, then the features that are generated are quality based features.

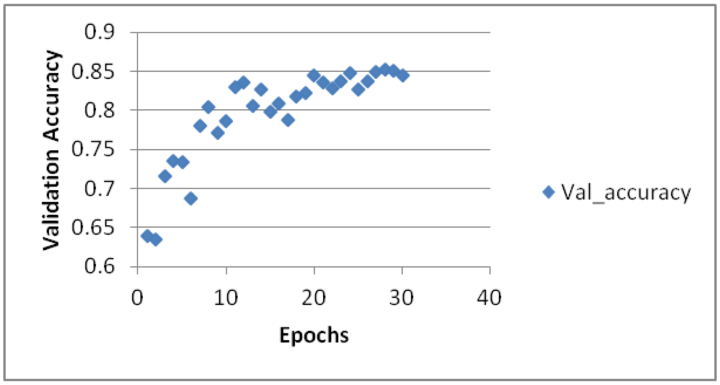


Figure 9: Validation accuracy of the Deep Classification model

As shown in figure 9, 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 embeddings or quality-related features for these images. The main idea as mentioned above as well 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. The diagrammatic workflow has been shown below in Figure 10.

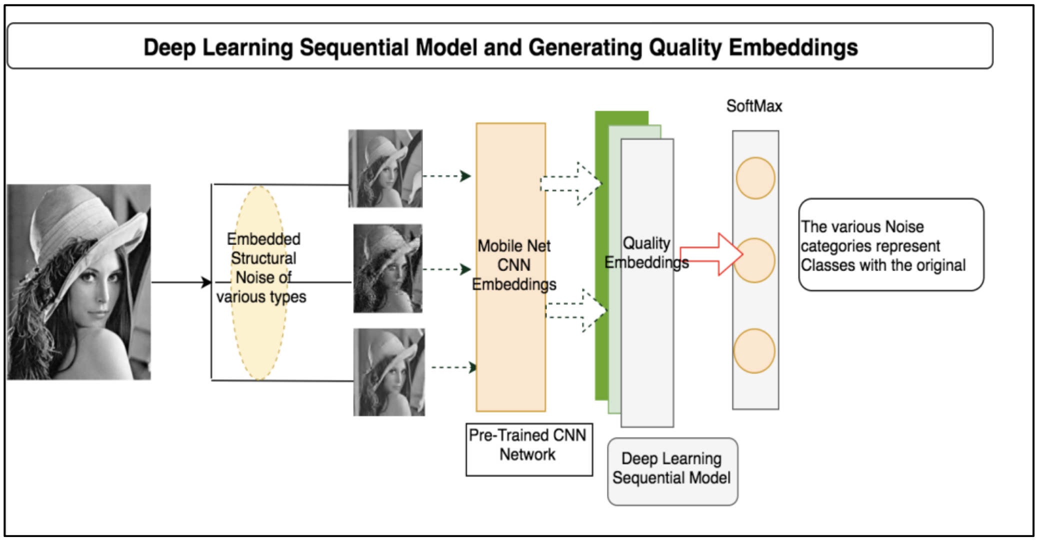

Deep Learning Sequential Model and Generating Quality Embeddings
SoftMax
Embedded Structural Noise of various types
Mobile Net CNN Embeddings
Quality Embeddings
The various Noise categories represent Classes with the original
Pre-Trained CNN Network
Deep Learning Sequential Model

Figure 10: Generating Quality Embeddings using Deep architecture

### Computation of Structural Similarity Scores and prediction using Ridge Regression model

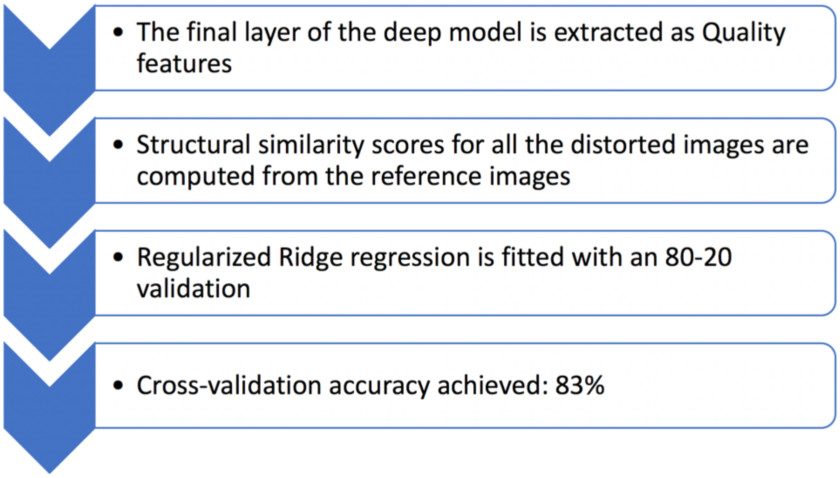
Once the quality related features have been extracted for the images, the Structural similarity scores for all the known synthetically generated distorted images and original images are computed from the reference images. So, for the true reference images the structural similarity will be 1 and as the distortion in the images increasing the structural similarity metric value decreases.

Then the quality embeddings for all the images have been extracted using the deep learning model described in Table5 and extracting the last but one layer weights.

Once the images are projected into quality dimensions, the quality embeddings have been taken as predictor variables and the Structural similarity scores computed for the same images as the response variables and a Ridge regression is fitted with an 80-20 validation and a validation accuracy of 83% is achieved by this methodology. So, now whenever a new image is there, the quality embeddings are extracted from the images by projecting them in the latent quality dimensions and then considering the same as a test feature for our Ridge regression model, the Structural similarity score for that image will be predicted using the model and based on which and a business decided threshold value, necessary actions will be taken. The ridge regression model equation is shown in Equation 5.

$$\begin{aligned}\hat{\beta}^{\text{ridge}} &= \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 \\ &= \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}} \underbrace{\|y - X\beta\|_2^2}_{\text{Loss}} + \lambda \underbrace{\|\beta\|_2^2}_{\text{Penalty}}\end{aligned} \tag{5}$$

Finally the ordering is done as per business requirements which complete the pipeline of our process and the flow has been shown below in Figure 11.



- The final layer of the deep model is extracted as Quality features
- Structural similarity scores for all the distorted images are computed from the reference images
- Regularized Ridge regression is fitted with an 80-20 validation
- Cross-validation accuracy achieved: 83%

Figure 11: Directional Flow for predicting Structure Similarity metric using Ridge regression

## 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 compiled in Table 6.

Table 6: 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 |

The statistical significance test has been performed to check if the increase in accuracy is statistical significant or not and hence a paired t-test has been done to do the same.

Table 7: Student's paired t-test for comparison of Cross-validation accuracy of models

| t-Test: Two-Sample Assuming Equal Variances | | |
|---|---|---|
| | *Multinomial logistic _Hog* | *Multinomial Logistic Mobilenet CNN* |
| 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 that MobileNet CNN features with Multinomial logistic regression performance is much superior and that has been tested in Table 7 via paired t-test.

Table 8: Validation accuracy 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%.

The final workflow of the pipeline has been shown below in Figure 12.

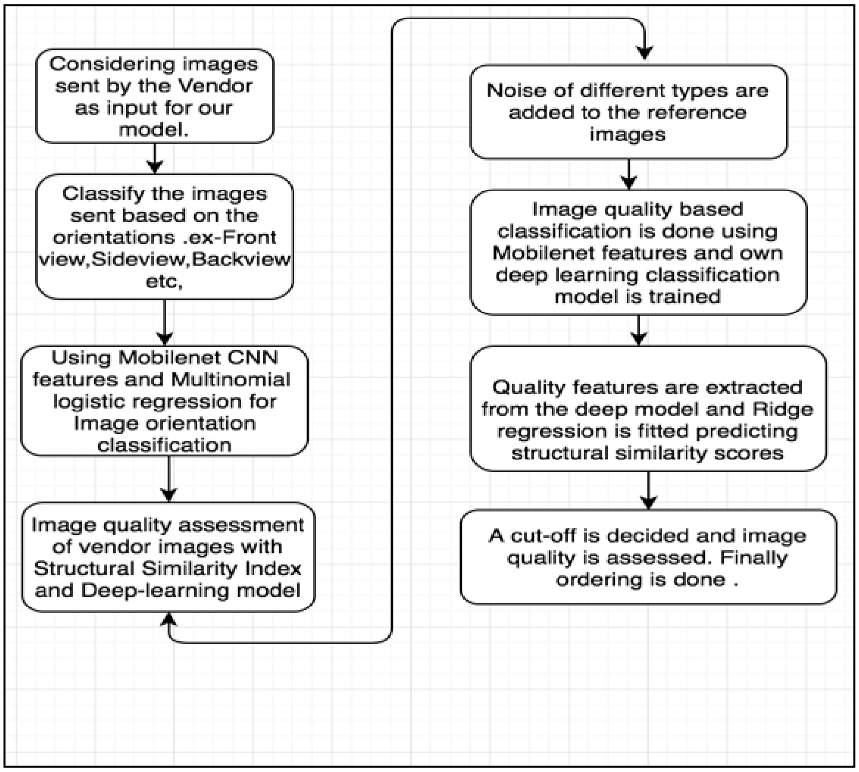

Considering images sent by the Vendor as input for our model.
Classify the images sent based on the orientations .ex-Front view,Sideview,Backview etc,
Using Mobilenet CNN features and Multinomial logistic regression for Image orientation classification
Image quality assessment of vendor images with Structural Similarity Index and Deep-learning model
Noise of different types are added to the reference images
Image quality based classification is done using Mobilenet features and own deep learning classification model is trained
Quality features are extracted from the deep model and Ridge regression is fitted predicting structural similarity scores
A cut-off is decided and image quality is assessed. Finally ordering is done .

Figure 12: Final Workflow of the Pipeline of Catalogue Management.

## IV. CONCLUSION

In this work, we have successfully build 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.

## ACKNOWLEDGMENT

We want to grasp this opportunity to convey our deepest gratitude to all of them who aided us in completing our paper. Our heartfelt gratitude to AICAAM team, for this opportunity. Our great thankfulness to Data and Analytics Team, Walmart International, for introducing us to this topic and guiding us throughout our paper. A special thanks to Ashish Gupta, Manager Walmart Labs as our internal guide of the paper, for his persistent encouragement and motivation towards us. We are thankful to all of them who directly and indirectly helped us in accomplishing this paper.

## REFERENCES

[1]. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 1. Curran Associates Inc., USA, 1097-1105.

[2]. Murinto, Murinto & Prahara, Adhi & Winiari, Sri & Pramudi Ismi, Dewi. (2018). Pre-Trained Convolutional Neural Network for Classification of Tanning Leather Image. International Journal of Advanced Computer Science and Applications. 9. 10.14569/IJACSA.2018.090129.

[3]. Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. 2012. Practical Bayesian optimization of machine learning algorithms. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'12), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 2. Curran Associates Inc., USA, 2951-2959.

[4]. Wang, Zhou & Bovik, Alan & Lu, Ligang. (2002). Why is image quality assessment so difficult? Proc IEEE Int Conf Acoustics Speech Signal Process. 4. IV-3313. 10.1109/ICASSP.2002.1004620.

[5]. Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

[6]. S. Bosse, D. Maniry, K. Müller, T. Wiegand and W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," in IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 206-219, Jan. 2018.

[7]. P. E. Rybski, D. Huber, D. D. Morris and R. Hoffman, "Visual classification of coarse vehicle orientation using Histogram of Oriented Gradients features," *2010 IEEE Intelligent Vehicles Symposium*, San Diego, CA, 2010, pp. 921-928.

[8]. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *CoRR, abs/1704.04861*.

[9]. Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[10]. C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9.

[11]. Douglas C. Montgomery, George C. Runger, *Applied Statistics and Probability for engineers*. 6th edition, Wiley India.