Machine Learning Project Update

Ronaldson Bellande
University of Massachusetts Lowell
Ronaldson_Bellande@student.uml.edu

Parisa Hajibabaee
University of Massachusetts Lowell
parisa_hajibabaee@student.uml.edu

Gregory LeMasurier
University of Massachusetts Lowell
gregory_lemasurier@student.uml.edu

I. Introduction

Image classification is used in industries ranging from healthcare, robotics, advertising and the automobile industry. A misclassification could have dire consequences, thus it is important to understand and compare the performance of various machine learning implementations to ensure that a reliable algorithm is used. In many machine learning algorithms the model has a possibility of not actually learning from the provided picture and instead is learning from the environment and that could cause major issues. In this project, we are going to explore and compare multiple methods to the problem of image classification. We will compare their performance to baselines provided by related works and the top performing algorithms to see which algorithms gives us the best accuracy while actually learning from the object at hand and not the environment.

In order to do a task of image classification, we need to determine a suitable classification system and then select training samples, do the image processing, feature extraction, selection of suitable classification approaches, post-classification processing, accurate learning objective and accuracy assessment.

Given an image and K categories, the task is to select one category label for each image. We have developed proof of concept algorithms for image classification using transfer learning, k-means clustering with logistic regression, and neural networks.

In this paper, we report our progress on three image classification algorithms. These algorithms include models from transfer learning, clustering and classification, and convolutional neural networks. Many related works have evaluated similar algorithms on the Pascal VOC datasets. We will compare the initial results of our algorithms to the reported results in these papers in the future. We will discuss potential causes for the differences in precision between algorithms in our final report. This report focuses mostly on our initial designs and our future plans for the project.

II. Problem Definition and Algorithm

A. Task Definition

For our project we have focused on the image classification problem. Image classification is used in industries including the healthcare industry, robotics, advertising and the automobile industry. In these scenarios, a misclassification could have dire consequences, thus its important that image classification algorithms are precise, have a low error rate and are learning from the object and not the environment.

We have chosen to use the Pascal VOC 2012 dataset [1]. This data set has about 17,000 images that belong to 20 different categories. These categories include: aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, pottedplant, sheep, sofa, train and tvmonitor. The dimentions of the original images are various so, we need to downsample them to a fixed resolution for the k-means clustering. We chosen to upsample the images for the remaining algorithms. This dataset provides us with image sets for training and validation runs. We will use these and we will also run our algorithms with a custom split of 80% training, 10% validation, and 10% test.

The inputs to our algorithms are images from the Pascal VOC 2012 dataset. Our algorithms will decide the most appropriate class for each image in our validation and test image sets. We will analyze our algorithms' performance through classification precision and error rate metrics.

B. Algorithm Definition

In this section we describe our preprocessing on the Pascal VOC 2012 dataset as well as the current state of all three of our algorithms, which use transfer learning, clustering with logistic regression, and neural networks.

1) Preprocessing: First we have to preprocess our data before we run our algorithms on it. All of our algorithms require our images to be in a directory that is associated with their class name, thus we have written a script to sort our images into their appropriate classes. Our clustering and transfer learning algorithms require images to be in png format. The Pascal VOC 2012 dataset contains jpg images, so we have written a script to convert all images to png format. For our clustering algorithm, we then had to downsample our images to improve memory consumption when we create our test and validation sets and to make all of the images the same resolution. We have resized all of our images to be 128x128 pixels using OpenCV's inter cubic resize function. While this does reduce memory consumption, it also could have an impact on our accuracy as some images may become distorted. We plan to report the impact of this downsampling method for the final project report. To ensure that this preprocessing can be replicated, all of the scripts will be included in our final project submission along with README instructions that describe how to run each script.

2) Clustering and classification: Clustering can be efficient approach to dimensionality reduction, in particular as a preprocessing step before a supervised learning algorithm like logistic regression classifier. Our clustering and classification algorithm is still in an early state. We have been based our implementation from image classification examples online that use clustering (MNIST dataset). Therefore, we plan to improve our clustering algorithm based the findings of related works in the future.

Currently our classification algorithm uses a logistic regression over the test and validation splits in the Pascal VOC 2012 dataset. Our logistic regression uses the following parameters: $C = 1.0, class\_weight = None, dual = False, fit\_intercept = True, intercept\_scaling = 1, l1\_ratio = None, max\_iter = 5000, multi_class =' ovr', n\_jobs = None, penalty =' l2', random\_state = 42, solver =' lbfgs', tol = 0.0001, verbose = 0, warm\_start = False$. These parameters were selected as suggested by online examples. For our final report, we plan to investigate and change these parameters to provide a more optimal solution. We also plan to evaluate using K-Means preprocessing followed by using logistic regression to try to reduce our error rate. Currently we plan to run our K-Means clustering with the following parameters: $algorithm =' auto', copy\_x = True, init =' k - means + +', max\_iter = 300, n\_clusters = 50, n\_init = 10, n\_jobs = None, precompute_distances =' auto', random\_state = 42, tol = 0.0001, verbose = 0)$ and we will adjust these parameters to try and improve our results.We also plan to use GridSearchCV to find the optimal number of clusters and hopefully get a significant accuracy boost. As an example of using clustering for dimensionality reduction, we tackled the digit MNIST dataset containing 1797 grayscale 8 x 8 images repreenting the digits 0 to 9. First we applied Logistic Regression model and got a baseline accuracy on the test set of 0.968. Then, we created a pipline that first cluster the training set into 50 clusters and we are using centroid to replace the images with their distances to these 50 clusters, then apply a LR model. So we got accuracy of 0.977. By doing so, we reduced the error rate from almost 3.1% to about 2.2%. Finally, we used GridSearchCV to find the optimal number of clusters and then got a significant accuracy boost of 0.982!

3) Convolutional Neural Network: The convolutional neural network (CNN), a class of deep learning neural networks, are widely used in image classification. A CNN, with an input layer, and output layer, and hidden layers with certain amount of neurons, tries to convolve learned features with input data and uses 2D convolutional layers. Very little preprocessing is needed in CNN in comparison with other image classification algorithms. We performed convolutional neural network for our task of image classification. The result will be presented in Section III-B.

4) Transfer Learning: Our transfer learning algorithm was created as a proof of concept. We plan to make adjustments to this algorithm and run this algorithm on the Pascal VOC 2012 dataset in the future. Our approach of transfer learning uses Tensorflow and the Inception V3 convolutional neural network architecture. It will be using trained model on the ImageNet dataset with classes and will be expended upon using a other datasets with different classes as well.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 67, 67, 32)        896
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 23, 23, 32)        9248
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 11, 11, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 4, 4, 64)          18496
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 2, 2, 64)          36928
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 1, 1, 64)          0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 1, 1, 128)         73856
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 1, 1, 128)         147584
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 1, 1, 128)         0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 1, 1, 256)         295168
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 1, 1, 256)         590080
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 1, 1, 256)         0
_________________________________________________________________
flatten (Flatten)            (None, 256)               0
_________________________________________________________________
dense (Dense)                (None, 256)               65792
_________________________________________________________________
dropout (Dropout)            (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               65792
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 257
_________________________________________________________________
activation (Activation)      (None, 1)                 0
=================================================================
Total params: 1,304,097
Trainable params: 1,304,097
Non-trainable params: 0
_________________________________________________________________
```

Fig. 1. This figures shows our model summary for our convolutional neural network

III. Experimental Evaluation

A. Methodology

In our analysis, the dependent variables include the algorithm's precision as well as the error rate. These results are dependent on the independent variables, which include our method of classification, which consist of transfer learning, clustering and logistic regression, and neural networks, as well as the parameters that we use in these algorithms. We also expect that our preprocessing method, for image downsampling for clustering and upsampling for the rest of the algorithms, may also impact the dependent variables. We will analyze these effects more in our final report.

The Pascal VOC 2012 dataset has predefined training and validation image sets. We are using these datasets so that we are consistent with related works. By using the predefined training and validation image sets, we eliminate any effects that could be caused by having variation in the training and validation sets, ultimately allowing us to compare the precision and error rate of our algorithms to the reported results in related work. We will also run a custom split of data which will be divided 80% training, 10% validation, and 10% test. We will investigate if and how this impacts our result. This dataset consists of images where the target object is off center and even partially occluded in some cases. This means that our training and validation sets will include some more realistic, harder cases.

With the Pascal VOC datasets, it is common practice to report the algorithms performance on each category as well. To do this, we are provided with subsets of training and validation images and categories which the related works, for the most part, report their performance on.

To analyze the performance of our algorithms we will measure our classification precision and error rate. These are the metrics that are often reported when evaluating classification on the Pascal VOC datasets. We will be reporting the classification precision and error rate on the general training and validation set, as well as the training and validation set for each classification.

B. Results

We currently do not have results for our transfer learning and clustering with logistic regression algorithms. Both have been proven to work on smaller sets of data. We are using these smaller sets of data to ensure that our algorithms are functioning properly before we start running full attempts, which take much longer. Both algorithms are currently ready to run attempts with the Pascal VOC dataset, however, we ran out of time to complete and analyze these results for this progress report. We will include our findings in our final report.

For our neural network algorithm, we have run an initial attempt on the Pascal VOC 2012 dataset. We used our custom 80-10-10 split and found that our average precision was about 4%. We believe this is due to the fact that we ran 20 epochs, we plan to run our algorithm with 200 epochs to see if that will improve our result. We will also evaluate our algorithm on the predetermined splits in the dataset to ensure that our custom split did not impact our results.

C. Discussion

We can not yet determine the optimal algorithms and parameters as we have not run full attempts on the Pascal VOC 2012 dataset.

For our final project report, we plan to analyze our transfer learning, clustering with logistic regression, and neural network algorithms. We will try to optimize these algorithms by tuning parameters and through methods used by some of the best performing algorithms for image classification. We will also identify and discuss any impact our preprocessing has had on our results. Finally we plan to compare our results to the best performing algorithms on this dataset, as well as other related works.

IV. Related Work

Many cutting edge algorithms are designed using image classification methods. Image classification has been utilized to recognize objects on the side of the road for assisted driving systems [2]. The healthcare industry has also benefited from image classification, including glioma cancer prediction [3] and early oral cancer diagnosis [4]. Image classification has also been used to provide navigation assistance to people who are blind [5].

Lu and Weng conducted a survey analyzing different classification methods for machine learning based on image classification algorithms [6]. There are many different approaches to image classification, including machine learning, transfer learning, reinforcement learning, deep learning, and neural networks.

A. Clustering and classification

Numerous clustering based algorithms have been developed for image classification on the Pascal VOC datasets. One group compared using a SVM (Support Vector Machine) to LDA (Latent Dirichlet allocation) on Pascal VOC 2007 using whitened-histogram orientation clustering [7]. They reported a table of their average precision in all categories. Though they didn't report any breakthrough performance, they proved that they could use LDA instead of SVM to reduce training time with minimal impact on performance.

Other work has presented new clustering methods for unsupervised image classification models. Caron et al. proposed DeepCluster, which is their new clustering method [8]. DeepCluster uses convolutional neural networks and builds off of k-means clustering by predicting cluster assignments and updating their weights [8]. They had analyzed their performance on the Pascal VOC 2007 dataset.

Clustering has also been used to remove outliers in image classification. Ge et al. proposed an image classification algorithm which used density based clustering to remove outliers, which were less similar than the object [9].

B. Convolutional Neural Networks

One group proposed five models, which utilized convolutional neural networks [10]. This solution reported a validation accuracy of 85.6% and a testing accuracy of 85.24% on their best model, which used a pre-trained AlexNet with fine tuned FC layers [10]. This group only reported results for general image classification on the Pascal VOC 2012 dataset, they did not evaluate their algorithm on the predefined training and validation splits.

A neural network algorithm, proposed by Bell et al., has improved the state of the art for the Pascal VOC 2012 [11]. They proposed Inside-Outside Net which uses a 2x stacked 4-directional Identity Recurrent Neural Network and multi-layer region of interest pooling [11].

C. Transfer Learning

Several groups have investigated using transfer learning on the Pascal VOC datasets. Huh et al. has proposed transfer learning from ImageNet onto the Pascal VOC 2007 and 2012 datasets [12]. They found that they could reduce the number of classes and images used in pre-training and still have a

Average Precision (AP %)

	mean	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor	submission date
CSAC-Net V1 [?]	**89.4**	96.4	**92.0**	94.0	92.2	**77.1**	90.5	87.3	95.2	**84.0**	**91.3**	78.1	94.1	94.6	93.4	95.2	**74.9**	91.9	**84.2**	94.3	86.4	09-Feb-2020
SRN+ [?]	88.8	98.2	89.6	92.9	**92.3**	69.3	**93.0**	**89.8**	95.9	80.2	87.8	81.2	94.1	**95.2**	**94.0**	**97.0**	71.8	90.4	80.3	**96.5**	**87.6**	02-Jul-2018
SFA_NET [?]	87.5	95.2	89.7	92.1	90.1	75.0	88.9	84.7	93.8	83.4	90.9	78.3	93.0	93.4	90.3	92.7	72.1	89.8	83.0	92.1	82.3	22-May-2018
SE [?]	86.5	98.3	86.4	92.7	92.0	67.1	90.9	84.6	95.6	75.9	84.5	**82.1**	94.3	93.1	92.6	96.1	62.5	88.0	71.9	96.3	85.1	19-Oct-2016
LIG_DCNN_FEAT_ALL [?]	85.4	**98.6**	86.0	93.4	92.2	65.4	91.0	83.6	95.5	73.4	82.1	79.6	**94.7**	92.9	92.1	95.0	59.4	87.4	67.8	96.0	82.7	08-Sep-2015
S&P_OverFeast_Fast_Bayes [?]	82.8	97.1	82.3	91.2	89.4	61.2	87.8	80.4	94.0	70.7	77.9	75.7	92.5	89.1	89.6	95.0	56.0	83.2	67.4	93.9	82.1	20-Nov-2014
NUSPSL_CTX_GPM_SCM [?]	82.2	97.3	84.2	80.8	85.3	60.8	89.9	86.8	89.3	75.4	77.8	75.1	83.0	87.5	90.1	95.0	57.8	79.2	73.4	94.5	80.7	30-Oct-2014
BCE_loss [?]	82.1	97.1	81.7	**94.8**	85.8	67.8	86.7	83.9	95.4	64.2	82.1	64.8	92.7	88.3	86.7	90.0	55.5	88.2	62.3	92.6	82.3	20-Jan-2018
Resnet [?]	80.7	98.4	81.1	92.9	88.7	57.1	87.5	73.3	**96.7**	63.4	90.1	64.0	94.4	95.1	93.0	76.8	43.8	**93.0**	67.3	93.1	65.2	25-Apr-2017

Fig. 2. This figures shows the best attempts, regarding average precision on the Pascal VOC 2012. The best result in each category is in bold text. The leftmost column is the name of the image classification used, with a hyperlink to their run. The [?] symbol is for the footnote on their website.

very similar result, whereas many people assumed that it was necessary to pre-train on the entire ImageNet dataset [12].

Other groups have also seen success with using transfer learning from ImageNet to the Pascal VOC datasets. Oquab et al. have proposed using transfer learning from ImageNet to reuse mid-level image features in image classification on the Pascal VOC datasets [13]. At the time, this method of transfer learning had outperformed the state of the art on the Pascal VOC 2007 and 2012 datasets [13]

D. Pascal VOC 2012 Leader Board

Out of these related works the best performance was:

The Pascal VOC challenge also has a leader board for the best performing algorithms, in each category, on their dataset. The leader board can be seen in Figure 2. We were unable to find papers associated with these runs in our initial search, we will conduct a more in-depth search and report our findings in our final report.

V. Future Work

For this progress report, we have discussed our preliminary work, which has established three working algorithms for image classification. One major limitation of our preliminary work, is that we have not yet run full tests on the Pascal VOC 2012 dataset, we have used smaller sets of data to prove that our algorithms work. For our final report, we will evaluate the performance of our algorithms on the Pascal VOC 2012 dataset. We will then tune our parameters and identify the pros and cons of each model.

We will also compare our results to the results published in related work, as well as the best recorded results on the Pascal VOC 2012 dataset. We will also look to make improvements on our algorithms, using inspiration from algorithms in related work that perform better than ours.

Specifically, for our clustering with logistic regression, we have not yet implemented clustering. By adding K-Means preprocessing we hope to reduce the error rate.

An updated timeline for our project can be seen below:

Task	Due date	Responsible
Progress report submission	8-Nov	Team
Progress presentation	9-Nov	Team
Implementation of transfer learning on whole dataset	12-Nov	Team
Implementation of logistic regression on whole dataset	12-Nov	Team
Finalizing neural network on whole dataset	12-Nov	Team
Tuning parameters and improving algorithms	17-Nov	Team
Comparing our results to the results of related work	19-Nov	Team
Modifying report	20-Nov	Team
Finalizing Report	5-Dec	Team
Final report submission	6-Dec	Team
Final presentation	7-Dec	Team

VI. Conclusion

In conclusion, we have discussed our progress on our transfer learning, clustering with logistic regression, and neural network models. These algorithms have all been validated using smaller datasets. We plan to analyze our algorithms on the Pascal VOC 2012 dataset and compare our algorithms' results to the results of the best performing algorithms and results discussed in related work. We will then tune our algorithm's parameters and improve our algorithms based on related works.

References

[1] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[2] H. Gao and L. Yuan, "Research on key technology of pavement object recognition based on machine learning," *Neural Computing and Applications*, vol. 32, no. 10, pp. 5483–5493, 2020.

[3] E. Pintelas, M. Liaskos, I. E. Livieris, S. Kotsiantis, and P. Pintelas, "Explainable machine learning framework for image classification problems: Case study on glioma cancer prediction," *Journal of Imaging*, vol. 6, no. 6, p. 37, 2020.

[4] P. R. Jeyaraj and E. R. S. Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *Journal of cancer research and clinical oncology*, vol. 145, no. 4, pp. 829–837, 2019.

[5] P. M. Kumar, U. Gandhi, R. Varatharajan, G. Manogaran, R. Jidhesh, and T. Vadivel, "Intelligent face recognition and navigation system using neural learning for smart security in internet of things," *Cluster Computing*, vol. 22, no. 4, pp. 7733–7744, 2019.

[6] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[7] B. Hariharan, J. Malik, and D. Ramanan, "Discriminative decorrelation for clustering and classification," in *European Conference on Computer Vision*. Springer, 2012, pp. 459–472.

[8] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 132–149.

[9] W. Ge, S. Yang, and Y. Yu, "Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1277–1286.
[10] S. Shetty, "Application of convolutional neural network for image classification on pascal voc challenge 2012 dataset," *arXiv preprint arXiv:1607.03785*, 2016.
[11] S. Bell, C. Lawrence Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2874–2883.
[12] M. Huh, P. Agrawal, and A. A. Efros, "What makes imagenet good for transfer learning?" *arXiv preprint arXiv:1608.08614*, 2016.
[13] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717–1724.