Image classification for advertisements using deep learning

Chun Hin Attic Lee
Applied Machine Learning
001183418

***Abstract*—Image classification unlike traditional machine learning, it requires more complex model and training and it is obviously more challenging. However, the reward potential application of this field is tremendous. In this work, two convolutional neural network (CNN) model, VGG16 and EfficientNetB1 were implemented for the advertisement image classification. Multiple machine learning technique were employed to enhance to model performance (e.g. Upsampling, Data augmentation) Data augmentaion also served as a systematic experimentation for the 2 methods. The performance of the models did not increase significantly as the data are class imbalance and the data quality are not satisfied for the CNN network.**

I. Introduction and related work

Image classification is challenging for machine to deal with, unlike human, machines do not have the ability to interpret the picture, notice the details, understand the context of a picture as a human does. However,it is an important task for the society as it has enormous potential and a lot of people can possibly get benefits from it. For instance, image classification can analyse medical images , which can facilitate the process of diagnosis and benefits both patients and doctors [1]. Applications of Image classification is everywhere in the society and strongly related to our daily lives. For example, Facebook's face recognition features in their social media, and facial recognition in the airport for faster boarding and security [2]. Image classification is also crucial in business world as company can make more powerful marketing strategy such as user preferences or trends to produce a better advertisements to attract people [3]. The development and employment of Image Classification has facilitate the work of complicated tasks and enhanced our quality of live.

In this work, a dataset of advertisements images will be used. The images in the dataset are in different size and the raw data has 39 possible topic categories. The original dataset has 5984 images and 39 different labels. After exploratory data analysis and data cleaning(which will be discussed below), the final dataset has 5630 and 17 labels for training set, test set, and validation set.

Image normalisation, upsampling, and image augmentation were carried out in this work as data preprocessing to improve the performance of the model. In this work, 2 machine learning implementation was conducted:(1) VGG16 model [4], (2)EfficientNetB1 (EffNet) [5]. Both implementation are CNN neural network which is the best practice on image classification. VGG16 served as a baseline model in this work as it is the state of the art CNN model that is widely available for transfer learning , EfficientNetB1 is a much stronger model with more layer, to enhance the accuracy for this work.

Image augmentation is the systematic experiment for both neural network. It creates various different version of a image to improve the performance of the model when encountered imbalance classes, small dataset, and prevent overfitting.

II. Ethical discussion

As Artificial Intelligence and machine learning evolves and permeates, it has been influencing our societies and our lives in multiple ways. Machine learning requires large amount and diverse type of data, the data collection for this purpose creates decent impacts to the society. Personal information are often used in machine learning such as computer vision and image classification, which risen the concern from the public whether the data collecting process has violate the ethical and legal boundaries [6].

Social started realised how important and how valuable their data is,people nowadays became more aware of their personal data that would potentially be collected and their usage, how would the data be collected, how would the data be used, who owns the data, etc. At a same time, people are embracing the machine learning system that utilised their data in a certain extends. There are numerous of application and technologies that used machine learning algorithm to improve their product and provide more convenience and useful tools or service to the social. For example, Siri, the personal assistant in everyone's iPhone, is driven by Deep Neural Network (DNN) to detect the user speech and keep improving with learning the user's voice [7] [8]. Moreover, facial recognition such as Facebook's DeepFace [9] and Google's Google AI [10] also provide a beneficial tools to the users which also 'trading' their personal data for a better service and product [11]. These type of applications not only used large dataset for processing and make machine learning prediction that benefits the user, but also continuously improving when people enjoy and have great experience in the products.The social rose the awareness of the data being used for image classification, but they also in charge of they own data and decided how it should be used. The tremendous growing and development of machine learning also creates controversy of privacy, and the legalisation of that topic. The image classification and ethical challenges on privacy, became an important topic and decisive factor for customers to choose their products and more legalisation has been made to protect our privacy [12].

III. Dataset preparation

In the raw data, there are 5984 images in the original dataset with 39 different labels , 111 images with missing values was found, no duplicated value was found during the exploratory data analysis. However, some disrupted images was discovered as there are 111 images that has 20 pixels of height and 20 pixels of width only, which is unusable for the machine learning process. For the labels distribution among the 39 labels, some category has a much larger amount of images than other category. For example, the 'clothing' category has 825 images in the dataset while 'gambling' only has 1 image in the dataset. In addition, there are one label called 'unclear' in the dataset, which is some images have not been categorised. As a result, this category can also considered as missing values.

In data cleaning, missing value including the 111 images and images with 'unclear' category was dropped during the data cleaning process. Labels of the dataset has also been re-categorised. As the dataset size is not large enough for the neural network model to achieve high accuracy, and the imbalanced distribution of the data among the labels. Some labels can be group into one category. For example, 'soda', 'coffee', 'alcohol', 'chocolate', 'chips', 'petfood', 'seasoning', 'baby', and 'cleaning' can be categorised into 'grocery', 'healthcare', 'other_service', and 'phone_tv_internet_providers' can be categorised to 'service'. As a result, the 39 labels reduced into 17 labels, most of it are grouped, some are excluded from the dataset such as 'unclear' category and the label 'charities' were also being excluded since it cannot be grouped into any of the categories and have small amount of data (i.e. less than 10 images). Image that belongs to 'security' are also regrouped into 'entertainment' with 'game' and 'gambling' since the image of this category are too ambiguous, that is not ideal for the neural network as it would be confused with another category.

The images had also been resized before get into training as the image size of the dataset are different, resize to 256 width and 256 height and keeping the image original aspect ratio with padding.

Before data splitting, label encoding is essential to ensure the data of labels are machine-readable data for machine learning algorithms to be operated more efficiently. One-hot encoding was carried out in this work to convert categorical data into numeric form by creating multiple dummy columns for every variables in the 'category' column into binary matrix. After one-hot encoding, the data are ready for data splitting. The dataset was split into train set, test set, and validation set. The dataset was first split into a 80% train and validation set , and 20% of test set. Then further splitting it into train set with a 80% and 20% for the validation set. The data were being shuffled to ensure that data have random distribution of data and avoid bias in the data.

The conducted data pre-prosseing practices were: (1) Data normalisation, by changing the range of pixels value in image to more normal, Normalising the images are crucial that the pixel values in image must be scaled and normalised from originally between 0 to 255 to 0 to 1 before input into the neural network, which reduced the redundancy of the data and ensure the input data has similar distribution. (2) Upsampling was performed in this work, as mentioned above, the dataset has a problem of class imbalance, upsampling duplicate images from the minority classes, to achieve a balanced classes among the dataset. (3) Image augmentation, which is a method that are use for enlarging the training dataset. Unlike upsamling, instead of making duplicate copy of the images, it creates variations of the image such as flipping the image upside down, flipping it back to front, adjusting the contrast, and adjusting the brightness of the images. This process improves and allows the model to generalise to the data.

IV. Methods

VGG16 is a convolution neural network (CNN) architecture deep learning model that has shown its significant in the computer vision work. VGG [4] was used as the baseline model of in this work. As VGG16 is the oldest state of the art CNN model that is widely available for transfer learning in Tensorflow. Transfer learning is a method in machine learning that where using a model that has already been trained for related problem as the beginning of the current task. As convolution neural network in computer vision requires long time to train on large dataset, straightly using a well-developed model for our own problem or integrate with a new model is an ideal approach when dealing with image classification problem [13].

V. Experiments and evaluation

Image augmentation was implemented as the systematic experimentation for this work. The images in the training set were duplicated with variations: the images will be randomly flipped upside down, back to front, random brightness with max delta 0.2, random contrast with 0.5 lower bound and 2.0 upper bound, random saturation with 0.75 lower bound and 1.25 upper bound, and random hue of the image with 0.1 max delta.

The purpose of implementing this technique as a systematic experimentation is to the tackle the imbalanced class problem of the dataset, through image augmentation to improve the diversity of the data and enhance the performance of the models. However, the outcome did not meet the expectation, both the VGG model and EfficientNetB1 model do not have significant improvement, it remains the similar level of performance. This might because of the original dataset is not only have class imbalanced problem, but also there are a lots of ambiguous images that the neural network might not be able to learn all the pattern. In addition, the dataset does not have large enough of data for the neural network. The machine learning metrics that used for evaluating the deep learning models are: Top k accuracy score, F-1 score, and accuracy score. Top k accuracy score is the number of time that the neural network predicted the correct labels among the top 5 labels, which is suitable for this task as this task is a multiclass classification problem with more than 15 classes. F-1 score was used because the correct prediction of the labels is important in this work. F-1 score is the harmonic mean of recall and precision score, it directly reflect how the performance of the prediction of the true positive result of the model. Accuracy score is used, as it calculate the rate of predicting the label correctly, which is also the objective of this task.

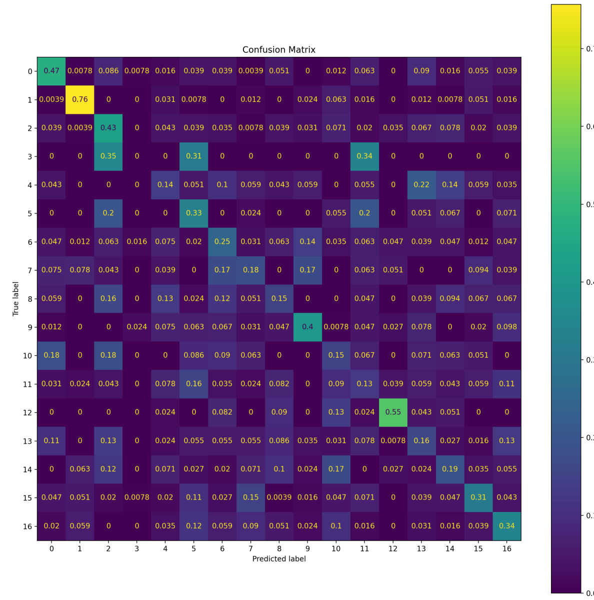

Figure 1. Confusion Matrix of VGG16 model. X-axis refers to prediction labels and Y-axis refers to the true labels .More brighter the colour shows it predicts the correct label of the class. Brighter the boxes in the diagonal, indicates a good performance of the model as it can predict the correct classes.

Table I
METRICS OF THE VGG16 AND EFFICIENTNETB1 MODEL AND THEIR SYSTEMATIC EXPERIMENTATION.

	VGG16	VGG16 w/Aug	EffNet	EffNet w/ Aug
Top K Accuracy	0.630	0.635	0.658	0.662
F1-score	0.282	0.264	0.333	0.327
Accuracy	0.290	0.277	0.329	0.320

For the evaluation, refers to Table 1, VGG16 model, the baseline model of this work, has 0.630 top k accuracy score, F1-score is 0.282, and accuracy score is 0.290. After the deployment of image augmentation, the top k accuracy score is 0.635,F1-score is 0.264 and accuracy score is 0.277.
The EfficientNetB1(EffNet) model has shown better performance than VGG16 model, with top k accuracy 0.658, 0.333 F1-score and 0.329 accuracy score before tuning. After the deployment with image augmentation, top k accuracy score is 0.662, F1- score is 0.327 ,accuracy score is 0.320.

One of the reason why the image augmentation did not work as expected, might be the data quality is not good enough for the neural network to learn. Therefore, despite the image augmentation is commonly known as a reliable tool to deal with small dataset, overfitting, and class imbalance, it still cannot improve the final peformance of the model. For instance, if the data are a blank white image, even if we flipped it, adjust the brightness of it, tuning the contrast of the image, the neural network performance will still remain the same because it could not learn from those meaningless data. In fact, the perfomance of the models had slightly improved after image augmentation, however it only had such influence on some of the classes, which means some data/images in those classes has better quality and the nerual network are more capable learn from those data, hence some classes on the training set has reasonable improvement on the performance of both VGG16 and EfficientNetB1 model, while some are still remain poor performance.

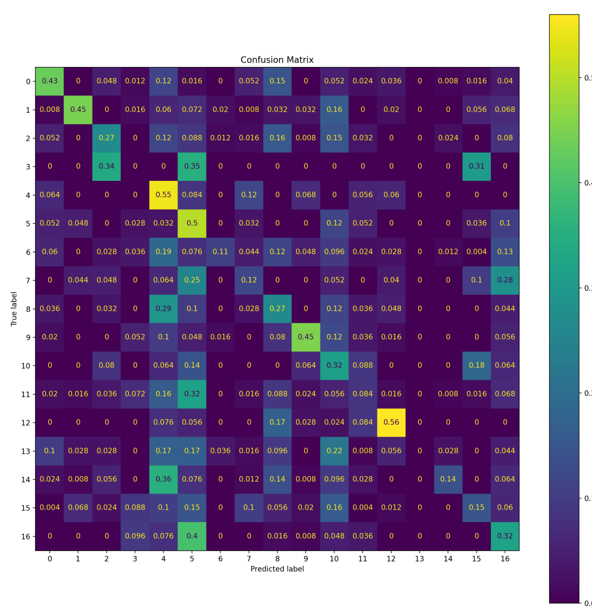

Figure 2. Confusion Matrix of VGG 16 model with image augmentation. It shows that the has slightly improvement on True Positives, but also True Negative.

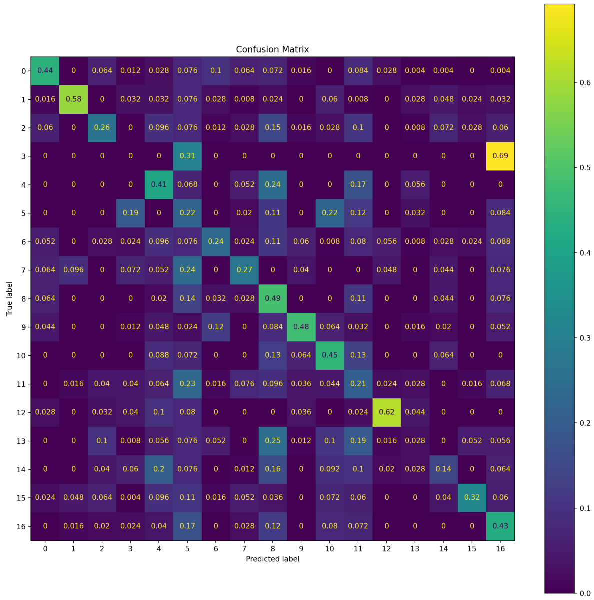

Figure 3. Confusion Matrix of EfficientNetB1 model . It shows that the has slightly improvement on True Positives, but also True Negative.

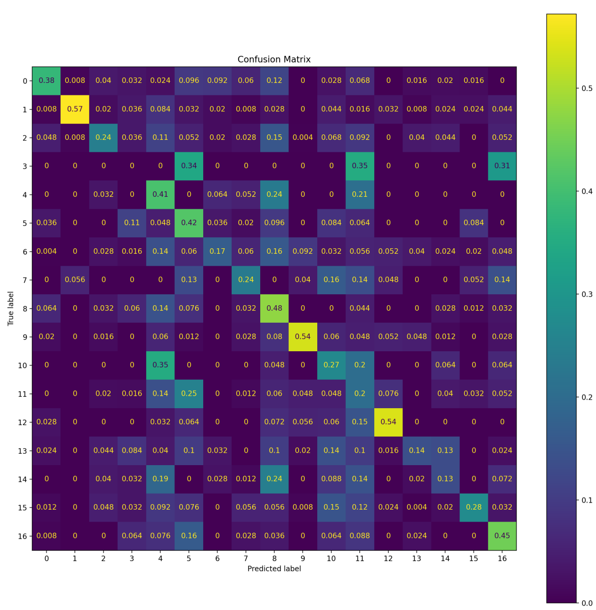

Figure 4. Confusion Matrix of EfficientNetB1 model with image augmentation. It shows that the has slightly improvement on True Positives, but also True Negative.

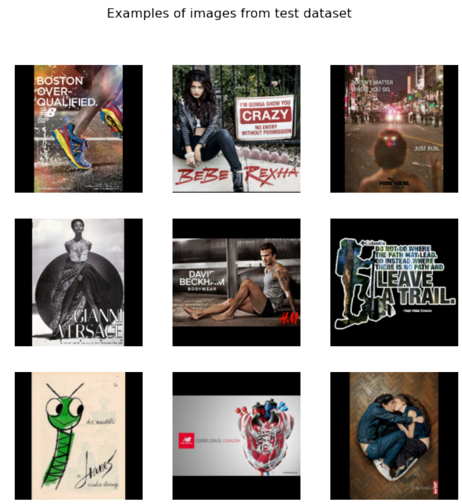

Figure 5. 9 examples images from the test dataset. The size of image has been resize in to identical size and remain the original aspect ratio of the image using padding

VI. Discussion and future work

This work has demonstrated the typical exploratory data analysis, data cleaning and data preparation on computer vision task. Multiple prepprossessing has been carried out to improve the performance such as upsampling and image augmentation. Yet, the performance still remain in a low level even despite most of the technique has been applied to attempt to boost the performance of the model. Due to the small amount of data, low variation in data per label and the ambiguity advertisement images, this model was not a robust enough handle these ambiguous advertisement images. As machine learning is a iterative process, further improvement of the model can be started from data collection. By collecting more data and data with better label quality, can positively improve the model performance.

VII. Conclusions

This work is a great demonstration of the real world use case of image classification and highlights the potential problems encounter during the development of machine learning system. It includes many technique of imagine classification such as image normalisation, image resize, neural network, etc. The highlight of this work is that even using the state of the art CNNs, with poor quality data, the machine learning model stills under perform. Thus, this emphasis the importance of data quality.

References

[1] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, 2014, pp. 844–848.

[2] Altexsoft, "Image recognition with deep neural networks and its use cases," AltexSoft, 12 2019. [Online]. Available: https://www.altexsoft.com/blog/image-recognition-neural-networks-use-cases/

[3] D. Baier, I. Daniel, S. Frost, and R. Naundorf, "Image data analysis and classification in marketing," *Advances in Data Analysis and Classification*, vol. 6, pp. 253–276, 10 2012.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[5] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
[6] C. F. Kerry, "Protecting privacy in an ai-driven world," Brookings, 02 2020. [Online]. Available: https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/
[7] S. Team, "Hey siri: an on-device dnn-powered voice trigger for apple's personal assistant," Apple Machine Learning Research, 10 2017. [Online]. Available: https://machinelearning.apple.com/research/hey-siri
[8] ——, "Personalized hey siri," Apple Machine Learning Research, 04 2018. [Online]. Available: https://machinelearning.apple.com/research/personalized-hey-siri
[9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
[10] G. AI, "Our approach to facial recognition – google ai," Google AI, 2021. [Online]. Available: https://ai.google/responsibilities/facial-recognition/
[11] J. C. Wong, "Google reportedly targeted people with 'dark skin' to improve facial recognition," The Guardian, 10 2019. [Online]. Available: https://www.theguardian.com/technology/2019/oct/03/google-data-harvesting-facial-recognition-people-of-color
[12] B. Morgan, "Whatsapp controversy shows just how much privacy matters to customers," Forbes, 06 2021. [Online]. Available: https://www.forbes.com/sites/blakemorgan/2021/01/14/whatsapp-controversy-shows-just-how-much-privacy-matters-to-customers/?sh=6ba821e8ff33
[13] J. Brownlee, "Transfer learning in keras with computer vision models," Machine Learning Mastery, 05 2019. [Online]. Available: https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/