GOTHENBURG UNIVERSITY

MACHINE LEARNING FOR STATISTICAL NLP: ADVANCED

LT2326

Can Detect: A Beer Can Classification Model

Author:
Bogdan LASZLO

1 Introduction

1.1 Motivation and Goals

The visual diversity and complexity is something often overlooked, even though it plays and important part in consumer choice. This project explores the potential of neural networks to recognize and interpret these aesthetic cues: shapes, colors, and designs, unique to each beer can. As beer cans shapes are for the most part uniform, the only possible way of identifying the various features is by looking at the visual design, namely the shapes and colours on the beer can itself. Thus, the goal of this experiment is to showcase the importance of beer can designs in signaling attributes to potential customers, such as brew type and taste and how they can also be perceived by neural networks as well.

I chose to undertake such a project as I notice that there wasn't any publicly available dataset concerning beer cans in particular and that it would be an interesting challenge for me to create the dataset myself. Not only that, but I find the applicability of neural networks in helping customers choose the best product for their tastes as something to strive for, as I personally consider that finding utilisation for the latest technologies and paradigms fit for the daily life aspects is an important factor in truly democratising it.

1.2 Background and State of Knowledge

The project's is based around the conception that beer manufacturers intentionally embed visual cues pertaining to their products attributes within beer cans, in order to signal and appeal to potential costumers based on their preferences. The importance of beer can design is underlined by [Vrontis, 1998] who states that "brewers rely on differentiation to create consumer preferences" as the beer market due to its nature is hard to innovate and there is a fierce competition among those participating in it. This sentiment is also shared by [Pilone et al., 2023] who presents in their paper that the product attributes that customers appreciate the most are can packaging and product taste. Based on these findings, I based my hypothesis on the manufacturer's incentive to invest significantly in creating an appealing can packaging signaling its taste, taste which is inherent to its brew type.

Figure 1: Showcase of beer can design elements

A similar experiment to the one found within this report was done by [Hohlfeld et al., 2022], which focused primarily on detecting beer bottles from various environments and classifying them by brand. Within it they employed a ResNet-18 model alongside SVMs for detection and classification, obtaining a detection and classification accuracy of almost 100%. Besides this, multiple other research forays have been done on the applications of computer vision into the field of food tech. A few notable research projects include [Dutta et al., 2016] with a novel approach using K-means clustering of of pixels based on their colors in order to measure the freshness of fish and [Gonzalez Viejo et al., 2016] where a feed-forward neural network was employed to asses images of beer foam in order to determine alcohol level.

2 Data

2.1 Data Collection

The dataset utilized within this experiment was collected specifically for it as publicly available resources proved to be insufficient for the amount of specificity that was needed for this project. Thus, in order to obtain the necessary data, images were collected both on-site from retailers (*System Bolaget* and *Hemköp*), the *System Bolaget* website and from public review websites such as *Untappd*, *Beermatch* and *Pint Please*. The resulting dataset contains 177 unique beers, 2167 beer images and 3068 total beer occurrences. As the intention was to publish the website publicly, the images were anonimized beforehand, removing identifiable characteristics that could bring harm to people's privacy.

The resulting images were annotated manually based on widely available information, with the following features being specified: *brand*, *name*, *brew_type*, *alc_content*, *keyword_1*, *keyword_2*, *keyword_3*, *bounding box*. The keyword features were selected by me, based on official information from the manufacturer when available, otherwise relying on information from *System Bolaget*'s website or from the previously mentioned review websites. The keywords are meant to offer a structured representation of the beer's taste, which is represented by this 3-dimensional variable with multiple values. The taste attribute is represented by the resulting multiple combinations of these flavour-specific keywords.

2.2 Data Processing

Before being used, the data was subjected to various techniques meant to reduce the amount of redundant data for this experiment in particular and to maximize the usability of the relevant data. Firstly, the images were cropped based on their *bounding box* attribute, in order to extract the image information relating only to the cans and their design. Afterwards, only the attributes *brew_type* and the taste keywords were kept, as the focus of the experiment falls on predicting taste-related attributes based on images.

Lastly, data augmentations techniques were applied upon the dataset in order to increase its suitability for training a model. Each extracted image was duplicated and then subjected to several transformations including hue change, rotation, and contrast change, thus providing enough *new features* to the image so that the new image would be fit for training on. The augmentations were applied only to the training dataset to minimize any potential overfitting. Thus, creating a training dataset of 4908 images and leaving 307 images each for both the validation dataset and the testing dataset.

3 Model

3.1 Architecture

The model's architecture is based on Boa's multi-label image classification model [Boa, 2023], which employs a ResNet15 model to run multi-label classification on various images of clothing, predicting attributes relating to clothing item type and color.

I have chosen to split the model in half, essentially creating 2 models, one for multi-class classification, targeting brew types and another one for multi-label classification, focusing on the taste keywords. This was strictly a choice motivated by my desire to keep the complexity low for training time reasons, but also to assure the acquisition of results in a timely manner.

For the multi-class model, I am using an 18-layer ResNet architecture, with 13 convolutional layers, 2 fully connected layers and 3 pooling layers. It uses the ADAM optimizer with a weight decay of 0.0001 and a maximum learning rate of 0.001. The fit function serves as the training loop, handling both training and validation for a set number of epochs. As the model is designed around predicting classes, it uses a *softmax* output activation function.

The model also incorporates batch normalization and ReLU activation after each convolutional layer, which helps in stabilizing the learning process and adding non-linearity to the model's capabilities. One of the key features of the ResNet architecture is the use of residual connections, which allow for easier training of deeper networks by enabling the flow of gradients directly through the network. This is particularly important in a model with 18 layers, as it helps to mitigate the vanishing gradient problem.

In addition to the standard components of the ResNet architecture, I have introduced modifications to adapt it more effectively to the specific requirements of my dataset. This includes adjusting the filter sizes and stride values in convolutional layers to better capture the features of the input images. The pooling layers, strategically placed after certain convolutional layers, aid in reducing the spatial dimensions of the feature maps, thereby decreasing the computational load and the risk of overfitting.

For the second model, the architecture is largely the same, utilizing a ResNet model of 18 total layers, the main difference being the utilisation of a *sigmoid* output activation function, as the goal of the model is to independently predict the taste labels.

3.2 Results

Concerning the results of the models, they seem to be in line with [Hohlfeld et al., 2022]'s findings, the model exhibiting a very high rate of correctly labeling and classifying the data.

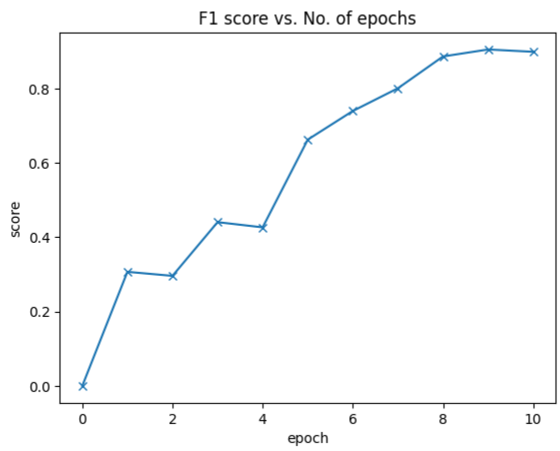

Figure 2: F1 score plotted over the number of epochs

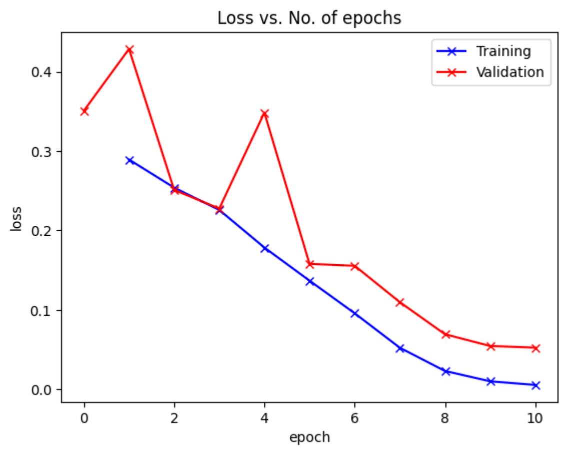

Figure 3: Loss plotted over the number of epochs

The multi-class model exhibited in figure 2 shows a final F1 score of 0.8993, accurately predicting the brew types of almost all beers in the testing dataset. Starting from an F1 score of just above 0.3, the score shows a sharp improvement in the initial epochs, indicating that the model is rapidly learning from the data. After the fifth epoch, the increase in the F1 score becomes more gradual, suggesting that the model continues to learn and improve its classification ability, but with diminishing returns as it approaches its performance ceiling. The trend plateaus slightly after the seventh epoch, with even a slight decrease at the tenth epoch, ending with an F1 score near 0.9. This high F1 score denotes a strong balance between precision and recall in the model's predictions, illustrating that the network is highly effective in accurately classifying the different types of beer brews by the end of the observed training period. The error bars, although relatively small, indicate some variance in the score across different training runs, which is normal in machine learning experiments.

The chart in figure 3 depicts the loss for both training and validation sets of a neural network model that classifies beer brew types across 10 epochs. The validation loss begins high and drops sharply after the first epoch, indicating rapid learning and significant improvement in the model's predictions. As the epochs progress, both training and validation losses continue to decline, suggesting that the model is generalizing well and not merely memorizing the training data. The validation loss closely tracks the training loss, which is a positive sign that the model is not overfitting. The convergence of the two losses and their consistent downward trend point to a stable training process. By the end of the 10th epoch, the losses have flattened out, implying that additional training may yield minimal gains in performance. The close alignment between the training and validation loss curves also reflects a well-tuned learning rate and model complexity that is well suited for the given task.

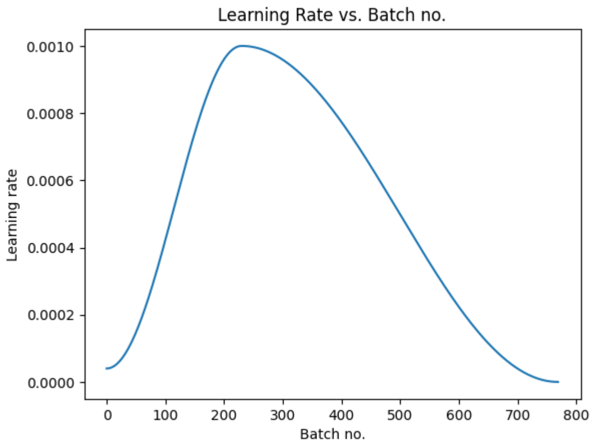

Figure 4: Learning rate plotted over the number of batches

Figure 4 illustrates the learning rate adjustment over a series of batches in training the neural network model for classifying the beer brew types. The learning rate starts at a low value, increases to a peak around the 200th batch, showing an accelerated learning phase where the model is be searching the solution space. After reaching the peak, the learning rate decreases steadily, which leads to a more refined approach to updating the model weights, helping the model in fine-tuning its parameters and converging to an ideal set of weights. Employing the one-cycle learning rate policy, is beneficial for training this model as it combines the advantages of both high and low learning rates, leading to better generalization and faster convergence. The smooth curve and the absence of sharp spikes indicate a well-controlled learning rate schedule throughout the training process.

The multi-label model's performance, as depicted in figure 5, presents a strong upward trend in accuracy across training epochs, culminating in an impressive F1 score of 0.9504 by the final epoch. This high score is indicative of the model's proficiency in labeling the flavors within the testing dataset. Initially, the model's F1 score starts at approximately 0.45, due to the initial learning curve. It then rises abruptly, though a subsequent dip at epoch 2 hints at a potential issue with one of the data batches. Despite this, the model quickly stabilizes and shows consistent improvement, suggesting a reliable performance with little variance. The consistent upward trajectory after the initial fluctuation points to the model's increasing stability and effectiveness in learning from the data throughout the training process.

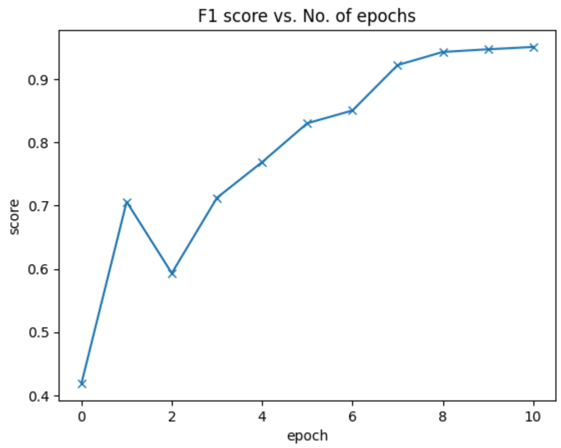

Figure 5: F1 score plotted over the number of epochs

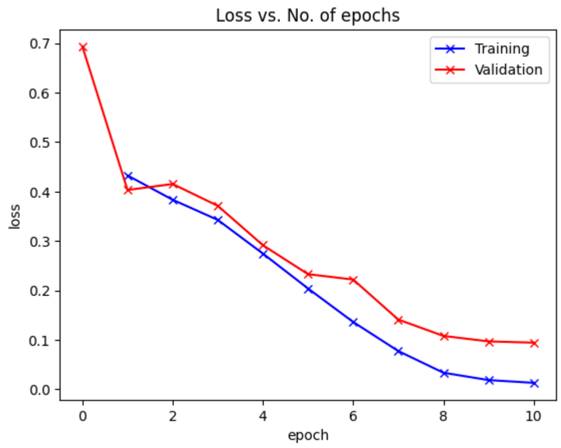

Figure 6: Loss plotted over the number of epochs

Figure 6 exhibits the loss values for both the training and validation phases of the model over 10 epochs. The training loss starts at a high value and decreases sharply, indicating rapid learning initially. Both training and validation losses continue to decrease over epochs, following a similar downward trajectory, which suggests that the model is generalizing well and not overfitting. The convergence of training and validation losses, along with the relatively smooth descent, ensuring a stable training process. The model's loss reduction slows down as it approaches the later epochs, typical of a learning process nearing its optimal state.

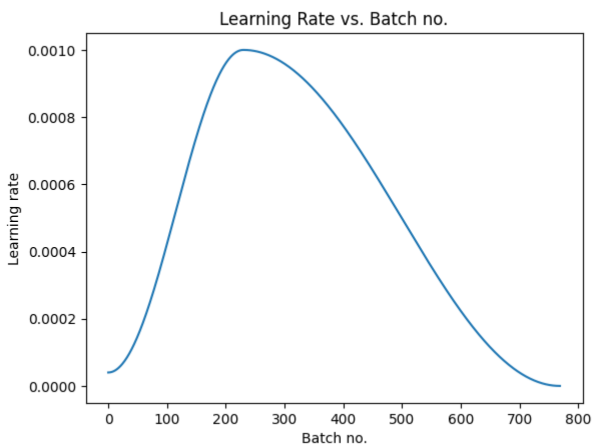

Figure 7: Learning rate plotted over the number of batches

The chart at figure 7 illustrates a learning rate scheduling strategy over the course of processing batches. It shows that the learning rate starts at a low value, gradually increases to a peak around the 200th batch, and then declines smoothly. This pattern is shows the initial ramp up of the learning rate to traverse the loss quickly and then decrease it to fine-tune the model's weights. The peak in the learning rate suggests a point where the model is allowed to explore the parameters more broadly before it begins to converge on a solution by gradually lowering the rate, thus reducing the learning steps to refine its parameters. This approach lead to a faster convergence and resulted in an adequate model performance by allowing the model to avoid local minima during the higher learning rate phase and then settle into a more optimal solution during the decay phase.

4 Conclusion

The experiment can be considered a success, as it fulfilled its expected goals and performance, that is to show that there is a direct and quantifiable link between a beer can's design and its inherent taste characteristics. Not just that but the model's excellent performance highlight just how intertwined is the beer's taste to its package design, as it can be picked-up easily even by a relatively simple neural network model, despite the apparent complexity of correctly representing flavour. This showcases in an objective way the amount of work that goes into creating such beer cans design, as they are not random, but instead are focused on delivering as much of the product's characteristics' information to a potential customer via visual means.

Regarding potential improvements, such models can be made more complex, allowing them to predict multiple design features directly, rather than through the 2 model approach presented in this report. Thus, not only could the brew type and the flavour be detected, but also the alcohol content, the brand and the beer's product name as well. Such a model could be further adapted so that it would match a customer's set preferences to a list of beers most closely matching the user's tastes, thus increasing their overall product satisfaction, by cutting down potential ambiguity in the customer's decision making process. This could be coupled with a more complex taste-quantifying framework, using multiple key-words, with a wider variety of labels meant to confer a higher degree of granularity when determining a beer's taste, for better describing and matching a beer to a compatible customer.

Overall, the model which I have showcased through this experiment could lay the groundwork for a more comprehensive one, which would find great use within the greater food-tech industry, bringing about a greater push towards personalized content in regards to beer consumption.

References

[Boa, 2023] Boa, E. (2023). Multi-label image classification pytorch gpu. Kaggle.

[Dutta et al., 2016] Dutta, M. K., Issac, A., Minhas, N., and Sarkar, B. (2016). Image processing based method to assess fish quality and freshness. *Journal of Food Engineering*, 177:50–58.

[Gonzalez Viejo et al., 2016] Gonzalez Viejo, C., Fuentes, S., Li, G., Collmann, R., Condé, B., and Torrico, D. (2016). Development of a robotic pourer constructed with ubiquitous materials, open hardware and sensors to assess beer foam quality using computer vision and pattern recognition algorithms: Robobeer. *Food Research International*, 89(1):504–513.

[Hohlfeld et al., 2022] Hohlfeld, P., Ostermeier, T., and Brandl, D. (2022). Classification of beer bottles using object detection and transfer learning. *arXiv preprint arXiv:2201.03791*.

[Pilone et al., 2023] Pilone, V., Di Pasquale, A., and Stasi, A. (2023). Consumer preferences for craft beer by means of artificial intelligence: Are italian producers doing well? *Beverages*, 9(1):26.

[Vrontis, 1998] Vrontis, D. (1998). Strategic assessment: The importance of branding in the european beer market. *British Food Journal*, 100(2):76–84.