**Fashion Item Classification**

**Introduction**

The digital era has brought about a transformative approach to retail, particularly in the fashion sector. The ability to quickly and accurately sort fashion items is no longer a luxury but a necessity for businesses to thrive in the online marketplace. Correctly categorized products enhance user navigation, promote efficient stock handling, and facilitate targeted marketing strategies. Addressing the challenge at its core, this paper presents the use of Convolutional Neural Networks (CNNs) as a solution for precise and swift fashion item classification, a crucial element for the success of e-commerce platforms.

Dataset: <https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small>

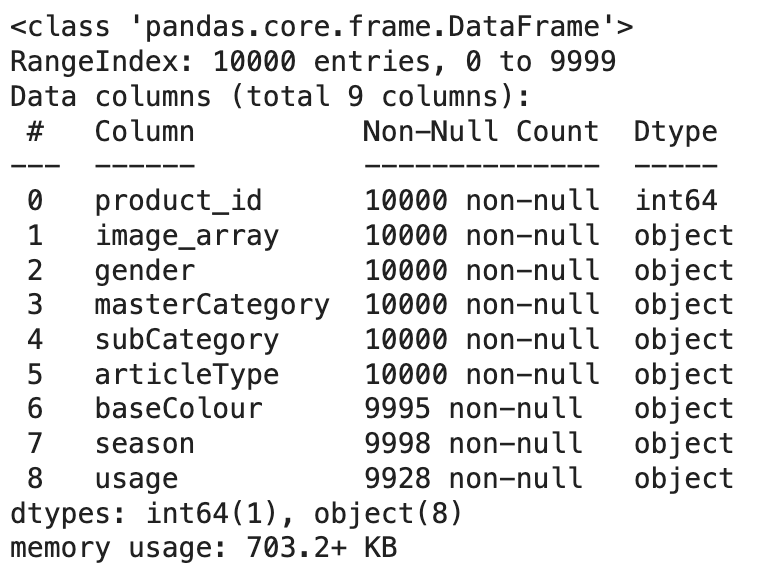
**Methodology**

*Data Handling Challenges:*

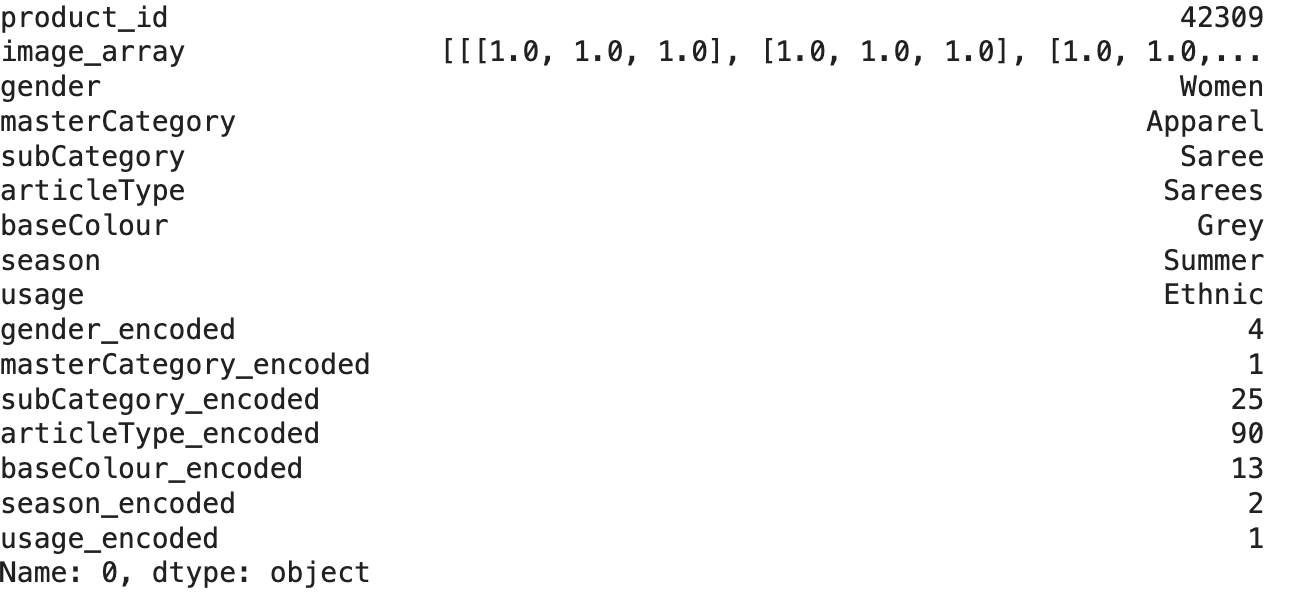
To effectively manage the large number of images required for our project, we developed a compression strategy. By compressing all images into a single ZIP file, we were able to minimize the space they occupied and simplify their transfer to Google Drive for storage. Once uploaded, we used Google Colab to decompress these files and organize them into a structured directory, making them readily accessible for the training process. This method ensured that our training environment was both efficient and well-organized, facilitating smoother model training sessions.

*Data Preprocessing:*

Our dataset includes a rich variety of images sourced from a leading fashion retail catalog. To prepare this data for the CNNs, each image was resized to a uniform resolution of 64x64 pixels. Our final raw data frame consisted of the following columns:



An example row looks like the following:



This row corresponds to the following image:



**Model Architectures**

* Model 1: Predicting the gender of the item.

*Target Classes:* ‘Male’, ‘Female’, ‘Boy’, ‘Girl’, ‘Unisex’. Total: 5.

Architecture: This model is structured primarily to predict gender based on images. It employs an initial set of convolutional layers with 32 filters each, capturing basic features such as edges and textures. This is followed by a max pooling layer to reduce the spatial dimensions and a dropout layer to minimize overfitting by randomly disabling 25% of the neurons. The network then flattens the pooled feature maps to form a single vector, which is fed into a dense layer of 128 neurons, again followed by a 50% dropout for further regularization. The output layer utilizes a softmax activation function, suitable for multi-class classification, which might assume various gender categories or related features.

* Model 2: Classify images into master categories.

*Target Classes:* 'Apparel', 'Footwear', 'Accessories', 'Personal Care', 'Free Items', 'Sporting Goods'. Total:6

Architecture: This model increases in complexity, beginning with 64 filters and expanding to 128 and then 256 in subsequent convolutional layers, enhancing its ability to capture more complex patterns. Max pooling and dropout layers are strategically placed to progressively reduce overfitting and dimensionality. The flattened output is processed through a dense layer with 256 neurons, reflecting the model's capability to manage complex data categorizations. The final softmax layer classifies the output into various master categories, indicating a broad classification scope.

* Model 3: Classifying subcategories within the master categories.

*Target Classes:*'Saree', 'Shoes', 'Flip Flops', 'Bottomwear', 'Topwear', 'Bags', 'Innerwear', 'Ties', 'Lips', 'Watches', 'Nails', 'Dress', 'Socks', 'Belts', 'Sandal', 'Headwear', 'Fragrance', 'Jewellery', 'Eyewear', 'Wallets', 'Accessories', 'Scarves', 'Loungewear and Nightwear', 'Makeup', 'Free Gifts', 'Skin', 'Eyes', 'Water Bottle', 'Skin Care', 'Apparel Set', 'Sports Equipment', 'Gloves', 'Cufflinks', 'Mufflers', 'Stoles', 'Hair', 'Shoe Accessories', 'Sports Accessories', 'Perfumes', 'Wristbands', 'Bath and Body', 'Umbrellas'. Total: 42.

Architecture: It mirrors the complexity of Model 2 but adapts for higher specificity required for subcategory differentiation. It also incorporates both dropout and L2 regularization in its dense layers, which are crucial for preventing overfitting given the finer granularity of classification expected. The softmax activation in the output layer manages the multi-class classification among the numerous subcategories.

* Model 4: Classifying subcategories within the master categories. (Demonstrating overfitting)

*Target Classes:* 'Saree', 'Shoes', 'Flip Flops', 'Bottomwear', 'Topwear', 'Bags', 'Innerwear', 'Ties', 'Lips', 'Watches', 'Nails', 'Dress', 'Socks', 'Belts', 'Sandal', 'Headwear', 'Fragrance', 'Jewellery', 'Eyewear', 'Wallets', 'Accessories', 'Scarves', 'Loungewear and Nightwear', 'Makeup', 'Free Gifts', 'Skin', 'Eyes', 'Water Bottle', 'Skin Care', 'Apparel Set', 'Sports Equipment', 'Gloves', 'Cufflinks', 'Mufflers', 'Stoles', 'Hair', 'Shoe Accessories', 'Sports Accessories', 'Perfumes', 'Wristbands', 'Bath and Body', 'Umbrellas'. Total: 42.

Architecture: This is an advanced configuration, potentially combining several features or targeting higher accuracy while classifying subcategories within the master categories. It uses a deep network architecture starting from 64 filters and extending up to 512. The heavy use of dropout, going up to 50%, suggests that it is designed to handle either a very large dataset or extremely complex features where overfitting is a significant risk. Multiple dense layers with a high count of neurons and additional dropout emphasize the model's capacity to learn and manage complex patterns deeply.

* Model 5: Classifying seasonality

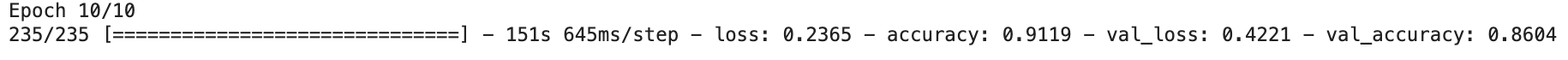
Target Classes: 'Summer', 'Fall', 'Winter', 'Spring', nan. Total: 4.

Architecture: This model has the same architecture with Model 3 which was the most successful model so far. We wanted to see if the high accuracy while classifying master categories would translate into high accuracy while predicting seasonality of fashion items.

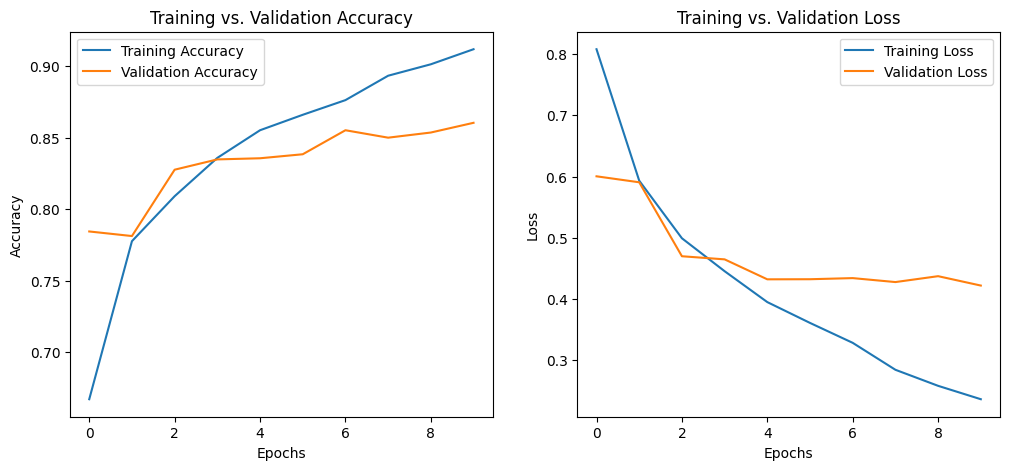
\*\*Across all models, ReLU activation is used for intermediate layers due to its efficiency in non-linear transformations and mitigating the vanishing gradient issue. Softmax is consistently employed in the output layer for multi-class classification across diverse categories. Regular use of dropout and max-pooling across the models helps in regularization and reducing spatial dimensions, respectively.

**Results**

* ***Model 1:***

******

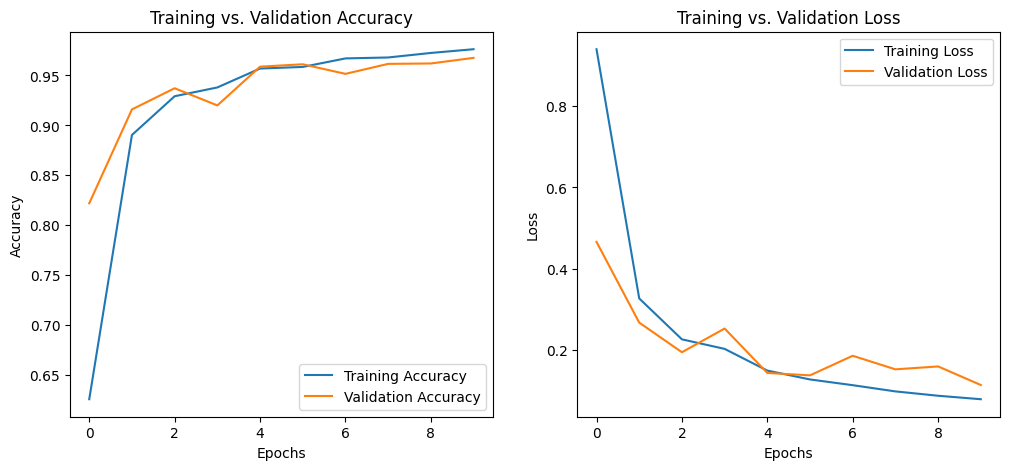
Achieved an accuracy of 86%, demonstrating the effectiveness of even a relatively simple CNN architecture. The close alignment of training and validation accuracies suggests that the model generalized well across new data, effectively avoiding overfitting.



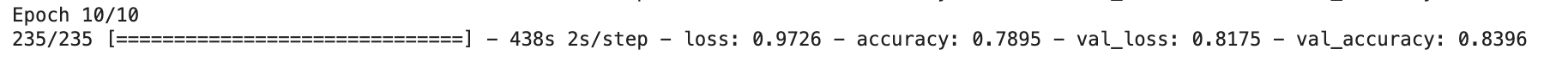
* ***Model 2:***

******

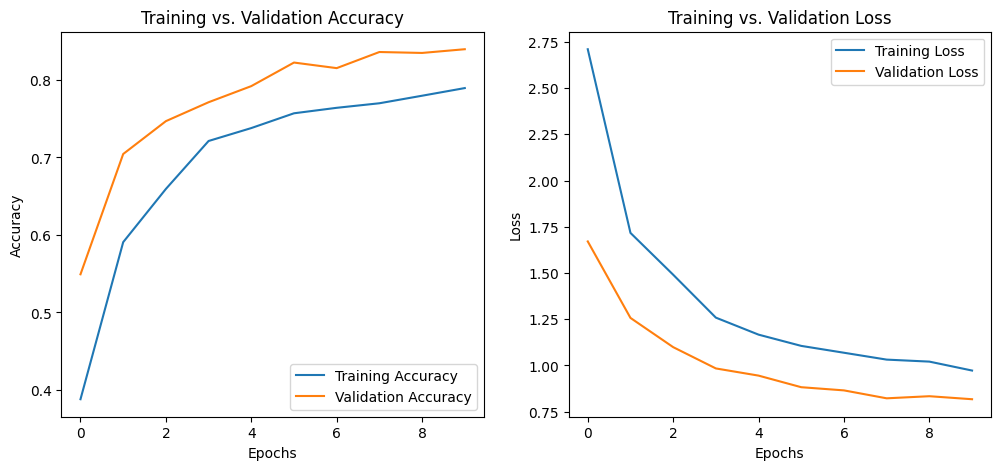
Showed an improvement in accuracy, supporting our hypothesis that a more complex network with additional filters is better equipped to capture the intricate details and structures within fashion images. The consistent decrease in loss throughout the training phases indicates effective learning and the model's improved ability to generalize from training data to unseen data. This model was the model that reached the highest validation accuracy.



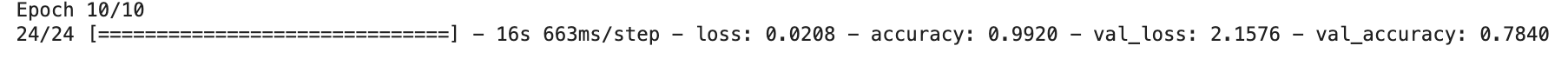
* ***Model 3:***

******

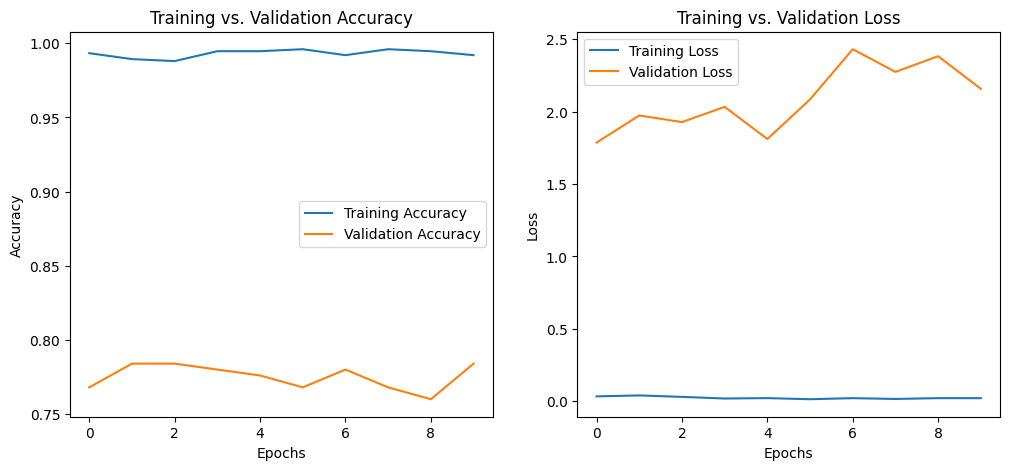
Although the validation performance was generally good, the appearance of a plateau in the accuracy suggests that there may be limits to further improvements in model complexity using the current architecture. The big drop in the accuracy shows how an increase in number of target classes makes it more difficult for the model to classify.



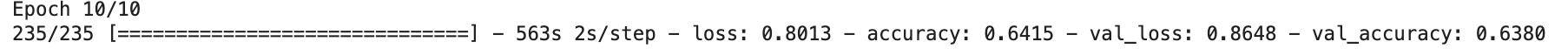
* ***Model 4:***



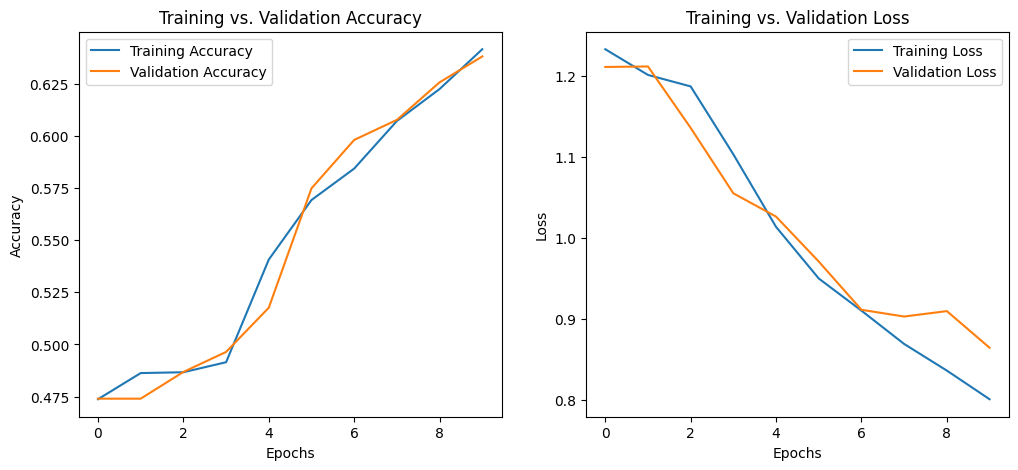
This model was built to show the effects of overfitting when the model is made too complex. The big gap between the training and validation accuracies show that modern does not generalize but instead memorizes the patterns and the noise of the training data.



* ***Model 5:***



Our final model was built on our most successful model so far, model 2. We decided to use this model to see if it can also differentiate between seasonality of the fashion items using the same architecture. The accuracies, both for training and validation, are relatively low, peaking at just over 62%. Such figures suggest that while the model is learning and improving, the overall ability of the model to accurately classify fashion items is modest. This level of accuracy indicates that the model may struggle with the complexity of the task or lack sufficient features to make more precise predictions.



**Conclusions**

As a baseline, Model 1's 86% accuracy underscores the potential of relatively straightforward CNN architectures to achieve substantial classification efficacy. Moving to Model 2, an enhancement in complexity resulted in the highest validation accuracy, reaffirming the hypothesis that a denser network with more filters can more adeptly capture complex image features. Model 3, however, encountered the bounds of architectural refinement, signaling diminishing returns with increased class targets, despite maintaining good validation performance. This was elucidated further by Model 4, which exhibited classic signs of overfitting, evidenced by a substantial disparity between training and validation accuracies. Model 5's attempt at seasonality classification using the architecture from the most successful Model 2 faced considerable challenges, indicating the nuanced nature of such classification tasks and suggesting areas for further research and model adjustments. Together, these outcomes demonstrate a diverse range of responses to CNN model complexities, offering valuable lessons for optimizing network architectures in the dynamic field of AI-assisted fashion retail.

To conclude, the findings suggest that while deeper networks are more capable of detecting detailed patterns, they must be meticulously regulated to prevent overfitting and to maintain their applicability to real-world data. Looking forward, these insights pave the way for future research focused on integrating various machine learning methodologies and leveraging pre-trained models. Such approaches could potentially unlock new capabilities in the automation and enhancement of fashion classification systems, further revolutionizing the field of fashion retail.