Image Classification for Fashion Items

**Introduction**:

In this the fast-paced world of fashion retail, the rapid and precise categorization of apparel and accessories is not just a convenience—it's a vital component of market leadership. By leveraging automated systems powered by cutting-edge image classification technology, retailers can vastly improve operational workflows and elevate the shopping experience to new heights. Our initiative is aimed towards this transformation, deploying state-of-the-art algorithms to identify and classify a diverse array of fashion items through image analysis. This allows for a more streamlined inventory management process and a user-friendly online shopping environment, setting a new standard in the fusion of fashion and technology.

**Problem**:

The problem was chosen to address the challenges many retailers face in manually categorizing large volumes of fashion items, which is time-consuming and prone to error. We aimed to automate this task which would speed up tasks and ensure consistency and accuracy in product categorization across platforms. This advancement has profound implications for resellers like Depop and Goodwill, which are inundated with an array of clothing items and face challenges in sorting and classifying them effectively. By automating these processes, we open new avenues for these companies to optimize their operations and capitalize on their vast inventory potential.

**Dataset**:

Our analysis is based on the Glami-1M dataset which comprises of over 1million labeled fashion images of fashion products from various online retail sources as well as 191 categorizations of the images and their descriptions. This dataset represents the vast complexity of the current fashion retail landscape. To streamline our approach and enhance the focus of our analysis, we mapped these categories into 28 general groupings, establishing a more manageable framework for model training. From this rich repository, we curated a representative 10% sample to train our models, thus laying down a robust foundation that captures the wide array of fashion items. This strategic selection is designed to improve the accuracy and reliability of our classification efforts, ensuring a model that is both versatile and adaptable to the multifaceted nature of fashion products.

**Methodology**:

Our approach to data analysis involved several stages:

* Data Standardization: The initial step involved processing the raw image data into a consistent format, resulting in standardized embeddings for input into various models, thereby enhancing the efficiency and effectiveness of the training phase.
* Model Experimentation:
  + **Random Forest Classifier**: We initiated our experiment with this model, applying it to data transformed into NumPy array format. It posted an initial accuracy score of 38%. However, through hyperparameter tuning, we fine-tuned its performance to reach an accuracy of 36.6%. The best results came from a decision tree depth of 100, requiring at least one sample per leaf, a minimum split of five, and a forest size of 300 trees (n\_estimators).
  + **Multi-Layer Perceptron**: Next, we used an **MLP** with three linear layers interspersed with two ReLU layers to train our model. This yielded a validation accuracy of 41%. (Appendix 2)
  + **Convolutional Neural Networks (CNN)**: A significant uptick in performance was observed upon deploying CNNs, which are particularly suited for image data. It achieved a validation accuracy of 79% and a test accuracy of 71% when trained on grayscale image, showcasing their robust feature extraction capabilities. The model’s capability decreased when trained on RGB images. (Appendix 3)
  + **Vision Transformers (ViT)**: The culmination of our experiments was the application of Vision Transformers, which set a new benchmark by attaining an accuracy of 85% (Appendix 4). This result underscored the transformative impact of attention-based models in processing complex image datasets.

For a detailed view of the CNN's performance metrics, please refer to our W&B report: <https://api.wandb.ai/links/ahmedfkbu/g1w6x4m7>

**Impact and Relevance**:

The image classification model we developed holds significant potential for entities dealing with large-scale fashion inventories:

* **Operational Streamlining**: By automatically classifying new stock, our model significantly reduces reliance on manual sorting. This not only speeds up inventory management but also introduces a new level of precision and efficiency in the categorization process.
* **Market Responsiveness**: Our model's ability to identify trends and quickly classify items like those in high demand empowers businesses to dynamically adapt their marketing strategies. This real-time market adaptability facilitates the development of finely tuned, data-driven marketing initiatives.
* **Enhanced Customer Experience**: With the integration of image-based search capabilities, customers can effortlessly find the products they're looking for, transforming their shopping journey into a seamless and enjoyable experience. This capability is not just a convenience; it’s a strategic enhancement that can drive sales and foster customer loyalty.

The strategic advantage afforded by our model goes beyond mere operational improvements; it is a transformative tool that can redefine market engagement strategies and drive growth, all while elevating the end-user experience to unprecedented heights.

**Challenges and Future Work**:

We encountered certain constraints and identified key areas for future exploration.

* Our examination was limited to 10% of the initial dataset, suggesting the potential for even greater accuracy and insights from a more expansive data set.
* The dataset's descriptive text remains unexplored. By applying Natural Language Processing (NLP) or advanced clustering techniques, we could enhance the model's classification capabilities, adding another layer of sophistication to the algorithm.
* Although Vision Transformers have shown remarkable results, their computational demands present a challenge. A balance must be struck between the efficiency offered by CNNs and the superior performance of Transformers.
* Autoencoders present a promising research direction, particularly in their potential to determine the authenticity of fashion products—a significant concern in the luxury goods sector.
* We encountered a strategic decision point where we had to choose between conducting hyper-parameter tuning for the CNN and MLP or integrating a novel model like the Vision Transformer due to computational constraints. Opting for the latter, we witnessed our model's accuracy surge to 85%. This significant improvement validated our choice, confirming that the introduction of the Vision Transformer was a pivotal and justified step in our methodology.

**Conclusion**:

This project presents the transformative potential of machine learning in reshaping the fashion retail landscape. By deploying image classification models, retailers can anticipate operational enhancements, strategic marketing advantages, and a superior customer interface. Our research, laying the groundwork for this technology's application, opens doors to new possibilities for efficiency and growth in the industry.

**Appendix 1**:

Sample images from our dataset

A long grey coat on a swinger

Description automatically generated A close-up of a watch

Description automatically generated A person in a vest

Description automatically generated

**Appendix 2**:

MLP Training Graphs

A graph of a graph of a graph

Description automatically generated with medium confidence

<https://api.wandb.ai/links/ahmedfkbu/f3savdhb>

**Appendix 3**:

CNN Training Graphs

A screenshot of a graph

Description automatically generated

<https://api.wandb.ai/links/ahmedfkbu/x62cku0o>

**Appendix 4**:

Transformer Evaluation

A black and white text

Description automatically generated

**Appendix 5**:

Code contribution

A screenshot of a computer

Description automatically generated