BINUS UNIVERSITY GRADUATE PROGRAM Master Program | Master of Computer Science

Image Classification for the Fashion Industry with Convolutional Neural Networks

Authors

Bob Sebastian
2502478723
bob.sebastian@binus.ac.id

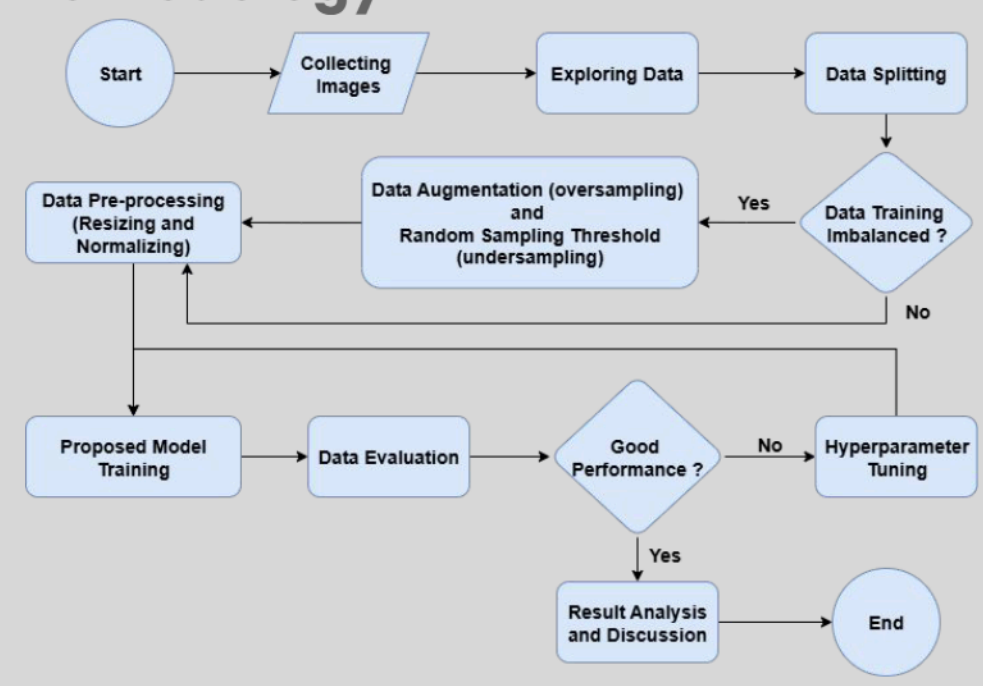
I Gede Putra Kusuma Negara, B.Eng., PhD
D5627
inegara@binus.edu

Background

In today's fast-growing e-commerce era, accurate product classification is vital, as misclassification can frustrate customers, reduce sales, and damage brand trust. The fashion industry is especially challenging since items vary in design, texture, color, and function, making manual categorization inefficient and inconsistent; for example, a handbag may sometimes be incorrectly labeled as a shoulder bag due to similar features (Smith et al., 2021). Fashion datasets also often suffer from **data imbalance**, where common categories like t-shirts or sneakers dominate while scarves or belts are underrepresented, causing model bias and poor minority-class performance.

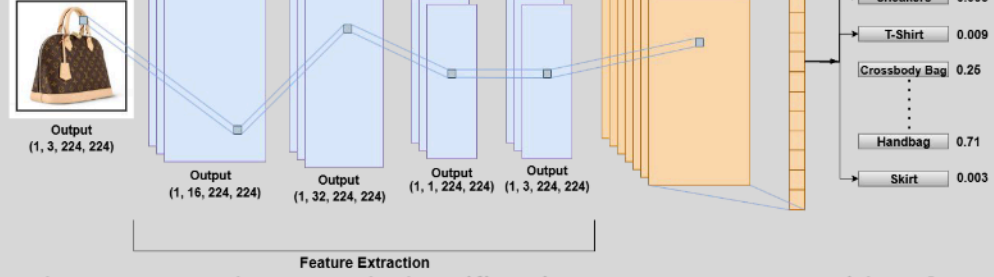
Deep learning, particularly Convolutional Neural Networks (CNNs), addresses these challenges by automatically performing **feature extraction** from raw pixels, learning hierarchical patterns, and recognizing subtle differences in texture, shape, and design (LeCun et al., 2015). CNNs reduce human error, improve scalability, enable real-time deployment in retail platforms, and mitigate imbalance with techniques like resampling or augmentation, ultimately enhancing classification accuracy, customer experience, and business decision-making (Zhang et al., 2022).

Methodology



This methodology outlines the process of building an image classification model using EdgeNet architecture for feature extraction. It begins with data collection and exploration, followed by splitting the dataset into training, validation, and testing sets. The data is carefully checked for imbalance, and data augmentation (oversampling) and random sampling (undersampling) are applied to balance it. The dataset is then pre-processed through resizing and normalization before being used to train the proposed model. After training, the model's performance is evaluated, and if results are unsatisfactory, hyperparameter tuning is performed iteratively to further improve performance. The process concludes with a final detailed analysis and discussion of the results.

Proposed Model

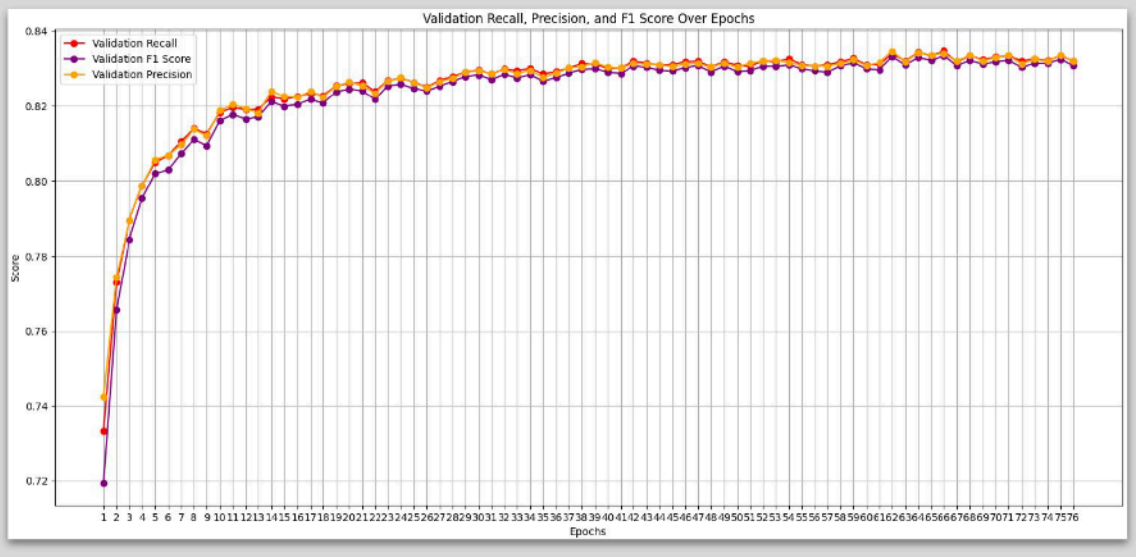


The proposed apparel classification system starts with **edge enhancement** using the **EdgeNet** architecture with three convolutional layers. The output is expanded to match the input requirements of the backbone model. The enhanced image is fed into 4 backbone model for **feature extraction**. Features are passed to a fully connected layer with 59 neurons representing the dataset classes. The model is fine-tuned from a pretrained network by replacing the final layer and retraining only the last layers using preprocessed and augmented apparel images. This allows the model to capture patterns textures and shapes in apparel images and improves classification accuracy on the target dataset.

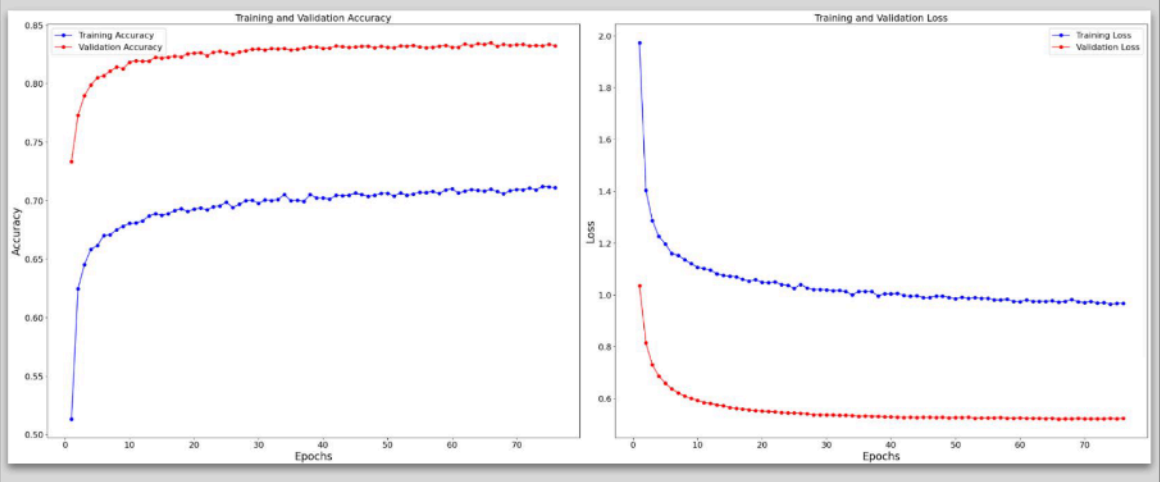
Result and Discussion

Epochs	Batch Size	Learning Rate	Dropout Rate	Weight Decay	*Optimizer*	Train		Validation	
						Accuracy	*Loss*	*Accuracy*	*Loss*
45	32	0.001	0.05	0.001	RMSProp	69.50%	0.98	82.80%	0.55
76	**32**	**0.001**	**0.005**	**0.01**	**Adam**	**71.19%**	**0.96**	**83.47%**	**0.52**
56	64	0.0001	0.1	0.001	SGD	68.80%	1.02	82.00%	0.57
58	64	0.0001	0.05	0.001	RMSProp	70.10%	0.97	83.00%	0.54

The model combines **BERT Pretrained Image Transformer (BeiT)** with **EdgeNet** and **data balancing**, which together enhance feature extraction and improve learning stability. EdgeNet strengthens edge information in images, allowing the backbone BeiT model to capture more meaningful visual patterns such as shapes, textures, and contours in apparel images. Data balancing ensures that all classes are equally represented during training, reducing bias toward majority classes and improving the model's ability to generalize across diverse apparel categories. Training accuracy ranges from 68.80% to 71.19%, while validation accuracy is higher, between 82.00% and 83.47%, showing that the model generalizes well beyond the training set. Training loss is moderately low (0.96–1.02), and validation loss is even lower (0.52–0.57), indicating the model learns efficiently without overfitting. The Adam optimizer with 76 epochs and batch size 32 produced the best result, achieving the highest validation accuracy of 83.47% with minimal loss, demonstrating that this combination effectively leverages pretrained knowledge, edge enhancement, and balanced data to maximize classification performance.



The validation evaluation of the BeiT model combined with EdgeNet and data balancing, using a learning rate of 0.0001 and the Adam optimizer, shows the development of recall, precision, and F1-score during training. Recall steadily increases, peaking around epoch 36, indicating the model is improving in identifying positive classes. Precision also trends upward, though with more fluctuations, suggesting occasional inaccurate positive predictions. The F1-score, as the harmonic mean of precision and recall, rises consistently and reaches its highest value at the end of training, demonstrating a good balance between precision and recall. Overall, these metrics indicate that the model becomes increasingly effective at classification, with all three measures showing significant improvement as training progresses.



The training of the BeiT model combined with EdgeNet and data balancing, using a learning rate of 0.0001 and the Adam optimizer, shows the progression of **accuracy and loss** over epochs. **Training accuracy** increases consistently, surpassing 40% by the final epoch, indicating that the model effectively learns from the training data. **Validation accuracy** also rises, though at a lower level and tends to plateau around 30%, suggesting room for further generalization. The training loss decreases significantly, reflecting stable learning, while validation loss also declines but more slowly. By the end of training, training loss reaches approximately 2.48, and validation loss is around 2.66, showing the model is learning effectively but still slightly underfitting the validation data.

Conclusion and Future Works

Overall, this study demonstrates that integrating **EdgeNet** as a feature extractor significantly enhances model performance, particularly when combined with **BeiT**, achieving the highest validation accuracy of **83.47%** with minimal loss. This confirms EdgeNet's effectiveness in optimizing feature extraction for BeiT. While data balancing positively influenced models such as Swin Transformer V2 and DeiT, its impact varied, highlighting that successful classification on imbalanced datasets relies on the **synergy of preprocessing, data balancing, and model architecture**. For future work, three key improvements are recommended: first, **fine-tuning hyperparameters** such as learning rate and dropout to optimize training and prevent overfitting; second, implementing **more diverse data augmentation** techniques to enrich dataset variability and enhance robustness; and third, addressing **imbalanced datasets** through advanced methods like SMOTE or Retrieval Augmented Graph to ensure fair class representation. Combining these strategies is expected to produce a more **accurate, robust, and reliable classification model.**

Selected References

1. Akbari, N., & Baniasadi, A. (2023). EDGE-Net: Efficient Deep-learning Gradients Extraction Network. Department of Electrical and Computer Engineering, University of Victoria, Canada
2. Deldjoo, Y., Quadrana, M., Garzotto, F., & Cremonesi, P. (2023). A review of modern fashion recommender systems. ACM Computing Surveys, 56(4), 81. https://doi.org/10.1145/3561470
3. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239