Fashion Image Classification: A Comparative Study of CNN Architectures

Abstract

This study investigates the development and comparison of two Convolutional Neural Network (CNN) architectures for fashion image classification, specifically focusing on distinguishing between casual, formal, and smart casual wear. The research addresses the critical challenge of extreme class imbalance in fashion datasets, with a distribution ratio of 513:35:1 across the three categories.

The methodology encompasses two distinct approaches: a baseline CNN model utilizing fundamental deep learning components, and an advanced model incorporating transfer learning with MobileNetV2, comprehensive data augmentation, and optimization techniques. Both models were evaluated using a dataset of over 36,000 fashion images, with performance assessed through multiple metrics including accuracy, precision, recall, and F1-score.

Results revealed that the baseline model achieved superior overall performance (96.75% accuracy) compared to the advanced model (93.91% accuracy), challenging the assumption that more complex architectures necessarily yield better results. However, both models struggled significantly with minority classes, particularly the smart casual category, despite implementing techniques for handling class imbalance (Johnson and Khoshgoftaar, 2019).

The study's findings contribute to the growing body of research on deep learning applications in fashion retail, highlighting the persistent challenges of extreme class imbalance and the limitations of conventional solutions. As noted by Liu et al. (2016), while CNNs show promise in fashion classification tasks, achieving robust performance across all clothing categories remains a significant challenge, particularly when dealing with imbalanced real-world datasets.

1. Introduction

Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have emerged as powerful tools for tackling classification challenges, demonstrating remarkable success in various computer vision tasks (Dargan et al., 2020).

Fashion image classification presents unique challenges due to the complex nature of clothing items, varying styles, and the inherent subjectivity in fashion categorization. The task becomes particularly challenging when dealing with nuanced categories such as distinguishing between casual, formal, and smart casual wear. As noted by Vijayaraj et al. (2022), deep learning approaches have shown promising results in fashion design and classification tasks, but significant challenges remain in achieving robust performance across different clothing categories.

This study focuses on developing and comparing two distinct CNN architectures for fashion image classification:

1. A baseline CNN model utilizing fundamental deep learning components including Conv2D, Dense, and MaxPooling2D layers
2. An advanced model incorporating transfer learning, data augmentation, and sophisticated optimization techniques

The research is particularly significant as it addresses real-world challenges in fashion image classification, where class imbalance is a common occurrence. As highlighted by Deldjoo et al. (2023), modern fashion recommender systems heavily rely on accurate image classification, making the development of robust classification models crucial for practical applications.

Our methodology follows a systematic approach to model development and evaluation, with particular attention to:

- Data preprocessing and organization
- Model architecture design and optimization
- Performance evaluation across different clothing categories
- Analysis of the impact of class imbalance on model performance

The primary objectives of this study are to:

1. Develop and evaluate a baseline CNN model for fashion image classification
2. Implement and assess an advanced model
3. Compare and analyse the performance of both approaches
4. Identify key challenges and potential solutions in fashion image classification

This research contributes to the growing body of literature on deep learning applications in fashion retail, as discussed by Maurício et al. (2023), who emphasize the importance of comparing different architectural approaches in image classification tasks. The findings from this study provide valuable insights into the effectiveness of various deep learning techniques in handling real-world fashion classification challenges.

This report proceeds by detailing our methodology and model architectures in Section 2, presenting experimental results and comparative analysis in Section 3, discussing key lessons learned in Section 4, before concluding with our findings in Section 5, with particular emphasis on addressing the challenges of extreme class imbalance in fashion image classification.

2. Methodology

2.1 Data Preprocessing and Organization

The dataset utilized in this study consists of fashion product images categorized into three distinct classes: Casual, Formal, and Smart Casual. Initial analysis revealed significant class imbalance, with 34,406 Casual, 2,345 Formal, and 67 Smart Casual images. This imbalanced distribution presents a common challenge in real-world machine learning applications (Johnson and Khoshgoftaar, 2019).

The preprocessing pipeline began with data organization and validation. Images were systematically organized into class-specific directories using a custom function that maintained data integrity while establishing a structured hierarchy. This approach aligns with best practices for deep learning dataset organization (Chollet, 2021).

To prepare the images for model training, several preprocessing steps were implemented:

1. **Image Standardization**: All images were resized to 64x64 pixels, striking a balance between computational efficiency and maintaining sufficient detail for classification. This resolution choice is supported by similar studies in fashion image classification (Liu et al., 2016).
2. **Pixel Normalization:** Pixel values were normalized to the range [0,1] by dividing by 255, a standard practice that helps stabilize the training process and accelerate convergence (Goodfellow et al., 2016).
3. **Data Validation:** A robust error-handling mechanism was implemented for CSV parsing to handle potential inconsistencies in the metadata file, ensuring reliable data loading across different environments.

The dataset was split into training and validation sets using a 80:20 ratio, with stratification to maintain class distribution. This split ratio is commonly used in deep learning applications to provide sufficient training data while retaining a representative validation set (Zhang et al., 2021).

The severe class imbalance observed in the dataset (approximately 513:35:1 ratio) necessitated careful consideration during model development. This imbalance is particularly challenging for the Smart Casual class, which could lead to biased model performance, a common issue in real-world classification tasks (He and Garcia, 2009).

2.2 Baseline Model Architecture

The baseline Convolutional Neural Network (CNN) was designed with a straightforward architecture, following established principles for image classification tasks. The model architecture was implemented using TensorFlow's Keras API, which provides a robust framework for deep learning model development (Chollet, 2021).

The baseline architecture consists of the following key components:

```
model = Sequential([
  layers.Rescaling(1./255, input_shape=(64, 64, 3)),
  layers.Conv2D(16, 3, activation='relu'),
  layers.MaxPooling2D(),
  layers.Conv2D(32, 3, activation='relu'),
  layers.MaxPooling2D(),
  layers.Flatten(),
  layers.Dense(32, activation='relu'),
  layers.Dense(3, activation='softmax')
])
```

The model begins with a rescaling layer to normalize pixel values to the range [0,1], a fundamental preprocessing step that helps stabilize training (Rizvi, 2024). The architecture then implements two convolutional blocks, each consisting of a Conv2D layer followed by MaxPooling2D operations. This design choice aligns with standard practices in computer vision tasks, where initial layers learn low-level features while deeper layers capture more complex patterns (Rizvi, 2024).

The first convolutional layer employs 16 filters with a 3x3 kernel size, followed by a second layer with 32 filters, progressively increasing the feature depth while reducing spatial dimensions through max pooling. This progressive feature extraction approach is widely recognized in CNN architectures for its effectiveness in capturing hierarchical patterns in image data (Goodfellow et al., 2016).

The model concludes with a flatten operation followed by two dense layers: a hidden layer with 32 units and ReLU activation, and an output layer with softmax activation for three-class classification. This architecture was deliberately kept simple to establish a baseline performance metric.

The model was compiled using categorical crossentropy loss, appropriate for multi-class classification tasks, and the Adam optimizer, which has shown robust performance across various deep learning applications (Kingma and Ba, 2014). The training process utilized a batch size of 64 and ran for 5 epochs, with validation data used to monitor performance and prevent overfitting.

2.3 Advanced Model Development

Building upon the baseline model's foundations, the advanced architecture incorporated several sophisticated techniques to enhance performance and address the challenges identified in the initial implementation.

Transfer Learning Implementation

The advanced model leveraged transfer learning using MobileNetV2, a lightweight yet powerful CNN architecture pre-trained on ImageNet. This choice was motivated by MobileNetV2's efficient design using inverted residuals and linear bottlenecks, which has demonstrated strong performance while maintaining computational efficiency (Sandler et al., 2018). The implementation was structured as follows:

```
base_model = MobileNetV2(
    weights='imagenet',
    include_top=False,
    input_shape=(64, 64, 3)
)
base_model.trainable = False


model = Sequential([
    base_model,
    layers.GlobalAveragePooling2D(),
    layers.Dense(64, activation='relu'),
    layers.Dropout(0.5),
    layers.Dense(3, activation='softmax')
])
```

Data Augmentation Strategy

To address the significant class imbalance and enhance model generalization, a comprehensive data augmentation pipeline was implemented (Shorten and Khoshgoftaar, 2019):

```
data_augmentation = Sequential([
    layers.RandomFlip("horizontal"),
    layers.RandomRotation(0.2),
    layers.RandomZoom(0.2),
    layers.RandomBrightness(0.2),
    layers.RandomContrast(0.2),
])
```

This augmentation strategy was specifically designed to create realistic variations in fashion images while preserving class-specific features. The techniques employed are supported by research in fashion image classification, showing effectiveness in handling varying viewpoints and lighting conditions (Liu et al., 2016).

Optimization Techniques

The training process was enhanced through several optimization strategies:

1. **Learning Rate Scheduling:** A reduce-on-plateau scheduler was implemented to dynamically adjust the learning rate, following best practices for deep model training (Smith, 2017):

```
tf.keras.callbacks.ReduceLROnPlateau(
  monitor='val_loss',
  factor=0.2,
  patience=2,
  min_lr=1e-6
)
```

2. **Early Stopping:** To prevent overfitting, an early stopping mechanism was employed with a patience of 3 epochs, monitoring validation loss (Goodfellow et al., 2016).
3. **Dropout Regularization:** A dropout rate of 0.5 was applied before the final dense layer, a value shown to be effective in preventing co-adaptation of features during training (Srivastava et al., 2014).

The model was compiled using the Adam optimizer with an initial learning rate of 1e-3, which has shown robust performance in similar transfer learning applications (Kingma and Ba, 2014).

3. Experimental Results and Analysis

3.1 Performance Metrics

The evaluation of both models employed a comprehensive set of metrics to assess their classification performance, following standard practices in machine learning evaluation (Grandini et al., 2020). The metrics were calculated using scikit-learn's implementation to ensure standardization and reliability.

Baseline Model Performance

The baseline CNN achieved the following results:

Overall Accuracy: 0.9675

Per-class Accuracy:

Casual: 0.9874

Formal: 0.6846

Smart Casual: 0.0000

Classification Report:

	Precision	recall	F1-score
Casual	0.98	0.99	0.98
Formal	0.77	0.68	0.73
Smart Casual	0.00	0.00	0.00

The baseline model demonstrated high accuracy for the majority class (Casual) but struggled with minority classes, particularly Smart Casual. This performance pattern is characteristic of models trained on imbalanced datasets without specific countermeasures (He and Garcia, 2009).

Advanced Model Performance

The transfer learning-based model with data augmentation showed different characteristics:

Overall Accuracy: 0.9391

Per-class Accuracy:

Casual: 0.9994

Formal: 0.0313

Smart Casual: 0.0000

Classification Report:

	Precision	recall	F1-score
Casual	0.94	1.00	0.97
Formal	0.78	0.03	0.06
Smart Casual	0.00	0.00	0.00

Following Sokolova and Lapalme's (2009) recommendations for imbalanced classification evaluation, we analyzed the following key metrics:

1. **Precision and Recall:** These metrics are particularly important in imbalanced scenarios, as they provide insight into the model's ability to handle minority classes (Powers, 2020). The advanced model showed high precision for the Casual class but struggled with minority classes despite the implemented augmentation techniques.
2. **F1-Score:** As a mean of precision and recall, the F1-score provides a balanced measure of model performance. The significant variation in F1-scores across classes (ranging from 0.97 for Casual to 0.00 for Smart Casual) indicates persistent challenges in handling class imbalance (Chicco and Jurman, 2020).
3. **Confusion Matrix Analysis:** Visual representation through confusion matrices revealed patterns of misclassification, particularly between Formal and Casual categories, suggesting potential feature overlap between these classes (Visa et al., 2011).

The evaluation metrics were computed on a held-out validation set comprising 20% of the total dataset, ensuring unbiased performance assessment. This validation strategy aligns with standard practices in deep learning evaluation (Goodfellow et al., 2016).

3.2 Comparative Analysis

A systematic comparison of the baseline and advanced models revealed interesting patterns in their respective performances and highlighted key insights into the challenges of fashion image classification with imbalanced data.

Model Performance Comparison

The comparative analysis showed unexpected results:

1. Overall Accuracy
    - Baseline Model: 96.75%
    - Advanced Model: 93.91%

While the advanced model incorporated more sophisticated techniques, it actually showed a slight decrease in overall accuracy. However, as noted by Gandler (2020), accuracy alone can be misleading when dealing with imbalanced datasets, as models can achieve high accuracy by simply predicting the majority class.

Class-Wise Performance Analysis

	Baseline		Advanced	
Class	Precision	Recall	Precision	Recall
Casual	0.98	0.99	0.94	1.00
Formal	0.77	0.68	0.78	0.03
Smart Casual	0.00	0.00	0.00	0.00

The comparison reveals several key findings:

1. **Majority Class Handling:** Both models showed strong performance on the Casual class, with the advanced model achieving perfect recall but slightly lower precision. This aligns with findings from He and Garcia (2009) regarding the tendency of models to favour majority classes.

2. **Minority Class Performance:** The advanced model's performance on the Formal class showed an interesting trade-off - higher precision (0.78) but significantly lower recall (0.03) compared to the baseline. This phenomenon is consistent with research showing that transfer learning can improve feature discrimination but may still struggle with class imbalance (Zhuang et al., 2020).

Training Efficiency

The models showed these characteristics in their training patterns:

1. **Convergence Speed**

    - Both models achieved stable performance within 5 epochs

2. **Resource Utilization**

    - The advanced model, using MobileNetV2, required more computational resources
    - The data augmentation pipeline introduced additional overhead during training

Impact of Advanced Techniques

The implementation of sophisticated techniques produced mixed results:

1. **Transfer Learning:** While MobileNetV2 provided a strong feature extraction backbone, it did not improve the handling of class imbalance, supporting findings by Johnson and Khoshgoftaar (2019) about the limitations of transfer learning in addressing class imbalance.
2. **Data Augmentation:** Despite implementing comprehensive augmentation strategies, the severe imbalance (513:35:1 ratio) proved too extreme for these techniques alone to overcome, aligning with observations by Shorten and Khoshgoftaar (2019) about the limitations of data augmentation in extreme imbalance scenarios.

3.3 Discussion of Results

The experimental results revealed several significant insights into the challenges and complexities of fashion image classification with severely imbalanced datasets, while also highlighting important considerations for model architecture selection.

Analysis of Success Factors

The baseline model's relatively strong performance can be attributed to several factors:

1. **Architectural Simplicity:** The straightforward CNN architecture demonstrated robust learning capability for the majority class, supporting the principle that simpler models can sometimes outperform more complex ones when dealing with imbalanced data (Nielsen, 2015).
2. **Feature Learning**: Both models showed strong capability in learning discriminative features for the Casual class, which aligns with findings from Liu et al. (2016) regarding the effectiveness of CNNs in capturing fashion-related features.

Limitations and Challenges

Several significant challenges emerged during the study:

1. **Extreme Class Imbalance**

The 513:35:1 ratio proved to be a fundamental challenge that neither model could fully overcome. This observation aligns with research by Buda et al. (2018), who demonstrated that CNN performance deteriorates significantly when class imbalance exceeds certain thresholds. The complete failure to classify the Smart Casual class (0% accuracy in both models) exemplifies this limitation.

2. **Transfer Learning Limitations**

Despite the theoretical advantages of transfer learning, our results align with findings from Wang and Deng (2018), who noted that pre-trained models might not always transfer effectively when:

- The target domain differs significantly from the source domain
- Extreme class imbalance exists in the target dataset

3. **Data Augmentation Effectiveness**

The limited improvement from data augmentation techniques supports observations by Shorten and Khoshgoftaar (2019) that traditional augmentation methods may be insufficient for addressing extreme class imbalance without additional balancing strategies.

Comparison with Published Results

Our findings can be contextualized within the broader literature:

1. **Fashion Classification Performance**

Recent studies on fashion image classification have reported varying results:

- DeepFashion benchmark: 73.95% accuracy (Liu et al., 2016)
- Fashion-MNIST: 93.7% accuracy (Xiao et al., 2017)

Our models' performance (96.75% and 93.91% overall accuracy) appears competitive, but these comparisons must be considered carefully due to:

- Different dataset compositions
- Varying levels of class imbalance
- Different classification objectives

2. **Imbalanced Learning Context**

Our results align with findings from Johnson and Khoshgoftaar (2019) regarding the persistent challenges of deep learning with extreme class imbalance, particularly:

- The tendency to favour majority classes
- The limitations of conventional data augmentation
- The need for specialized techniques for handling severe imbalance

4. Lesson Learned

Key Insights

The experimental outcomes provided several valuable insights into deep learning approaches for imbalanced fashion image classification:

1. **Model Complexity Trade-offs**

Our findings support the observation by Occam's Razor principle in machine learning (Raschka, 2018) that increased model complexity does not necessarily lead to better performance. The simpler baseline model demonstrated comparable or better performance in some aspects compared to the more sophisticated approach.

2. **Class Imbalance Impact**

The extreme class imbalance (513:35:1) proved to be the most significant challenge, highlighting the limitations of conventional deep learning approaches. This aligns with findings from Krawczyk (2016), who emphasized that class imbalance remains one of the most pressing challenges in real-world applications.

Technical Challenges

Several technical challenges emerged during the study:

1. **Data Augmentation Limitations**

While data augmentation was implemented following best practices, its effectiveness was limited in addressing extreme class imbalance. This supports findings by Shorten and Khoshgoftaar (2019) that traditional augmentation techniques alone may be insufficient for severely imbalanced datasets.

2. **Transfer Learning Considerations**

The use of MobileNetV2 revealed that transfer learning success is not guaranteed, particularly when:

- Domain shift exists between source and target datasets
- Extreme class imbalance is present

This observation aligns with research by Wang et al. (2019) on the limitations of transfer learning in specialized domains.

5. Conclusion

This study has provided valuable insights into the challenges and complexities of developing deep learning models for fashion image classification, particularly in the context of severe class imbalance. Through comparison of baseline and advanced architectures, several significant conclusions have emerged.

The experimental results demonstrated that architectural complexity does not necessarily translate to improved performance in imbalanced classification scenarios. The baseline CNN model, with its simpler architecture, achieved comparable and in some cases superior performance (96.75% vs 93.91% overall accuracy) compared to the more sophisticated transfer learning approach. This finding aligns with the principle of parsimony in machine learning, as discussed by Raschka and Mirjalili (2019), who emphasize that simpler models often provide more robust solutions in real-world applications.

A critical observation from this study is the persistent challenge of extreme class imbalance in fashion datasets. Despite techniques such as transfer learning and data augmentation implemented, both models struggled significantly with minority classes, particularly the Smart Casual category. This outcome supports the findings of Johnson and Khoshgoftaar (2019), who identify class imbalance as a fundamental challenge in deep learning applications.

The research also highlighted the limitations of conventional approaches to addressing class imbalance. As noted by Zhou et al. (2020), while data augmentation and transfer learning are powerful tools, they may be insufficient when dealing with extreme imbalance ratios such as the 513:35:1 distribution observed in our dataset. This suggests the need for more specialized techniques specifically designed for handling severe class imbalance in fashion classification tasks.

In conclusion, while deep learning continues to show promise in fashion image classification, significant challenges remain in developing robust models capable of handling real-world data distributions. The findings from this study contribute to the growing body of knowledge in this field and provide valuable insights for researchers and practitioners working on similar challenges in fashion analytics and e-commerce applications.

References

1. Buda, M., Maki, A. and Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural networks, 106, pp.249-259.

2. Chicco, D. and Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21, pp.1-13.

3. Chollet, F., 2021. Deep Learning with Python Second Edition, Shelter Island.

4. Dargan, S., Kumar, M., Ayyagari, M.R. and Kumar, G., 2020. A survey of deep learning and its applications: a new paradigm to machine learning. Archives of Computational Methods in Engineering, 27, pp.1071-1092.

5. Deldjoo, Y., Nazary, F., Ramisa, A., Mcauley, J., Pellegrini, G., Bellogin, A. and Noia, T.D., 2023. A review of modern fashion recommender systems. ACM Computing Surveys, 56(4), pp.1-37.

6. Gandler, G.Z. (2020) 'Training models on imbalanced data', Towards Data Science. Available at: https://towardsdatascience.com/training-models-on-imbalanced-data-561fa3f842b5

7. Goodfellow, I., 2016. Deep learning.

8. Grandini, M., Bagli, E. and Visani, G., 2020. Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.

9. He, H. and Garcia, E.A., 2009. Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), pp.1263-1284.

10. Johnson, J.M. and Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. Journal of big data, 6(1), pp.1-54.

11. Kingma, D.P., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

12. Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Progress in artificial intelligence, 5(4), pp.221-232.

13. Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X., 2016. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1096-1104).

14. Maurício, J., Domingues, I. and Bernardino, J., 2023. Comparing vision transformers and convolutional neural networks for image classification: A literature review. Applied Sciences, 13(9), p.5521.

15. Nielsen, A., 2015. Neural networks and deep learning.

16. Powers, D.M., 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

17. Raschka, S., 2018. Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808.

18. Raschka, S. and Mirjalili, V., 2019. Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. Packt publishing ltd.

19. Rizvi, M.S.Z. (2024) 'Image Classification Using CNN', Analytics Vidhya Blog. Available at: https://www.analyticsvidhya.com/blog/2020/02/learn-image-classification-cnn-convolutional-neural-networks-3-datasets/

20. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

21. Shorten, C. and Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. Journal of big data, 6(1), pp.1-48.

22. Smith, L.N., 2017, March. Cyclical learning rates for training neural networks. In 2017 IEEE winter conference on applications of computer vision (WACV) (pp. 464-472). IEEE.

23. Sokolova, M. and Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information processing & management, 45(4), pp.427-437.

24. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), pp.1929-1958.

25. Vijayaraj, A., Vasanth Raj, P.T., Jebakumar, R., Gururama Senthilvel, P., Kumar, N., Suresh Kumar, R. and Dhanagopal, R., 2022. Deep learning image classification for fashion design. Wireless Communications and Mobile Computing, 2022(1), p.7549397.

26. Visa, S., Ramsay, B., Ralescu, A.L. and Van Der Knaap, E., 2011. Confusion matrix-based feature selection. Maics, 710(1), pp.120-127.

27. Wang, J., Chen, Y., Yu, H., Huang, M. and Yang, Q., 2019, July. Easy transfer learning by exploiting intra-domain structures. In 2019 IEEE international conference on multimedia and expo (ICME) (pp. 1210-1215). IEEE.

28. Wang, M. and Deng, W., 2018. Deep visual domain adaptation: A survey. Neurocomputing, 312, pp.135-153.

29. Xiao, H., Rasul, K. and Vollgraf, R., 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747.

30. Zhang, C., Bengio, S., Hardt, M., Recht, B. and Vinyals, O., 2021. Understanding deep learning (still) requires rethinking generalization. Communications of the ACM, 64(3), pp.107-115.

31. Zhou, B., Cui, Q., Wei, X.S. and Chen, Z.M., 2020. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9719-9728).

32. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. and He, Q., 2020. A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(1), pp.43-76.