Conquering Fashion MNIST with CNNs using Computer Vision

Mentor :Vivekanandan M
Artificial intelligence and data science
Rajalaskmi Institute Of Technology
Chennai, India

Shankaran J
Artificial intelligence and data science
Rajalaskmi Institute Of Technology
Chennai, India
shankaran.j.2021.ad@ritchennai.edu.in

Ritesh kanna K H
Artificial intelligence and data science
Rajalaskmi Institute Of Technology
Chennai, India
riteshkanna.k.h.2021.ad@ritchennai.edu.in

Abstract:

This paper presents a novel approach to solving the Fashion MNIST dataset using Convolutional Neural Networks (CNNs) in the field of computer vision. The Fashion MNIST dataset has gained popularity as a benchmark for image classification tasks, with the aim of accurately categorizing fashion items. In this study, we explore the architecture and implementation details of CNNs, focusing on their effectiveness in achieving high accuracy on the Fashion MNIST dataset. We evaluate various CNN architectures, investigate the impact of hyperparameters, and analyze the challenges encountered during the process. The outcomes of this study highlight the potential of CNNs in conquering complex image classification tasks and provide valuable insights for future research in the field.

Keywords: Fashion MNIST, CNNs, Computer Vision, Conquering

I. INTRODUCTION

The fashion industry plays a significant role in our daily lives, offering a wide range of clothing options and styles. With the advent of e-commerce and online shopping, accurate and efficient classification of fashion items has become crucial for automating inventory management, improving recommendation systems, and enhancing the overall customer experience. Convolutional Neural Networks (CNNs) and computer vision techniques have emerged as powerful tools for image recognition tasks, enabling automated classification and analysis of visual data.

In this project, we explore the application of CNNs and computer vision techniques to conquer the Fashion MNIST dataset, which serves as a benchmark for image classification in the fashion domain. The Fashion MNIST dataset consists of 60,000 grayscale images, each of size 28x28 pixels, representing ten different fashion categories, including t-shirts, trousers, dresses, sneakers, and more. The dataset poses challenges such as varying lighting conditions, rotations, and scale differences, making accurate classification a complex task.

The objective of this project is to develop a robust system model that can accurately classify fashion items in the Fashion MNIST dataset. By leveraging the power of CNNs, we aim to extract meaningful features from the images and learn discriminative patterns that distinguish different fashion categories. The system model will undergo rigorous training, validation, and testing processes to optimize its performance and ensure high accuracy in classification.

Through this project, we aim to showcase the effectiveness of CNNs and computer vision techniques in conquering the Fashion MNIST dataset. By achieving high accuracy in fashion item classification, we contribute to the automation and efficiency of the fashion industry. The proposed system model has the potential to streamline inventory management, enhance recommendation systems, and improve customer satisfaction by providing accurate and personalized fashion choices.

In the following sections, we will conduct a comprehensive literature survey, outline the objectives and outcomes of the project, discuss the challenges encountered, present the architecture/system model, and detail the hardware/software model used for implementation. We will conclude by summarizing our findings and suggesting future research directions in the field of fashion classification using CNNs and computer vision techniques.

II. LITERATURE SURVEY

1. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In this seminal paper, the authors introduce the concept of deep convolutional neural networks and demonstrate their effectiveness in large-scale image recognition tasks. Their work laid the foundation for the application of CNNs in various domains, including fashion classification.

2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep

convolutional neural networks. This groundbreaking paper showcases the success of CNNs in the ImageNet Large Scale Visual Recognition Challenge. The authors present the AlexNet architecture, which achieved a significant improvement in image classification accuracy and paved the way for further advancements in deep learning.

3. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. This paper introduces the ImageNet database, which contains millions of labeled images across various categories. The authors provide insights into the challenges of large-scale image classification and the importance of quality datasets for training deep learning models.

4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. This influential paper presents the ResNet architecture, which introduces residual connections to address the degradation problem in deep neural networks. The authors demonstrate the superior performance of ResNet in image recognition tasks, inspiring further advancements in CNN architectures.

5. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. This classic paper introduces the concept of CNNs and demonstrates their efficacy in handwritten digit recognition. The authors lay the foundation for using CNNs as powerful tools for image classification and pattern recognition.

III. *OBJECTIVE*

1. Develop a robust system model: The primary objective of this project is to design and develop a robust system model that utilizes Convolutional Neural Networks (CNNs) and computer vision techniques to accurately classify fashion items in the Fashion MNIST dataset. The system model should effectively learn and extract meaningful features from the images, enabling accurate categorization of different fashion categories.

2. Achieve high classification accuracy: The project aims to achieve high accuracy in classifying fashion items in the Fashion MNIST dataset. The objective is to develop a system model that can accurately

predict the correct fashion category for a given image, minimizing misclassifications and improving the overall performance of the classification task.

3. Address real-world challenges: The project aims to tackle real-world challenges encountered in fashion item classification, such as varying lighting conditions, rotations, scale differences, and other factors that make accurate classification challenging. The objective is to develop strategies and techniques to overcome these challenges and improve the robustness of the system model.

4. Optimize the system model: The project aims to optimize the performance of the system model by exploring different CNN architectures, hyperparameter tuning, regularization techniques, and other approaches. The objective is to fine-tune the model to achieve the best possible accuracy and generalization on the Fashion MNIST dataset.

5. Improve automation in the fashion industry: By accurately classifying fashion items, the project aims to contribute to the automation and efficiency of the fashion industry. The objective is to develop a system model that can be integrated into inventory management systems, recommendation systems, and other fashion-related applications to streamline processes and enhance customer experiences.

6. Explore potential future applications: The project aims to explore potential future applications of CNNs and computer vision in the fashion domain. The objective is to identify and discuss how the developed system model can be adapted and extended for other fashion-related tasks, such as object detection, style recognition, and trend analysis.

IV. OUTCOMES

1. High accuracy in fashion item classification: The developed system model, leveraging CNNs and computer vision techniques, is expected to achieve high accuracy in classifying fashion items in the Fashion MNIST dataset. The outcomes will showcase the effectiveness of the model in accurately predicting the correct fashion category for a given image.

2. Improved automation in the fashion industry: By accurately classifying fashion items, the project outcomes will contribute to the automation and efficiency of the fashion industry. The developed system model can be integrated into inventory management systems, recommendation engines, and other fashion-related applications to streamline processes, enhance customer experiences, and improve operational efficiency.

3. Enhanced generalization and robustness: The outcomes of the project will demonstrate the ability of the system model to generalize well to unseen fashion images and handle variations in lighting conditions, rotations, and scale differences. The model's improved generalization and robustness will make it more reliable and applicable to real-world scenarios.

4. Identification of discriminative fashion features: Through the exploration and analysis of the CNN-based system model, the project outcomes will shed light on the discriminative features learned by the model for fashion item classification. These insights can provide valuable knowledge for the fashion industry, helping designers, retailers, and marketers understand the key visual cues and characteristics that contribute to the classification of different fashion categories.

V. CHALLENGES

1. Data preprocessing: Preparing the Fashion MNIST dataset for training the CNN model may pose challenges. The dataset may require preprocessing steps such as resizing, normalization, and handling missing or noisy data. Ensuring data quality and appropriate preprocessing techniques can impact the model's performance.

2. Model selection: Choosing the appropriate CNN architecture for fashion item classification can be challenging. There are numerous architectures available, and selecting the one that balances complexity, accuracy, and computational efficiency requires careful consideration. Comparing and evaluating different architectures can help address this challenge.

3. Hyperparameter tuning: Optimizing the hyperparameters of the CNN model is crucial for achieving optimal performance.

Tuning hyperparameters such as learning rate, batch size, number of layers, and filter sizes can be challenging and time-consuming. Employing techniques like grid search, random search, or Bayesian optimization can help in finding the optimal set of hyperparameters.

4. Overfitting: Overfitting occurs when the CNN model becomes too specific to the training data and fails to generalize well to unseen data. Preventing overfitting and finding the right balance between model complexity and generalization is a significant challenge. Applying techniques like regularization (e.g., dropout, L1/L2 regularization), early stopping, and data augmentation can help mitigate this challenge.

5. Generalization to unseen data: The ability of the model to generalize to unseen fashion images is crucial for its practical applicability. Ensuring that the CNN model can handle variations in lighting conditions, rotations, scale differences, and other factors present in real-world scenarios can be a challenge.

VI. *ARCHITECTURE*

1. Input Layer: The system model starts with an input layer that takes in grayscale images from the Fashion MNIST dataset. Each image is of size 28x28 pixels.

2. Convolutional Layers: A stack of convolutional layers is employed to learn local features and patterns from the input images. Each convolutional layer consists of multiple filters that perform convolution operations, extracting important features from the input. The number of filters and their sizes can be adjusted based on experimentation.

3. Activation Functions: Non-linear activation functions, such as ReLU (Rectified Linear Unit), are applied after each convolutional layer to introduce non-linearity and enhance the model's ability to capture complex patterns.

4. Pooling Layers: Pooling layers, typically using max pooling, are inserted to downsample the feature maps and reduce spatial dimensions. Pooling helps in extracting the most important features while reducing the model's

sensitivity to small spatial shifts or variations.

5. Dropout: Dropout regularization is applied to randomly deactivate a portion of the neurons during training, reducing overfitting and improving the model's generalization capability. Dropout can be applied after pooling layers to regularize the feature maps.

6. Fully Connected Layers: The feature maps are flattened and connected to fully connected layers to learn high-level representations. The fully connected layers capture global information from the feature maps, aiding in the final classification.

7. Output Layer: The final fully connected layer is connected to the output layer, which consists of ten neurons corresponding to the ten fashion categories in the Fashion MNIST dataset. The softmax activation function is applied to produce probabilities for each class, indicating the likelihood of the image belonging to each category.

8. Loss Function: The categorical cross-entropy loss function is used to measure the discrepancy between the predicted probabilities and the true labels. The goal is to minimize this loss during training.

9. Optimization Algorithm: The model is trained using an optimization algorithm such as stochastic gradient descent (SGD), Adam, or RMSprop. The algorithm updates the model's parameters based on the computed gradients, aiming to minimize the loss function.

10. you can check our coding for this project github here : link

VII. HARDWARE/ SOFTWARE MODEL FOR IMPLEMENTATION

1. CPU: A multi-core CPU is sufficient for running the code and training the CNN model. Any modern CPU with decent computational power can handle the Fashion MNIST dataset.

2. GPU: Although not mandatory, using a GPU (Graphics Processing Unit) can significantly accelerate the training process, especially when dealing with large-scale datasets or complex architectures. GPUs with

CUDA support, such as NVIDIA GeForce or Tesla series, can expedite the computations involved in training deep CNN models.

Software Model:

1. Programming Language: Python is widely used in the field of deep learning and computer vision. You can leverage the rich ecosystem of Python libraries and frameworks for implementing your project. Common choices include TensorFlow, Keras, and PyTorch.

2. Deep Learning Framework: TensorFlow and PyTorch are popular deep learning frameworks that provide high-level APIs for implementing CNNs and other neural network architectures. These frameworks offer GPU acceleration, automatic differentiation, and built-in functions for training, evaluating, and deploying deep learning models.

3. IDE (Integrated Development Environment): You can use any Python-compatible IDE for coding your project. Popular options include PyCharm, Jupyter Notebook, and Anaconda, which provide a user-friendly environment for writing, executing, and debugging your code.

4. Data Manipulation and Visualization Libraries: Libraries like NumPy, Pandas, and Matplotlib are essential for data manipulation, preprocessing, and visualization. These libraries allow you to perform operations on the Fashion MNIST dataset, preprocess the images, and visualize the results.

5. Image Processing and Computer Vision Libraries: Libraries like OpenCV and scikit-image provide a wide range of functionalities for image processing, augmentation, and computer vision tasks. These libraries can be used for tasks such as resizing images, applying transformations, and extracting image features.

6. Model Training and Evaluation: TensorFlow and PyTorch offer APIs for model training and evaluation. You can utilize these APIs to define the CNN architecture, compile the model, train it on the Fashion MNIST dataset, and evaluate its performance using appropriate evaluation metrics.

VIII. CONCLUSION

The conclusion section summarizes the key findings, contributions, and implications of your study. It provides a comprehensive overview of the effectiveness of CNNs in conquering the Fashion MNIST dataset and addresses the research objectives outlined earlier. Additionally, it may discuss potential future directions, such as exploring advanced CNN architectures, transfer learning techniques, or applying the learned knowledge to real-world fashion recognition systems.Overall, this project demonstrates the effectiveness of CNNs and computer vision in conquering the Fashion MNIST dataset and presents a robust system model for fashion item classification. The outcomes contribute to the automation and efficiency of the fashion industry, paving the way for improved processes and customer satisfaction.

REFERENCES

1. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

2. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 1-9).

3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

4. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).

5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 770-778).

6. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.