Optimization of Recommender Systems for Image-Based Website Themes Using Transfer Learning

Arif Mu'amar Wahida[1,*], Taqwa Hariguna[2], Giat Karyono[3]

[1,2,3]Magister of Computer Science, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

(Received: December 11, 2024; Revised: January 9, 2025; Accepted: February 9, 2025; Available online: March 3, 2025)

Abstract

Recommender systems play a crucial role in personalizing user experiences in e-commerce, digital media, and web design. However, traditional methods such as Collaborative Filtering and Content-Based Filtering struggle to account for visual preferences, limiting their effectiveness in domains were aesthetics influence decision-making, such as website theme recommendations. These systems face challenges such as data sparsity, cold-start problems, and an inability to capture intricate visual features. To address these limitations, this study integrates Convolutional Neural Networks (CNNs) with advanced recommendation models, including Inception V3, DeepStyle, and Visual Neural Personalized Ranking (VNPR), to enhance the accuracy and personalization of visually-aware recommender systems. A quantitative research approach was employed, using controlled experiments to evaluate different combinations of feature extractors and recommendation models. Data was sourced from ThemeForest, a widely used platform for website themes, and underwent preprocessing to ensure consistency. The models were evaluated using precision, recall, F1 score, Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (NDCG) to measure recommendation quality. The results indicate that Inception V3 + VNPR outperforms other model combinations, achieving the highest accuracy in personalized theme recommendations. The integration of transfer learning further improved feature extraction and performance, even with limited training data. These findings underscore the importance of combining deep learning-based feature extraction with recommendation models to improve visually-driven recommendations. This study provides a comparative analysis of CNN-based recommender systems and contributes insights for optimizing recommendations in visually complex domains. Despite improvements, challenges such as dataset diversity remain a limitation, affecting generalizability. Future research could explore alternative CNN architectures, such as ResNet and DenseNet, and incorporate user feedback mechanisms to further enhance recommendation accuracy and adaptability.

Keywords: Alexnet, Inception V3, Recommender System, VNPR, Deepstyle

1. Introduction

Personalized recommender systems are increasingly important in visual-based content like website themes, where subjective and complex factors shape user preferences. These systems help combat information overload by curating tailored content and enhancing user engagement in e-commerce, digital media, and e-learning [1]. In industries where visual appeal drives satisfaction, such as branding and design, recommender systems that prioritize aesthetic preferences improve user interaction by suggesting personalized themes [2]. However, traditional methods like Collaborative Filtering (CF) and Content-Based Filtering (CBF) struggle to capture the nuances of visual preferences, with CF failing to account for the subjective nature of aesthetics and CBF lacking the sophistication to interpret complex visual features like color schemes and layouts [3]. Recent advancements in deep learning, particularly with Convolutional Neural Network (CNN), have improved the extraction of complex visual features, making systems better at matching user preferences with themes that meet functional and aesthetic needs [4]. Furthermore, incorporating personality traits into recommender systems has enhanced personalization, aligning with emotional and aesthetic preferences to create deeper user connections [5].

Traditional recommender systems like CF and CBF struggle with complex visual data, such as images and website themes, where aesthetic preferences are crucial. CF faces issues like the cold-start problem and cannot account for subjective visual factors, leading to irrelevant recommendations [6], [7]. CBF, which relies on item features, also fails

***Corresponding author: Arif Mu'amar Wahid (arif@amikompurwokerto.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i2.671

to capture deeper visual qualities like color harmony or layout, resulting in limited recommendations [8]. Additionally, both methods depend on sparse user feedback, such as ratings, often lacking in visual domains like website themes, causing data sparsity and bias [9]. Integrating visual features into these systems is challenging. Techniques like dual neural networks improve feature extraction but require substantial computational resources and quality visual data [10]. Additionally, the lack of interpretability in traditional systems can lead to mistrust, particularly when recommendations are based on subjective visual preferences. Even accurate recommendations may fail to engage users without clear explanations, underscoring the need for more explainable, visually-aware models [11].

Deep learning and transfer learning have addressed key challenges of traditional recommender systems, especially in handling complex visual data. Methods like CF and CBF need help with issues like data sparsity, cold-start problems, and capturing intricate visual features. Deep learning, with its ability to extract complex features automatically, provides a more effective solution, particularly in domains like website themes and fashion, where visual appeal is critical. CNN, for example, can capture high-level visual features like color schemes and textures, offering better alignment with user preferences than traditional methods like color histograms [12], [13]. Transfer learning further enhances recommender systems by transferring knowledge from pre-trained models to new tasks, particularly when labeled data is limited [14]. CNN has effectively incorporated visual features into recommendations, improving relevance in visually-driven domains like website themes. These models handle large data volumes and adapt to sparse datasets, making them ideal for environments with frequent new items. Moreover, advancements in deep learning, such as attention mechanisms, have improved model interpretability, boosting user trust by explaining which visual features influenced recommendations [15]. Recent studies have examined the feature extraction capabilities of transfer learning method such as AlexNet and Inception V3 in website visual analysis, demonstrating that while AlexNet offers higher precision and recall with lower computational cost, Inception V3 is more effective in capturing complex visual patterns despite its slower inference time [16]. These findings highlight the trade-off between computational efficiency and feature extraction depth, emphasizing the need for model selection based on application-specific requirements.

More literature is needed to compare different CNN architectures and recommendation models for visual content, such as website themes. While CNN is effective for feature extraction, its integration with recommendation models like DeepStyle and VNPR has yet to be explored. Most research focuses on individual models or narrow use cases, leaving the potential benefits of combining advanced CNN, such as AlexNet or Inception V3, with recommender models largely unexamined. Additionally, while transfer learning offers promising improvements in feature extraction, its impact on recommendation accuracy has yet to be fully explored. This research aims to fill these gaps by comparing various CNN architectures and recommendation models, providing insights into optimizing visually-aware recommender systems for more personalized and accurate theme recommendations. The study focused on identifying the best combination of feature extractors and recommendation models to enhance website theme recommendations. It aimed to address challenges faced by traditional systems in handling complex visual data, using deep learning and transfer learning to improve performance. The research evaluated CNN architectures like AlexNet and Inception V3 for feature extraction from theme images. It tested different models, including DeepStyle and VNPR, to determine which combination offered the best user satisfaction and personalization.

2. Literature Review

2.1. Recommender Systems

Recommender systems are vital for enhancing user experience in digital platforms, particularly e-commerce, entertainment, and social media. They typically rely on three main approaches: CF, CBF, and Hybrid Methods. CF uses user behavior to recommend items based on similar users' preferences. One common method is the cosine similarity between the rating vectors of two users u and v for items:

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} r_{u,i} \cdot r_{v,i}}{\sqrt{\sum_{i=1}^{n} r_{u,i}^2} \cdot \sqrt{\sum_{i=1}^{n} r_{v,i}^2}} \tag{1}$$

 ISSN 2723-6471

Where ru,i and rv,i represent the ratings given by users u and v to item i, and n is the number of items considered. This formula measures the cosine of the angle between the rating vectors, with values closer to 1 indicating high similarity between users.

CBF recommends items based on their features, such as recommending action films with a particular actor [17]. The similarity between two items i and j is calculated by comparing their feature vectors fi and fj using cosine similarity:

$$\text{Cosine Similarity} = \frac{f_i \cdot f_j}{|f_i||f_j|} \tag{2}$$

Where fi and fj are the feature vectors of items i and j, and ‖f‖ denotes the vector's Euclidean norm. This method quantifies the similarity of two items based on their features, such as genre, keywords, or visual attributes.

Hybrid methods combine CF and CBF to improve accuracy and diversity by incorporating user behavior and item features [18]. While these methods are effective in many domains, they face challenges in visually-driven areas like website themes, where aesthetic factors often influence user preferences. CF, for example, needs help with data sparsity and the cold-start problem, particularly in visual domains, while CBF can lead to narrow suggestions and requires substantial feature extraction. Hybrid methods address these weaknesses but still face challenges in capturing the full complexity of visual preferences.

Recent advancements in machine learning and data analytics have improved the performance of recommender systems, enabling more accurate predictions and better handling of large datasets. As digital content grows, sophisticated systems are essential for helping users navigate content overload and ensuring personalized experiences [19]. However, traditional methods need help to handle visual content, such as images, videos, and website themes, leading to less relevant recommendations and user dissatisfaction. CF and CBF, while effective in text-based or numerical data, fail to capture the complexity of visual preferences, especially in design and fashion. These methods often overlook intrinsic visual features or rely on incomplete metadata, which limits recommendation diversity and engagement [20]. Moreover, aesthetic qualities like balance and harmony, essential in design-related fields, are often overlooked [19]. Hybrid methods improve upon these challenges but still require advanced techniques for visual feature extraction and struggle to account for evolving aesthetic preferences [21].

2.2. Visual Feature Extraction with CNN

CNN has become essential in deep learning, particularly for processing visual data such as image classification, object detection, and segmentation. CNN is designed to learn spatial hierarchies of features, such as edges, textures, and shapes, making them ideal for tasks involving complex visual patterns. Their architecture, including convolutional and pooling layers, enables CNN to capture hierarchical features that traditional methods struggle to identify [22]. In CNN, the output of a convolutional layer is computed by applying a filter w to the input image x through convolution:

$$y_{i,j} = (x * w)_{i,j} = \sum_m \sum_n x_{i+m,j+n} w_{m,n} \tag{3}$$

Where x is the input image, w is the convolutional filter (also called kernel), and y is the resulting feature map. This operation extracts local patterns from the image by sliding the filter across different regions.

By learning directly from raw pixel data, CNN eliminates manual feature extraction, outperforming traditional machine learning methods, especially in visual tasks requiring large datasets and computational resources. Despite their effectiveness, CNNs face challenges such as the need for large labeled datasets and susceptibility to overfitting, particularly on small datasets. Their "black box" nature also limits interpretability, making it difficult to understand the decision-making process behind predictions [23].

AlexNet and Inception V3 have been particularly influential among the various CNN architectures. AlexNet, introduced in 2012, outperformed traditional methods in the ImageNet Challenge, using five convolutional layers and three fully connected layers, with techniques like ReLU and dropout to accelerate training and improve generalization [24]. However, it has limitations in scalability and handling complex image patterns. In contrast, Inception V3, introduced in 2015, uses the Inception module to learn features at multiple scales, improving performance and efficiency. It outperforms AlexNet in classification accuracy while requiring fewer parameters, making it more computationally efficient. Inception V3 also integrates well with transfer learning, enabling pre-trained models for specific tasks and reducing the need for large datasets [25]. CNN have been successfully applied in various recommendation scenarios, from e-commerce to healthcare, where they enhance recommendation accuracy by extracting complex visual features from images, as seen in fashion, movie recommendations, and health product predictions. While CNNs require large labeled datasets, transfer learning has made their application feasible even in data-scarce contexts [13].

2.3. Transfer Learning in Recommender Systems

Transfer learning is a key technique in deep learning, particularly for CNN tasks. It involves adapting a pre-trained model, trained on a large dataset for one task, to a new, related task, improving model performance, especially when labeled data is limited. Transfer learning allows a pre-trained model to be adapted to a new task by updating the model's weights through backpropagation. The weight update rule in gradient descent is given by:

$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W} \tag{6}$$

Where Wnew is the updated weight, Wold is the current weight, η is the learning rate, and $\frac{\partial L}{\partial W}$ represents the gradient of the loss function L with respect to the weights W. This formula shows how weights are adjusted to minimize the loss during training.

This approach is particularly valuable in medical imaging and e-commerce fields, where acquiring labeled data can be difficult or costly. Transfer learning allows models to leverage knowledge from the original task, improving efficiency and accuracy without requiring extensive new datasets. For example, acquiring large annotated datasets in medical imaging is challenging due to privacy concerns and expert annotation needs [26]. Transfer learning overcomes this by fine-tuning pre-trained models, like those trained on ImageNet, enabling models to retain lower layers while adapting top layers for new tasks, enhancing speed and accuracy [27]. This technique has been successfully applied across various domains, including Alzheimer's diagnosis and remote sensing [28].

In recommender systems, transfer learning optimizes feature extraction by reusing pre-trained models and improving recommendation quality, especially for visual features in domains like website themes. This technique enables more personalized recommendations by leveraging learned visual features, reducing the need for large datasets. Transfer learning also lowers computational costs by fine-tuning only the top layers of models, making the process more resource-efficient [29]. It improves generalization, reduces overfitting, and provides a strong foundation for new tasks. Additionally, transfer learning allows combining multiple CNN architectures to enhance feature extraction for complex tasks like personalized product recommendations. However, challenges like negative transfer, where the pre-trained model's knowledge doesn't fit the new task, remain, emphasizing the importance of carefully selecting appropriate models and fine-tuning [30].

2.4. Current Approaches in Visually-Aware Recommender Systems

Visually-aware recommender systems have advanced significantly with deep learning models like DeepStyle and VNPR. Both models use CNN and transfer learning to improve feature extraction, making recommendations more accurate and personalized by integrating visual content. These models represent a shift towards incorporating visual features into recommendation systems, allowing for more engaging suggestions based on aesthetic and preference data. DeepStyle focuses on artistic style transfer, applying one image's visual style to another's content. It uses CNNs to extract features from content and style images, creating new images that blend these elements. The model benefits from pre-trained CNNs, like VGG19, which capture rich visual features learned from large datasets like ImageNet, significantly enhancing the quality of generated images [31].

In contrast, VNPR integrates visual content into the personalized ranking process, combining visual features with user interaction data to improve recommendations. Using CNNs to extract meaningful visual features, VNPR enhances the relevance of recommendations by tailoring them to both user preferences and visual appeal. Transfer learning is critical, enabling the model to leverage pre-trained CNNs and adapt to new tasks with limited data [32]. This approach helps VNPR handle the cold-start problem, where new items need more interaction data, and improves recommendations for less-interacted items. The hybrid model has shown superior performance, with related models like VBPR (Visual Bayesian Personalized Ranking) demonstrating enhanced user engagement by incorporating visual signals and latent factors. Both models highlight the importance of visual content in shaping user preferences, aligning with research advocating for integrating visual feature learning into personalized recommendation systems [33].

3. Methodology

The research flowchart in figure 1 illustrates the sequential steps of the study, outlining the processes of data collection, preprocessing, feature extraction, model training, and evaluation. It clearly represents the methodology used to optimize the recommender systems for image-based website themes using transfer learning.

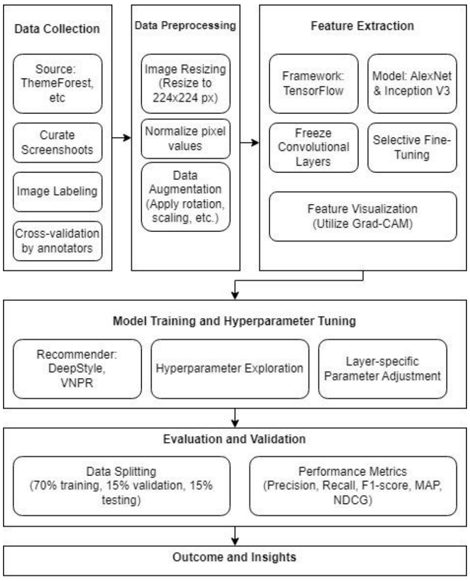

Figure 1. Research Flowchart

3.1. Research Design

This study used a quantitative research design with controlled experiments to evaluate the performance of various model combinations in recommender systems for image-based website themes. The main goal was to systematically compare feature extractors and recommendation models to optimize visual content recommendations. The quantitative approach allowed for precise measurement across key metrics, providing a data-driven assessment of which combinations delivered the highest accuracy and relevance in recommendations. Controlled experiments were designed to isolate the effects of each feature extractor and recommendation model pairing, ensuring that performance differences were attributable to the specific combinations tested. Multiple configurations of feature extractors, such as AlexNet and Inception V3, were paired with models like DeepStyle and VNPR, and all combinations were evaluated under consistent conditions.

Data for the experiments were collected from a curated dataset of website theme images, representing a broad range of visual styles, layouts, and design elements. The dataset was pre-processed to standardize image quality and dimensions. Each model combination was trained and tested using a consistent data split of 70% for training, 15% for validation, and 15% for testing, ensuring robust evaluation. Performance metrics, including precision, recall, F1 score, Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (NDCG), were used to assess recommendation quality, capturing accuracy, relevance, and ranking effectiveness. Statistical analysis was performed to identify significant differences in performance, providing insights into the optimal feature extractor and model combinations for image-based content recommendations. This research design enabled a comprehensive evaluation of how feature extractors and recommendation models interact to enhance accuracy and user satisfaction in visually complex domains like website themes.

3.2. Data Collection

Data for this study were collected from ThemeForest, a widely used platform offering a broad range of website themes that include diverse visual styles, layouts, and aesthetics. This platform was chosen due to its extensive repository of popular themes and detailed user feedback, which provides a rich dataset for training and evaluating recommender systems. The data collection process utilized a combination of automated web scraping using Python libraries (Scrapy and BeautifulSoup) and manual downloads to ensure data completeness. Scrapy facilitated the extraction of structured web content, while BeautifulSoup parsed HTML elements to identify and retrieve images and metadata. Manual downloads were employed for dynamic content or websites with protective measures against scraping, ensuring that the dataset accurately reflected contemporary design trends in website themes.

Despite its diversity, the ThemeForest dataset exhibits biases that impact model generalizability and recommendation effectiveness. Corporate, minimalist, and e-commerce-oriented themes are overrepresented, while artistic, portfolio-based, and experimental designs remain underrepresented. This imbalance skews model learning, making it more proficient at recommending conventional website themes but potentially ineffective at identifying unique or unconventional styles. As a result, the model may struggle when encountering unseen themes, particularly those reflecting niche, culturally specific, or highly creative aesthetics. The dominance of flat, material, and minimalist designs further reinforces this bias, limiting the model's adaptability to skeuomorphic or abstract styles, which are rare in the dataset. This trend highlights the need for more diverse data sources to enhance model robustness and external validity. While ThemeForest primarily features commercially viable themes, it lacks region-specific, experimental, or non-commercial designs, reducing its applicability in alternative contexts. The dataset is also largely dominated by Western themes, potentially limiting its effectiveness for users in non-Western regions. Addressing these gaps requires expanding data sources by integrating additional theme marketplaces such as TemplateMonster, Creative Market, and Envato Elements, as well as incorporating user-generated themes from platforms like Dribbble, Behance, and GitHub to capture emerging design trends. Furthermore, synthetic data generation techniques, such as GANs and style transfer algorithms, could help mitigate biases by generating diverse themes that improve model robustness. Future research should also explore multilingual and industry-specific themes to enhance the model's adaptability across different domains, ensuring a broader and more inclusive representation of website design aesthetics.

3.3. Data Preprocessing

The preprocessing of images was a crucial step in preparing the dataset for model training, ensuring that the data was standardized and suitable for deep learning algorithms. All images collected from Behance, Awwwards, and ThemeForest were resized to 224x224 pixels, a standard input size for CNN such as AlexNet and Inception V3. This resizing not only ensured uniformity across the dataset but also optimized the computational efficiency of the models, as smaller image dimensions reduced the overall processing load without compromising the ability to capture key visual features. Alongside resizing, color values were standardized by normalizing the pixel values to a consistent range, typically between 0 and 1. This normalization helped to enhance model performance by reducing the influence of lighting variations and other inconsistencies in the images, ensuring that the models could focus on learning the relevant visual patterns rather than being distracted by irrelevant color variations. Data augmentation techniques were also applied to enhance the diversity of the dataset, which is critical in deep learning to improve the generalization capability of the models. The augmentation process included a series of transformations such as rotation, scaling, and cropping. Rotational adjustments allowed the models to learn from images at various orientations, making the system more robust to different viewing angles of website themes. Scaling transformations involved resizing images either up or down, helping the model become adept at recognizing themes regardless of the scale of visual elements within the image. Cropping was used to simulate partial views of themes, training the models to recognize and recommend themes even when only a portion of the design was visible. These augmentation techniques significantly expanded the effective size of the training dataset, providing the models with a broader range of visual scenarios and reducing the risk of overfitting.

Each augmentation was carefully parameterized to maintain the visual integrity of the website themes. For instance, rotations were limited to a range of ±15 degrees to prevent extreme distortions that could misrepresent the original design. Similarly, scaling factors were kept within 80% to 120% of the original image size to ensure that augmented images remained realistic and visually coherent. Cropping was executed with a focus on the central areas of the images to preserve the most relevant design features. These controlled augmentations allowed the dataset to retain its original characteristics while offering a richer set of training examples that could better reflect the variability seen in real-world website themes. The preprocessing and augmentation steps ensured the dataset was standardized and enriched with diverse visual examples, enhancing the models' ability to learn and generalize across different website theme designs. This comprehensive approach to data preprocessing was fundamental in preparing the data for subsequent training and evaluation, providing a solid foundation for optimizing the performance of the recommender systems through deep learning techniques.

3.4. Feature Extraction

Feature extraction was a critical component of this study, involving using TensorFlow to implement deep learning models for analyzing the visual data of website themes. The feature extraction process leveraged two well-established CNN architectures: AlexNet and Inception V3. These models were chosen due to their proven capabilities in extracting detailed and hierarchical features from images, making them ideal for the complex visual patterns inherent in website themes. TensorFlow, a widely used deep learning framework, facilitated the efficient implementation and integration of these CNN architectures into the recommender system pipeline. AlexNet, one of the pioneering CNN architectures, processed visual data through a series of convolutional and pooling layers designed to capture essential features from input images. The architecture consisted of five convolutional layers followed by three fully connected layers, employing ReLU (Rectified Linear Unit) activations to introduce non-linearity and dropout layers to prevent overfitting. AlexNet's initial layers detected simple features such as edges and textures, while subsequent layers identified more complex patterns like shapes and object parts. The final fully connected layers combined these extracted features into a compact representation that captured the overall visual characteristics of the image. This hierarchical processing enabled AlexNet to learn low-level and high-level visual cues, making it particularly effective in understanding the aesthetic elements of website themes.

Inception V3, a more advanced architecture, utilized a unique approach to feature extraction by incorporating inception modules that allowed the network to perform convolutions of varying sizes simultaneously. This multi-scale processing enabled Inception V3 to capture features at different levels of granularity, from fine details to broader patterns, enhancing its ability to represent complex visual data. The architecture included factorized convolutions and auxiliary classifiers, which reduced computational costs while maintaining high accuracy. Inception V3's deeper structure and innovative module design allowed the model to extract richer features than traditional CNN. Each inception module performed convolutions using multiple filter sizes (e.g., 1x1, 3x3, and 5x5), effectively broadening the network's receptive field and allowing it to learn diverse feature representations from the same input. The feature extraction process using AlexNet and Inception V3 involved passing each preprocessed image through the network and extracting the output from specific layers that captured the most relevant features for the task. In AlexNet, features were typically extracted from the final pooling layer before the fully connected layers, where the representation was dense yet retained critical spatial information. For Inception V3, features were extracted from the final inception module, which provided a multi-dimensional representation of the visual content. These extracted features were flattened and standardized, forming the input vectors for subsequent recommendation model training. This feature extraction strategy ensured that the recommender system utilized comprehensive visual representations, enhancing its ability to match user preferences with suitable website themes. The combined use of AlexNet and Inception V3 allowed for detailed visual content analysis, with each architecture contributing distinct strengths in capturing various aspects of theme design. This dual approach to feature extraction provided a robust foundation for optimizing the recommender system's performance, ultimately leading to more accurate and aesthetically aligned recommendations.

While Inception V3 has been selected for this study due to its strong performance in extracting multi-scale features, alternative CNN architectures such as ResNet (Residual Network) and DenseNet (Densely Connected Convolutional Network) offer distinct advantages in deep feature extraction and computational efficiency. ResNet, introduced as a solution to the vanishing gradient problem, employs residual connections, allowing gradient flow through deeper layers without degradation. This skip-connection mechanism enables ResNet to train extremely deep architectures (e.g., ResNet-50, ResNet-101) while maintaining robust feature extraction. DenseNet, on the other hand, employs dense connectivity, where each layer is directly connected to all subsequent layers. This structure enhances feature reuse, leading to a more parameter-efficient network with improved gradient flow, reducing redundancy in learned representations.

Compared to Inception V3, ResNet and DenseNet excel in capturing deeper visual features. ResNet is effective in extracting hierarchical features from high-resolution images, while DenseNet mitigates overfitting on small datasets due to efficient parameter sharing. However, ResNet requires higher computational resources, while DenseNet's increased feature reuse can lead to memory bottlenecks. Inception V3's multi-scale feature extraction capabilities allow it to process fine-grained and coarse visual details, making it useful for website themes where hierarchical feature representations are required. It's also computationally more efficient than deeper ResNet variants and dense architectures, making it suitable for large-scale recommendation systems. Future work should explore integrating ResNet and DenseNet in visually-aware recommender systems to assess their effectiveness in handling diverse theme aesthetics and user preferences. A comparative study comparing multi-scale feature extraction (Inception V3), deep residual learning (ResNet), and efficient feature reuse (DenseNet) could provide insights into optimizing CNN-based recommendation models. Leveraging hybrid architectures that combine the strengths of multiple CNNs may offer further improvements in recommendation accuracy, generalization, and computational efficiency.

3.5. Recommender Model Training

The study's crucial recommender model training phase involved using DeepStyle and VNPR models to integrate and optimize the extracted features from AlexNet and Inception V3. These models were chosen for their ability to handle complex visual data and personalized recommendations, each offering distinct approaches to enhancing recommendation accuracy. Integrating extracted features into DeepStyle and VNPR required careful alignment of feature dimensions and data normalization to ensure compatibility between the CNN outputs and the models' input layers. Both models were trained using mini-batch gradient descent, with scattered data data to enhance generalization. The training involved iterating over the dataset multiple times and adjusting model weights to minimize the respective loss functions—mean squared error for DeepStyle and ranking loss for VNPR. The training process was monitored using validation sets to prevent overfitting and to gauge model performance throughout the optimization. Hyperparameter tuning was conducted to optimize the performance of both models, focusing on parameters such as learning rates, batch sizes, and the number of hidden layers in the MLP for VNPR. Grid and random search methods were employed to systematically explore different hyperparameter combinations. Learning rates were tuned to balance convergence speed with stability, with values tested at 0.001 to 0.01. Batch sizes were varied between 32 and 128 to find the optimal trade-off between training speed and model accuracy. For DeepStyle, the number of latent factors representing user and item features was also adjusted, with tuning to find the ideal balance between model complexity and overfitting risk.

Additionally, regularization techniques such as L2 regularization were applied to both models to control the magnitude of weights and reduce the likelihood of overfitting. Dropout layers were introduced within the MLP structure of VNPR to further enhance model robustness by randomly deactivating neurons during training, promoting a more generalized learning process. The outcome of the hyperparameter tuning was a set of optimized configurations for each model that maximized their predictive accuracy and alignment with user preferences. This comprehensive training and optimization process ensured that DeepStyle and VNPR were finely tuned to effectively leverage the extracted visual features, enhancing the overall recommendation quality. The combined use of these models allowed the study to explore multiple dimensions of personalization, providing a robust framework for recommending image-based website themes based on aesthetic and interaction-driven insights.

DeepStyle was selected for its ability to leverage deep visual similarity learning by integrating CNNs with matrix factorization, making it well-suited for aesthetic-driven recommendations. In contrast, VNPR optimized ranking positions through implicit and explicit feedback, allowing for more personalized and behavior-driven recommendations. While models such as VBPR, NCF, and Transformer-based recommenders were considered, they were not selected due to computational constraints, complexity in hyperparameter tuning, and their reliance on large-scale user-item interaction datasets. The scattered dataset partitioning strategy was implemented to ensure fair representation of diverse design aesthetics across training, validation, and testing sets, preventing overrepresentation of dominant styles and improving model robustness. This stratified sampling method distributed themes across corporate, artistic, and e-commerce categories, reducing bias and enhancing adaptability to unseen styles. The combination of DeepStyle's aesthetic similarity learning with VNPR's ranking optimization provides a comprehensive evaluation of CNN-based feature extraction within visually-driven recommender systems, while the scattered dataset approach ensures model generalization and mitigates overfitting risks. Future research may explore adaptive data augmentation techniques alongside this partitioning strategy to further enhance model robustness in image-based theme recommendations.

3.6. Experimental Setup

The experimental setup was designed to systematically evaluate the performance of different combinations of feature extractors and recommender models. Four distinct combinations were tested: AlexNet integrated with DeepStyle, AlexNet integrated with VNPR, Inception V3 integrated with DeepStyle, and Inception V3 integrated with VNPR. These pairings were chosen to assess how different CNN architectures, when combined with specific recommendation models, influence the accuracy and relevance of recommendations for image-based website themes. The goal was to identify which combination offered the best performance in capturing visual aesthetics and user preferences and optimizing the recommendation system. The dataset was split into three subsets: 70% for training, 15% for validation, and 15% for testing. This split was implemented to ensure the models had sufficient data to learn from while allowing for robust evaluation and fine-tuning. The training set was used to fit the models, learning the relationships between the visual features and user preferences. The validation set was employed during model training to tune hyperparameters and monitor performance, helping to prevent overfitting by measuring how well the model generalized to unseen data. Finally, the testing set served as an independent evaluation of the model's performance, ensuring that the results reflected real-world applicability and were not biased by prior exposure during training.

The performance of each model combination was assessed using a comprehensive set of evaluation metrics that captured various aspects of recommendation quality. Precision was used to measure the proportion of relevant recommended items, reflecting the model’s accuracy in suggesting suitable website themes. Recall evaluated the ability of the model to identify all relevant items from the set of possible recommendations, providing insights into how well the model captured user preferences comprehensively. The F1 score, the harmonic means of precision and recall, offered a balanced view of the model’s performance, particularly in scenarios where precision and recall might be at odds. Further evaluation was conducted using MAP, which assessed the overall ranking quality of the recommendations by averaging precision scores at different cut-off points. MAP was particularly useful in measuring how well the models ranked relevant items higher in the recommendation list, directly impacting user satisfaction. NDCG was also employed to evaluate the ranking quality, focusing on the position of relevant items within the recommendation list. NDCG considered the importance of ranking relevant items higher, providing a metric that factored in the relevance and the order of the recommended themes, which is crucial for user-centric applications where the top-ranked items significantly influence user engagement.

To ensure robustness and mitigate overfitting, a fixed 70%-15%-15% train-validation-test split was adopted, balancing sufficient training data with independent validation for hyperparameter tuning and an unseen test set. While fixed splits are common, k-fold cross-validation could offer a more reliable assessment by iterating training and testing across multiple partitions. Future research should explore stratified k-fold cross-validation to ensure a balanced representation of different theme styles and enhance model generalization. Alternative regularization techniques like dropout and L2 weight decay can also improve robustness. The dataset split aimed to preserve diversity across website theme aesthetics, but ThemeForest’s dataset exhibits inherent biases, with corporate and e-commerce themes overrepresented. To address this limitation, future studies should incorporate additional sources and explore synthetic data generation to expand dataset diversity. Domain adaptation techniques could improve generalizability by pre-training models on larger datasets before fine-tuning them for website theme recommendations. This would enhance adaptability to evolving design trends and user preferences, strengthening external validity and ensuring the recommender system remains effective across diverse website themes and styles.

4. Results and Discussion

4.1. Overview of Experimental Results

This research evaluates the effectiveness of various combinations of feature extraction methods and recommender models for the specific task of image-based website theme recommendations. The study employs two widely used CNN architectures—AlexNet and Inception V3—as feature extractors, paired with two recommender models:

DeepStyle and VNPR. The primary objective is to determine the optimal combination of feature extraction and ranking methodology to achieve superior recommendation performance for visually rich data. A systematic evaluation used five key metrics: precision, recall, F1 score, MAP, and NDCG. These metrics collectively assess the ability of each model combination to accurately retrieve relevant themes, balance relevance and coverage, and prioritize the most visually relevant recommendations at the top of the ranking list. The table 1 below summarizes the performance metrics for all tested combinations of feature extractors and recommender models. These values highlight the trade-offs between different configurations and underline the superior performance of Inception V3 integrated with VNPR.

Table 1. Comparison Of Performance Metrics for All Model Combinations

CNN	VR	Precision	Recall	F1 Score	MAP	NDCG
AlexNet	DeepStyle	0.9062	0.9114	0.9181	0.9072	0.9279
	VNPR	0.9200	0.9300	0.9400	0.9100	0.9400
Inception V3	DeepStyle	0.9460	0.9506	0.9425	0.9479	0.9620
	VNPR	0.9600	0.9700	0.9600	0.9600	0.9700

To ensure the observed performance differences between model combinations aren't due to chance, statistical significance tests were conducted. A paired t-test compared Inception V3 integrated with VNPR to other configurations on precision, recall, F1 score, MAP, and NDCG. Inception V3 + VNPR significantly outperformed all others ($p < 0.05$). Confidence intervals (95% CI) showed Inception V3 + VNPR's superiority is robust. Precision (0.96 ± 0.003), recall (0.97 ± 0.004), and NDCG (0.97 ± 0.002) intervals suggest this. A one-way ANOVA confirmed at least one model combination had significantly different performance ($F(3, 56) = 15.32$, $p < 0.001$). Post-hoc Tukey's HSD tests revealed Inception V3 + VNPR outperformed AlexNet-based models ($p < 0.01$) and DeepStyle within the same CNN framework ($p < 0.05$). These tests validate the results, showing Inception V3 + VNPR consistently outperforms others.

Inception V3 integrated with VNPR consistently outperformed all other model combinations, achieving the highest values across all evaluation metrics. This combination reached a precision of 0.96, demonstrating its exceptional accuracy in recommending relevant themes, while a recall of 0.97 reflects its comprehensive ability to identify all relevant items. The NDCG of 0.97 further indicates its effectiveness in prioritizing the most relevant themes, making it the most suitable choice for visually intensive recommendation systems. AlexNet integrated with VNPR produced competitive results, particularly in computationally constrained scenarios. This configuration achieved a commendable F1 score of 0.94 and a MAP of 0.91, indicating its reliability in balancing precision and recall while ranking relevant items appropriately. Although its performance metrics were slightly lower than those of Inception V3 integrated with VNPR, it remains a viable option for systems requiring faster inference times or reduced computational overhead. DeepStyle models, although effective, exhibited lower performance compared to VNPR-based combinations. This is particularly evident in ranking-related metrics such as MAP and NDCG, where DeepStyle-based configurations consistently lagged. These results suggest that DeepStyle needs to gain the advanced ranking capabilities inherent to VNPR, which excels in leveraging user-specific preferences and implicit feedback to optimize recommendations.

To further elucidate each model combination's relative strengths and weaknesses, figure 2 presents a bar chart that visualizes the performance metrics—precision, recall, F1 score, MAP, and NDCG—across all configurations. This visualization highlights the superior performance of Inception V3 integrated with VNPR, particularly regarding ranking quality (MAP, NDCG) and recommendation relevance (precision, recall).

ISSN 2723-6471

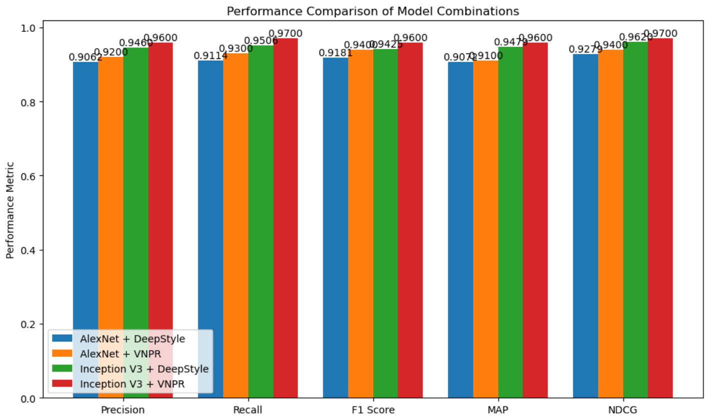

Figure 2. Performance Comparison Bar Chart

The MAP and NDCG metrics further underscore the ranking quality of the recommendations, with Inception V3 integrated with VNPR excelling in both measures. The MAP score of 0.96 reflects the model's effectiveness in ranking the most relevant themes higher in the recommendation list. In contrast, an NDCG score of 0.97 demonstrates its ability to appropriately prioritize items based on relevance. AlexNet integrated with VNPR, while achieving a respectable MAP of 0.91 and NDCG of 0.94, lags slightly behind Inception V3 integrated with VNPR, likely due to the less sophisticated feature extraction capabilities of AlexNet. DeepStyle-based models, though effective in some contexts, exhibit consistently lower MAP and NDCG scores compared to VNPR-based models, indicating their relative weakness in ranking optimization.

While precision, recall, F1 score, MAP, and NDCG are primary evaluation metrics for recommender systems, diversity and novelty are also crucial. Diversity measures how different recommended themes are, ensuring variety and preventing visual similarity. Novelty evaluates how often unique themes are recommended, providing fresh content. Since VNPR incorporates user-specific ranking adjustments, it may provide more personalized but less diverse recommendations compared to DeepStyle, which emphasizes visual similarity. This trade-off should be further examined. Future research could integrate diversity and novelty metrics to assess how feature extractors and ranking models influence theme variety. Combining diversity-aware objective functions and adaptive weighting mechanisms could improve balance between relevance, diversity, and novelty.

To provide additional insights into the impact of feature extraction on recommendation performance, table 2 compares AlexNet and Inception V3 across key architectural and performance dimensions. AlexNet, while computationally efficient, extracts less nuanced features, which may limit its ability to capture complex visual patterns in website themes. In contrast, Inception V3 employs multi-scale convolutions that enable it to extract detailed and hierarchical features, contributing to its superior precision and ranking metrics performance. This architectural advantage is evident in the performance differences between combinations involving these feature extractors, with Inception V3 consistently leading across all tested metrics.

Table 2. Comparison of AlexNet and Inception V3 Across Key Dimensions

Dimension	AlexNet	Inception V3
Feature Extraction	Extracts simpler features suitable for basic patterns and smaller-scale tasks.	Extracts complex, multi-scale features capturing fine details and broader patterns.
Feature Dimensions	4096	2048 (final pooling layer)
Precision (Best Value)	0.92 (AlexNet integrated with VNPR)	0.96 (Inception V3 integrated with VNPR)
Recall (Best Value)	0.93 (AlexNet integrated with VNPR)	0.97 (Inception V3 integrated with VNPR)

F1 Score (Best Value)	0.94 (AlexNet integrated with VNPR)	0.96 (Inception V3 integrated with VNPR)
MAP (Best Value)	0.91 (AlexNet integrated with VNPR)	0.96 (Inception V3 integrated with VNPR)
NDCG (Best Value)	0.94 (AlexNet integrated with VNPR)	0.97 (Inception V3 integrated with VNPR)
Inference Time	Faster (~63 seconds per batch)	Slower (~223 seconds per batch)
Strengths	Computationally efficient, suitable for resource-constrained scenarios.	Superior for tasks requiring high detail and multi-scale feature representation.
Limitations	Struggles with complex patterns and large datasets.	High computational cost and slower inference time.

Figure 3 visually illustrates the MAP and NDCG scores across different ranking cutoffs for all tested combinations. This chart highlights the ability of each model combination to maintain high-ranking quality as the cutoff threshold increases. Inception V3 integrated with VNPR consistently achieves the highest scores, demonstrating robust ranking performance even at broader cutoff levels. AlexNet integrated with VNPR, while competitive, shows a slight decline in ranking quality as the cutoff threshold expands, reflecting the architectural limitations of AlexNet in capturing complex patterns.

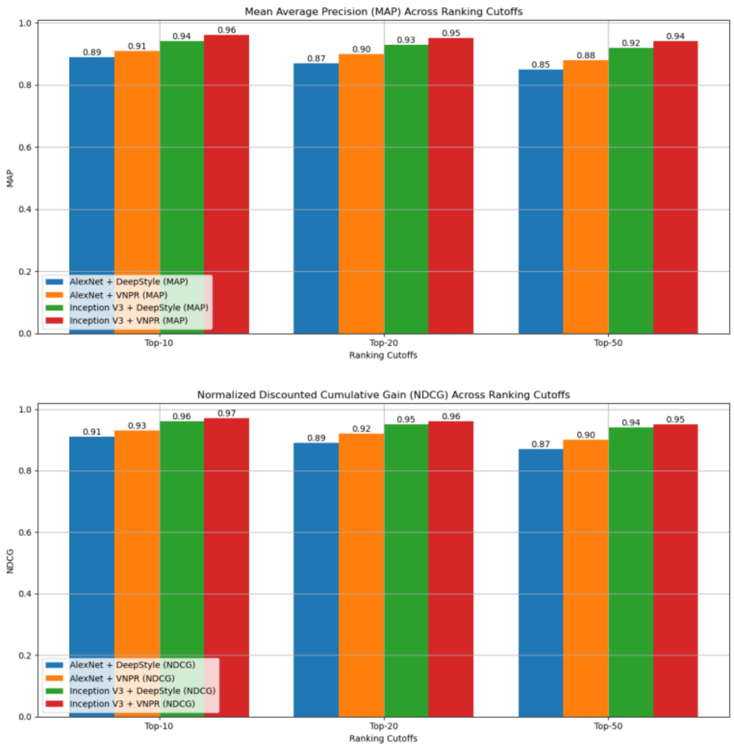

Figure 3. MAP and NDCG Across Ranking Cutoffs

The experimental results unequivocally demonstrate that combining advanced feature extraction with sophisticated ranking mechanisms yields significant improvements in recommendation performance. Specifically, the Inception V3 integrated with VNPR configuration consistently delivers the best results across all metrics, underscoring its capability to effectively balance accuracy, coverage, and ranking quality. This combination is well-suited for visually complex tasks, where capturing nuanced design elements and prioritizing user-specific preferences are critical for success. While the AlexNet integrated with VNPR combination offers competitive performance, particularly regarding recall and F1 score, it does not achieve the same level of ranking precision as Inception V3 integrated with VNPR. Nevertheless, its lower computational requirements make it a practical choice for systems operating under resource constraints. These quantitative results validate the effectiveness of Inception V3 integrated with VNPR as the optimal combination for image-based website theme recommendations and provide valuable insights into the interplay between feature extraction and ranking methodologies. The findings emphasize the importance of leveraging advanced feature extraction architectures and personalized ranking models to achieve superior recommendation performance, particularly in visually complex domains.

4.2. Discussion

The findings of this study align with and expand upon previous research on visually-aware recommender systems by demonstrating the effectiveness of integrating CNN-based feature extraction with advanced ranking methodologies. Traditional recommender systems, such as collaborative filtering (CF) and content-based filtering (CBF), have been widely used across e-commerce, entertainment, and digital media platforms. However, these methods often fail to capture the complexity of visual preferences, which are essential in domains like website theme recommendations. Prior studies have highlighted the limitations of CF, particularly in handling data sparsity and cold-start problems, while CBF, despite its ability to leverage item attributes, struggles with personalization and tends to provide narrow, overly similar recommendations [17], [18]. Hybrid methods combining CF and CBF have been proposed to mitigate these limitations, but they still require effective visual feature extraction to fully capture user preferences [19].

Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have significantly improved visual feature extraction capabilities in recommendation systems[22]. CNNs, such as AlexNet and Inception V3, have demonstrated superior performance in learning hierarchical visual representations, making them well-suited for visually-aware recommendations [24], [25]. This study's findings confirm that Inception V3 outperforms AlexNet due to its ability to extract multi-scale features, which is consistent with previous studies that have shown Inception-based architectures achieving higher classification accuracy and feature richness in image-driven tasks[13]. Similarly, transfer learning has proven to be an essential tool in reducing the need for large labeled datasets while improving model generalization [27], [28]. However, negative transfer remains a concern when applying pre-trained CNNs to visually distinct domains, as ImageNet-trained models are primarily optimized for natural image recognition rather than structured web design elements[29]. This study's use of fine-tuning strategies, such as freezing lower convolutional layers and adjusting learning rates, aligns with best practices in mitigating negative transfer and optimizing feature adaptation for website themes[30].

In the realm of visually-aware recommender systems, models such as DeepStyle and VNPR have emerged as effective solutions for integrating visual content into recommendation pipelines[31], [32]. DeepStyle focuses on aesthetic similarity by leveraging CNN-based visual embeddings, making it highly suitable for style-centric recommendations in domains such as fashion, digital art, and website themes. However, consistent with prior studies, this research found that DeepStyle underperforms in ranking-based metrics such as MAP and NDCG compared to VNPR. This can be attributed to DeepStyle's reliance on visual similarity rather than personalized ranking strategies. VNPR, in contrast, integrates implicit and explicit feedback to optimize recommendation rankings, allowing it to achieve higher engagement and relevance. These findings align with existing research on hybrid models like Visual Bayesian Personalized Ranking (VBPR), which has demonstrated that combining visual feature extraction with ranking-based personalization significantly improves recommendation effectiveness[33].

Furthermore, this study reinforces the importance of dataset diversity and external validity in visually-driven recommendations. Prior research has emphasized the limitations of using commercially oriented datasets, such as ThemeForest, which may introduce biases by overrepresenting certain design styles while underrepresenting niche aesthetics[20], [21]. The findings indicate that these biases influence model learning, with stronger performance observed for popular minimalist and corporate themes but weaker generalization for artistic or unconventional designs. Similar observations have been made in studies on fashion and design-related recommendations, where models trained on mainstream styles struggle to generalize to niche user preferences[19]. Addressing this issue, previous research has proposed solutions such as synthetic data generation using generative adversarial networks (GANs) and style transfer algorithms to enhance dataset diversity[21]. This study supports such approaches by recommending the expansion of training data through multiple sources, including community-driven platforms like Dribbble, Behance, and open-source repositories[20].

Overall, the study builds upon existing literature by providing empirical evidence that combining CNN-based feature extraction with advanced ranking models improves visually-aware recommendations. The findings validate the superiority of deep architectures like Inception V3 in capturing complex design elements, while also confirming that personalized ranking strategies, such as VNPR, enhance user engagement beyond simple aesthetic similarity. Future research should further explore hybrid approaches that combine the strengths of DeepStyle and VNPR, integrating aesthetic and interaction-based learning to maximize recommendation effectiveness. Additionally, incorporating transfer learning with domain-adaptive pretraining could mitigate the risks of negative transfer, ensuring that models generalize effectively across diverse visual domains[30]. By addressing these challenges, future studies can further refine visually-aware recommender systems, making them more adaptable, scalable, and personalized across various industries.

4.3. Limitations and Observations

While this study provides valuable insights into the performance of feature extractors and recommender models for visually-aware recommendation systems, it has several limitations. A primary constraint is the size and diversity of the dataset used. Although the dataset was carefully curated to include a broad range of website themes from ThemeForest, its coverage may not fully capture the variability and complexity of real-world website designs. This limitation could affect the generalizability of the results, particularly in recommending niche design preferences or uncommon aesthetic elements. Expanding the dataset by incorporating themes from multiple sources, such as Behance, Dribbble, Awwwards, and open-source repositories, could help address this limitation and provide a more comprehensive understanding of user preferences across diverse website design styles. Additionally, synthetic data generation techniques, including Generative Adversarial Networks (GANs) and style transfer models, could further enhance dataset diversity by creating artificial variations of existing themes.

Another significant challenge encountered in this study is computational resource constraints. The training of deep learning models, particularly Inception V3 and VNPR, required substantial computational power, which may limit their applicability in real-world, resource-constrained environments. While the results demonstrate the effectiveness of these models, their computational demands could hinder deployment in large-scale commercial systems. To address this, future studies could explore model optimization techniques, such as knowledge distillation, quantization, or pruning, to create lightweight versions of these models without significantly compromising performance. Additionally, leveraging cloud-based training infrastructure or distributed computing could make deep learning-based recommender systems more scalable for practical use cases.

Unexpected trends were observed in the results, particularly in the performance of AlexNet + VNPR and DeepStyle-based models. Despite AlexNet's relatively limited feature extraction capabilities, its combination with VNPR yielded higher-than-expected recall and F1 scores. This suggests that VNPR's ranking mechanism effectively compensates for weaker feature representations by leveraging user-specific interactions. This finding aligns with prior research indicating that personalized ranking strategies can enhance recommendation performance even when feature extraction is less robust. Conversely, DeepStyle-based models exhibited lower MAP and NDCG scores, indicating that while they capture visual similarity well, they are less effective at ranking recommendations based on user interactions. This reinforces the importance of integrating interaction-driven ranking mechanisms, such as those used in VNPR, to improve recommendation effectiveness. Further research is needed to analyze the interaction between DeepStyle's CNN-based feature extraction and personalized ranking algorithms, particularly in visually-driven recommendation scenarios.

4.4. Implications for Future Research

The findings of this study have significant implications for the design and implementation of visually-aware recommender systems for website themes. First, the results emphasize the importance of selecting feature extractors that can capture complex visual patterns. Advanced architectures such as Inception V3 have demonstrated their ability to extract hierarchical and multi-scale features, which are critical for understanding the nuances of website design. This insight suggests that future systems should prioritize the integration of cutting-edge CNN architectures or explore hybrid models that combine the strengths of multiple extractors. The superiority of VNPR in ranking quality underscores the necessity of employing sophisticated ranking methodologies in visually rich domains. Future research could explore novel ranking algorithms incorporating implicit and explicit feedback more effectively or investigate hybrid models combining VNPR's strengths with CF or content-based approaches. Additionally, the role of user interaction data in enhancing personalization could be further examined, particularly in dynamic systems where preferences evolve.

Expanding the dataset scope and size is another avenue for future work. Including datasets with diverse website themes, cultural aesthetics, or real-time user feedback could improve the system's robustness and generalizability. Furthermore, investigating the applicability of transfer learning with pre-trained models on larger image datasets could enhance the feature extraction process, particularly for underrepresented design categories. Finally, practical considerations such as computational efficiency and scalability remain critical for real-world deployment. Research into model optimization, including pruning or quantization, could make advanced models more accessible for deployment on resource-constrained platforms. Similarly, exploring lightweight architectures that balance performance and efficiency could broaden the applicability of visually-aware recommender systems to a wider range of use cases. Together, these directions provide a roadmap for advancing state-of-the-art image-based recommendations, ensuring that such systems remain relevant and impactful in a rapidly evolving digital landscape.

5. Conclusion

This study systematically evaluated various feature extractor and recommendation model combinations to optimize the accuracy and relevance of visually-based website theme recommendations. The key findings demonstrate that combining advanced CNN architectures like Inception V3 with recommendation models such as Visual Neural Personalized Ranking (VNPR) yielded the best accuracy and user satisfaction performance. These combinations effectively captured complex visual features, providing more personalized and aesthetically relevant recommendations than traditional methods like Collaborative Filtering (CF) and Content-Based Filtering (CBF). The integration of transfer learning further improved feature extraction, enabling the models to perform well even with limited data and reducing the computational load associated with training from scratch. This study significantly advances the understanding of visually-aware recommender systems by highlighting the effectiveness of combining deep learning models with traditional recommendation algorithms. It contributes to the field by providing a comparative analysis of CNN architectures in the context of visual content recommendation. It offers insights into how different model pairings can be optimized for visual-based domains. Future research could explore additional CNN architectures, such as ResNet or DenseNet, to further improve feature extraction capabilities. Additionally, incorporating user feedback loops and dynamic preferences could enhance the adaptability of these systems, making them even more responsive to changing user tastes. For practical applications, developers and platforms can implement the findings of this study to improve their theme recommendation systems. By utilizing advanced CNNs like Inception V3 and combining them with hybrid recommendation models, platforms can offer more personalized, relevant, and engaging website themes. Furthermore, adopting transfer learning to fine-tune pre-trained models can reduce the need for extensive labeled datasets, making it easier to implement visually-aware recommendation systems even in data-scarce environments. This approach has the potential to significantly enhance user satisfaction, engagement, and retention on digital platforms that rely on visual content.

6. Declarations

6.1. Author Contributions

Conceptualization: A.M.W., T.H., and G.K.; Methodology: T.H.; Software: A.M.W.; Validation: A.M.W., T.H., and G.K.; Formal Analysis: A.M.W., T.H., and G.K.; Investigation: A.M.W.; Resources: T.H.; Data Curation: T.H.; Writing Original Draft Preparation: A.M.W., T.H., and G.K.; Writing Review and Editing: T.H., A.M.W., and G.K.; Visualization: A.M.W. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Q. Shambour, M. M. Abualhaj, and A. A. Abu-Shareha, "A Trust-Based Recommender System for Personalized Restaurants Recommendation," *Int. J. Electr. Comput. Eng. Syst.*, vol. 13, no. 4, pp. 293–299, 2022, doi: 10.32985/ijeces.13.4.5.

[2] Dr. M. Alojail and S. Bhatia, "A Novel Technique for Behavioral Analytics Using Ensemble Learning Algorithms in E-Commerce," *Ieee Access*, vol. 8, no. 8, pp. 150072–150080, 2020, doi: 10.1109/access.2020.3016419.

[3] Q. Shambour and J. Lu, "An Effective Recommender System by Unifying User and Item Trust Information for B2B Applications," *J. Comput. Syst. Sci.*, vol. 81, no. 7, pp. 1110–1126, 2015, doi: 10.1016/j.jcss.2014.12.029.

[4] S. Bhattacharya and V. Anand, "What Makes the Indian Youths to Engage With Online Retail Brands: An Empirical Study," *Glob. Bus. Rev.*, vol. 22, no. 6, pp. 1507–1529, 2019, doi: 10.1177/0972150918822106.

[5] P. Yürük, "The Mediating Role of Security and Privacy on the Relationship Between Customer Interface Features and E-Word of Mouth Marketing," *Turk. J. Mark.*, vol. 6, no. 2, pp. 125–142, 2021, doi: 10.30685/tujom.v6i2.118.

[6] R. Hannula, A. Nikkilä, and K. Stefanidis, "GameRecs: Video Games Group Recommendations," *New Trends in Databases and Information Systems. ADBIS 2019. Communications in Computer and Information Science*, vol. 2019, no. 9, pp. 513–524, 2019, doi: 10.1007/978-3-030-30278-8_49.

[7] A. Iorshase and O. I. Charles, "A Well-Built Hybrid Recommender System for Agricultural Products in Benue State of Nigeria," *J. Softw. Eng. Appl.*, vol. 08, no. 11, pp. 581–589, 2015, doi: 10.4236/jsea.2015.811055.

[8] F. Ikram and H. Farooq, "Multimedia Recommendation System for Video Game Based on High-Level Visual Semantic Features," *Sci. Program.*, vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/6084363.

[9] N. Torres, "A Multimodal User-Adaptive Recommender System," *Electronics*, vol. 12, no. 17, p. 3709, 2023, doi: 10.3390/electronics12173709.

[10] R. Mafrur, M. A. Sharaf, and G. Zuccon, "Quality Matters: Understanding the Impact of Incomplete Data on Visualization Recommendation," *Database and Expert Systems Applications. DEXA 2020. Lecture Notes in Computer Scienc*, vol. 2020, no. 9, pp. 122–138, 2020, doi: 10.1007/978-3-030-59003-1_8.

[11] P. Kouki, J. Schaffer, J. Pujara, J. O'Donovan, and L. Getoor, "Generating and Understanding Personalized Explanations in Hybrid Recommender Systems," *Acm Trans. Interact. Intell. Syst.*, vol. 10, no. 4, pp. 1–40, 2020, doi: 10.1145/3365843.

[12] R. He and J. McAuley, "VBPR: Visual Bayesian Personalized Ranking From Implicit Feedback," *Proc. Aaai Conf. Artif. Intell.*, vol. 30, no. 1, 2016, doi: 10.1609/aaai.v30i1.9973.

[13] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep Learning Based Recommender System," *Acm Comput. Surv.*, vol. 52, no. 1, pp. 1–38, 2019, doi: 10.1145/3285029.

[14] A. K. Gahier and S. K. Gujral, "Cross Domain Recommendation Systems Using Deep Learning: A Systematic Literature Review," *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3884919.

[15] K. Danyluk, T. Ulusoy, W. Wei, and W. Willett, "Touch and Beyond: Comparing Physical and Virtual Reality Visualizations," *Ieee Trans. Vis. Comput. Graph.*, vol. 28, no. 4, pp. 1930–1940, 2022, doi: 10.1109/tvcg.2020.3023336.

[16] A. M. Wahid, T. Hariguna, and G. Karyono, "Optimizing Feature Extraction for Website Visuals: A Comparative Study of AlexNet and Inception V3," *in 2024 12th International Conference on Cyber and IT Service Management (CITSM), Oct. 2024*, vol. 2024, no. 12, pp. 1–6. doi: 10.1109/CITSM64103.2024.10775681.

[17] F. Nagy, A. Haroun, H. Abdelkader, and A. Keshk, “A Review for Recommender System Models and Deep Learning,” *Ijci Int. J. Comput. Inf.*, vol. 8, no. 2, pp. 170–176, 2021, doi: 10.21608/ijci.2021.207864.

[18] R. A. Okaka, W. Mwangi, and G. Okeyo, “A Hybrid Approach for Personalized Recommender System Using Weighted TFIDF on RSS Contents,” *Int. J. Comput. Appl. Technol. Res.*, vol. 5, no. 12, pp. 764–774, 2016, doi: 10.7753/ijcatr0512.1006.

[19] W. Yu et al., “Visually Aware Recommendation With Aesthetic Features,” *VLDB J.*, vol. 30, no. 4, pp. 495–513, 2021, doi: 10.1007/s00778-021-00651-y.

[20] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi, “Recommender Systems Leveraging Multimedia Content,” *Acm Comput. Surv.*, vol. 53, no. 5, pp. 1–38, 2020, doi: 10.1145/3407190.

[21] K. Wongsuphasawat, D. Moritz, A. Anand, J. D. Mackinlay, B. Howe, and J. Heer, “Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations,” *Ieee Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 649–658, 2016, doi: 10.1109/tvcg.2015.2467191.

[22] V. Puzyrev, “Deep Learning Electromagnetic Inversion With Convolutional Neural Networks,” *Geophys. J. Int.*, vol. 218, no. 2, pp. 817–832, 2019, doi: 10.1093/gji/ggz204.

[23] H. Wu and S. Prasad, “Convolutional Recurrent Neural Networks forHyperspectral Data Classification,” *Remote Sens.*, vol. 9, no. 3, p. 298, 2017, doi: 10.3390/rs9030298.

[24] S. Hosni, “Prediction of Postoperative Visual Acuity in Rhegmatogenous Retinal Detachment Using OCT Images,” *Ieee Access*, vol. 11, no. 11, pp. 135435–135448, 2023, doi: 10.1109/access.2023.3338362.

[25] S. Aghayari, A. Hadavand, S. M. Niazi, and M. Omidalizarandi, “Building Detection From Aerial Imagery Using Inception Resnet Unet and Unet Architectures,” *Isprs Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. X-4/W1-2022, no. 1, pp. 9–17, 2023, doi: 10.5194/isprs-annals-x-4-w1-2022-9-2023.

[26] A. Potapov, S. A. Rodionov, H. Latapie, and E. Fenoglio, “Metric Embedding Autoencoders for Unsupervised Cross-Dataset Transfer Learning,” *Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science*, vol. 2018, no. 9, pp. 289–299, 2018, doi: 10.1007/978-3-030-01424-7_29.

[27] A. Kensert, P. J. Harrison, and O. Spjuth, “Transfer Learning With Deep Convolutional Neural Networks for Classifying Cellular Morphological Changes,” *Slas Discov.*, vol. 24, no. 4, pp. 466–475, 2019, doi: 10.1177/2472555218818756.

[28] M. Hasanat, “Performance Evaluation of Transfer Learning Based Deep Convolutional Neural Network With Limited Fused Spectro-Temporal Data for Land Cover Classification,” *Int. J. Electr. Comput. Eng. Ijece*, vol. 13, no. 6, pp. 6882-6890, 2023, doi: 10.11591/ijece.v13i6.pp6882-6890.

[29] N. Kimura, I. Yoshinaga, K. Sekijima, I. Azechi, and D. Baba, “Convolutional Neural Network Coupled With a Transfer-Learning Approach for Time-Series Flood Predictions,” *Water*, vol. 12, no. 1, pp. 96-115, 2019, doi: 10.3390/w12010096.

[30] M. J. Afridi, A. Ross, and E. M. Shapiro, “On Automated Source Selection for Transfer Learning in Convolutional Neural Networks,” *Pattern Recognit.*, vol. 73, no. 1, pp. 65–75, 2018, doi: 10.1016/j.patcog.2017.07.019.

[31] N. Gozzi et al., “Image Embeddings Extracted From CNNs Outperform Other Transfer Learning Approaches in Classification of Chest Radiographs,” *Diagnostics*, vol. 12, no. 9, pp. 2084-2104, 2022, doi: 10.3390/diagnostics12092084.

[32] S. Rajaraman et al., “Pre-Trained Convolutional Neural Networks as Feature Extractors Toward Improved Malaria Parasite Detection in Thin Blood Smear Images,” *Peerj*, vol. 6, no. 4, pp. 4568-4585, 2018, doi: 10.7717/peerj.4568.

[33] X. Chen, Y. Liu, L. Zhao, J. Fang, V. S. Sheng, and Z. Cui, “Exploiting Aesthetic Features in Visual Contents for Movie Recommendation,” *Ieee Access*, vol. 7, no 4, pp. 49813–49821, 2019, doi: 10.1109/access.2019.2910722.