Dual-Branch Convolutional Neural Network for Image Comparison in Presentation Style Coherence

Maria Vlahova-Takova
Technical University of Sofia, Sofia, Bulgaria
mariavlahova.92@gmail.com (corresponding author)

Milena Lazarova
Technical University of Sofia, Sofia, Bulgaria
milaz@tu-sofia.bg

Received: 10 November 2024 | Revised: 29 December 2024 and 29 December 2024 | Accepted: 4 January 2025

 DOI: https://doi.org/10.48084/etasr.9571

ABSTRACT

Image comparison is an important task that is part of the pipeline in many different computer vision applications. Maintaining style coherence across presentation slides is essential for professionalism and effective communication. Inconsistent design elements, such as varying fonts, colors, borders, and logo placements, can disrupt the visual flow and diminish the overall impact. This study introduces a novel approach to automate the validation of presentation slide coherence using a Dual-Branch Convolutional Neural Network. The model is trained to calculate a similarity score between image slides based on key design parameters, including font consistency, color schemes, border styles, and layout alignment. The proposed CNN architecture is specifically designed to compare two inputs representing slide images for binary classification. Unlike traditional Siamese networks that rely on identical branches and a distance metric for feature comparison, the proposed dual-branch architecture concatenates feature embeddings from two specialized branches and processes them through fully connected layers for final classification, allowing more targeted and nuanced feature extraction and coherence evaluation. The model was evaluated on a custom image dataset comprising 6000 images synthesized following specific design guidelines for style coherence of image features to ensure consistency and variety in the dataset while maintaining a balance for comparative tasks. The experimental results demonstrate significant improvements over the baseline Siamese network across all key metrics. Specifically, the proposed model achieved an accuracy of 0.85 compared to 0.81 for the baseline Siamese network, Jaccard similarity 0.76 vs 0.72, Kappa coefficient 0.69 vs 0.62, and ROC AUC 0.87 vs 0.81. Additionally, precision increased from 0.73 to 0.77 and the F1-score reached 0.87, reflecting a stronger balance between precision and recall. This work provides a significant contribution to automated design evaluation, offering a flexible and modular architecture that supports multi-view analysis and captures intricate visual patterns and discrepancies. By addressing key limitations of traditional approaches, the proposed model provides a robust tool to ensure style coherence in professional presentations, paving the way for more efficient and accurate design validation processes.

Keywords-image similarity; presentation advisor; image processing; presentation coherence; neural networks

I. INTRODUCTION

In today's digital world, keeping a consistent design is essential for ensuring professionalism and clarity, especially in business presentations, websites, and marketing materials. When design elements such as fonts, colors, borders, or logos do not match across slides, the result can be confusing and unprofessional, potentially impacting the audience's understanding and engagement. This is especially relevant in business settings, where clarity in communication is crucial.

Image coherence is fundamental in a presentation advisor tool as it ensures visual consistency and alignment across all slides in a presentation. A cohesive design enhances the overall professional appearance, allowing audiences to focus on content without being distracted by style variations. Consistency in fonts, colors, and layouts reinforces the message and creates a smooth flow, which is essential for an effective presentation. For a teaching tool, demonstrating these principles is crucial to help users create informative and visually appealing presentations, ultimately improving the communication of ideas.

Manual consistency checks for each slide can be time-consuming and labor-intensive. However, recent advances in Artificial Intelligence (AI) can automate this process. AI can analyze slides to catch subtle design discrepancies and calculate a similarity score based on elements such as fonts, colors, and layout. By generating a score from 0 to 1, where a higher value reflects stronger design coherence, AI helps ensure that presentations remain visually aligned with branding and other guidelines. Previous research has underscored the need for such tools, as design inconsistencies in presentations often reduce effectiveness and professionalism across sectors from academia to business [1].

Style coherence can be interpreted differently by designers. Elements such as gradients, transparency, and textures are subjective, meaning that what appears consistent to one designer may not be the same for another. Achieving a cohesive style requires managing multiple parameters, including colors, fonts, borders, logos, spacing, and alignment, all of which must be evaluated consistently. Therefore, the following elements should be taken into account:

- Color analysis: Techniques like histograms or metrics, such as color distance in RGB/HSV space, are typically used to detect similarities in color schemes [2, 3].
- Font analysis: Consistency in font types, sizes, and styles is usually checked using Optical Character Recognition (OCR) [4].
- Border and shape consistency: A key aspect of style is maintaining consistent border thickness, patterns, or shapes across designs to ensure visual coherence [5].
- Logo detection and placement: For a cohesive style, logos must be consistently sized and positioned, helping to maintain brand uniformity [6].
- Layout and alignment: Consistency requires analyzing the arrangement and alignment of elements, ensuring the overall layout is balanced and visually harmonious [7, 8].

The paper introduces a novel approach that utilizes a dual-branch CNN specifically designed for image coherence assessment by innovatively targeting the key elements of visual design, namely color schemes, fonts, borders, and layouts, to quickly detect and flag inconsistencies, distinguishing from standard time-intensive manual checks. In contrast to the traditional usage of Siamese CNNs with two identical branches that utilize a distance metric to compare embeddings and determine similarity between the inputs, the proposed dual-branch CNN concatenates the feature embeddings from each branch and passes them through fully connected layers for final classification, allowing for more specialized feature extraction. Thus, the proposed CNN architecture enables parallel feature extraction while preserving effective feature integration that allows it to identify even subtle differences that contribute to the overall style coherence. The modularity and flexibility of the architecture further support multi-view analysis. In addition, the incorporation of dedicated layers for regularization within the proposed architecture mitigates overfitting, enhances model generalizability, and fosters robust learning, even when working with limited datasets.

II. RELATED WORKS

The feature detection and similarity metrics used in image comparison for evaluation of image similarity include Oriented Fast and Rotated BRIEF (ORB), Earth Mover's Distance (EMD), Structural Similarity Index (SSIM), and Pixel Similarity. ORB is a fast and efficient algorithm for detecting and describing image key points [9]. It uses a corner detection based on Features from Accelerated Segment Test (FAST) to find important points in the images and BRIEF as a feature descriptor to represent the area around each key point with a compact binary string. EMD measures the similarity between two images by calculating the minimum "work" required to transform one image's distribution into another, often comparing color histograms [10]. It is effective in handling distribution shifts, making it robust for comparing images based on color or texture. However, EMD is computationally expensive, particularly for large images. It is commonly used in color-based image retrieval and texture comparison tasks. The SSIM compares two images based on structural features, luminance, and contrast, reflecting how humans perceive image quality [11]. Rather than focusing on pixel-level differences, it provides a more balanced comparison. However, SSIM is less effective when images experience geometric transformations, such as rotation or scaling. It is widely used in tasks such as image quality assessment, video compression analysis, and image restoration. Pixel similarity compares two images by directly matching corresponding pixel values, using methods such as Mean Squared Error (MSE) or normalized cross-correlation [12]. It is simple and fast, making it easy to compute. However, it is sensitive to noise and slight shifts and does not capture image structure or patterns. This method is best suited for exact image-matching tasks, such as in medical imaging or surveillance, where minimal differences are expected.

With recent advances in deep learning architectures, deep-ranking approaches are also widely used for image similarity evaluation. In [13], deep ranking was presented for fine-grained image similarity, utilizing a deep learning model based on a triplet-based CNN architecture. This model learns to rank the similarity of images by processing a triplet consisting of a query image, a positive image, and a negative image. This approach effectively captures subtle distinctions between similar images and employs a multiscale method to analyze features at various resolutions. However, it requires large amounts of labeled training data and significant computational resources for training. The model is particularly useful in applications like search-by-example, where identifying fine-grained differences is essential, such as distinguishing between different species of animals or plant varieties.

CLIP (Contrastive Language-Image Pretraining) is a multimodal similarity measure that evaluates images based on their semantic content utilizing both text and image encoders [14]. It transforms images and their corresponding textual descriptions into a shared vector space and measures similarity using cosine similarity. This approach provides a deeper semantic understanding, surpassing traditional pixel or structural comparisons by incorporating semantic relationships between images. However, CLIP's effectiveness depends on the availability of high-quality textual descriptions. It is particularly well suited for tasks such as semantic image search, image classification, and recommendation systems that integrate visual and textual information.

OASIS (Online Algorithm for Image Similarity) is an online learning algorithm specifically designed for scalable image similarity learning, employing a large-margin criterion and hinge loss [5]. It excels at handling large datasets, capable of learning from millions of images while utilizing sparse matrices for efficient computation and storage. However, OASIS may struggle with cases requiring dense representations or very fine-grained distinctions between images. Its primary applications include image retrieval and visual search tasks in large-scale scenarios. This makes OASIS particularly effective for environments where rapid and scalable image comparison is essential. In [15], an autoencoder based on a DNN was used to identify and improve image similarity tasks. This approach operates in an unsupervised learning setup, where the model autonomously learns image features. The autoencoder comprises an encoder that compresses input images to latent representations and a decoder that reconstructs images from them. The cosine similarity measure is employed on latent vectors to quantify the similarity between images. Principal Component Analysis (PCA) also reduces dimensionality, improving similarity computations. Dropout, batch normalization, and data augmentation are integrated to address overfitting, enhancing generalization. During training, hyperparameter tuning and network optimization increased performance. The results showed that this autoencoder framework surpassed traditional CNNs in similarity assessment.

Siamese neural networks, comprising two identical feedforward NNs, are also widely used for similarity evaluation and vector comparison in many practical applications in computer science and statistics [16]. In [17], a Siamese network architecture called sHybridNet was discussed for image matching, aiming to enhance image similarity assessment across large datasets using deep learning techniques. This model consists of two identical branches that process two images independently utilizing a contrastive loss function to minimize the distance between similar images while maximizing it for dissimilar ones. Built on a hybrid CNN architecture, it outputs feature vectors from a specific layer after processing images through multiple convolutional layers. The method outperformed traditional techniques in image retrieval and showed strong generalization on unseen data, highlighting its effectiveness in challenging conditions such as occlusions and varying lighting. Thus, sHybridNet is a valuable tool for image matching with the potential for further improvement through larger, accurately labeled datasets.

III. SYSTEM ARCHITECTURE AND DESIGN METHODOLOGY

A. *Image Similarity Model Training and Evaluation Pipeline*

The method comprises the following stages:

- Dataset generation: At this stage, a synthetic dataset was created containing images with various attributes.
- Image preprocessing: All images in the dataset were resized to a uniform size, and the pixel values were normalized for consistent input data.
- Dataset split: the dataset was split into a training set (60%), a validation set (20%), and a test set (20%).
- Training the dual-branch model: The proposed dual-branch model was trained on the training set, while the validation set was used to monitor and adjust the model.
- Evaluation: The dual-branch model was evaluated using the test dataset based on several performance metrics.

The similarity score provided by the trained model, ranging from 0 to 1, provides a quantitative measure of style coherence, where 1 indicates full consistency. The developed similarity assessment model supports the automated validation of presentations by providing an objective criterion for style consistency. Through the evaluation result generated by the model, users can quickly identify whether two slides follow the same style and whether they are visually and thematically connected. This enables easy detection of inconsistencies and optimization of the design, leading to more consistent and effective presentations.

B. *Image Dataset*

A custom synthetic image dataset was created following specific design guidelines to ensure style coherence. These guidelines consider multiple visual attributes, including background color, font type, text color, logos, and borders, to generate a wide range of stylistic combinations. The dataset consists of 6000 images, each carefully designed to reflect unique stylistic elements while maintaining overall consistency and providing a rich foundation for comparative style analysis tasks.

Each image in the dataset features a distinct background color and includes randomly generated text ranging from 5 to 50 characters in length. Depending on the chosen configuration, an image can contain embedded images, logos, and borders, with up to three text areas placed dynamically within it. This variety ensures a comprehensive representation of potential styles. To maintain balance, the dataset was evenly divided between pairs of images that share the same style and those that differ. Two images are considered identical in style if they use the same font type, border presence and style, background color, logo design, and logo placement (in cases where both images contain logos).

The image generation process was based on a diverse selection of design elements: nine font types, six background colors, six text colors, three logos, five additional images, and four logo positions (top left, top right, bottom left, and bottom right). Each image was standardized to a size of 224×224 pixels, while logos - when present - are consistently resized to 10×10 pixels for visual harmony. Borders can be optionally added and customized by color and width to provide even greater stylistic variety. Figure 1 shows some examples of the generated images.

Fig. 1. Sample images from the generated dataset.

C. Image Preprocessing

The image preprocessing pipeline aims to standardize the images in the dataset and prepare them for input into the model. The process begins by loading the image from a given path. Once the image is loaded, it is resized to a consistent size of 224×224 pixels. This ensures uniformity in the dataset, which is crucial for image processing tasks, particularly when feeding images into machine learning models that require a fixed input size. After resizing, the pixel values are normalized, converting them from the range [0, 255] to the range [0, 1]. The normalization process aims to improve the model performance by ensuring that the pixel values are on a similar scale, which can make model training faster and more accurate.

D. Model Architecture

To solve the image coherence problem, a dual-branch CNN was designed to compare two inputs that represent slide images for binary classification. Figure 2 and Table I describe the Dual-Branch CNN architecture.

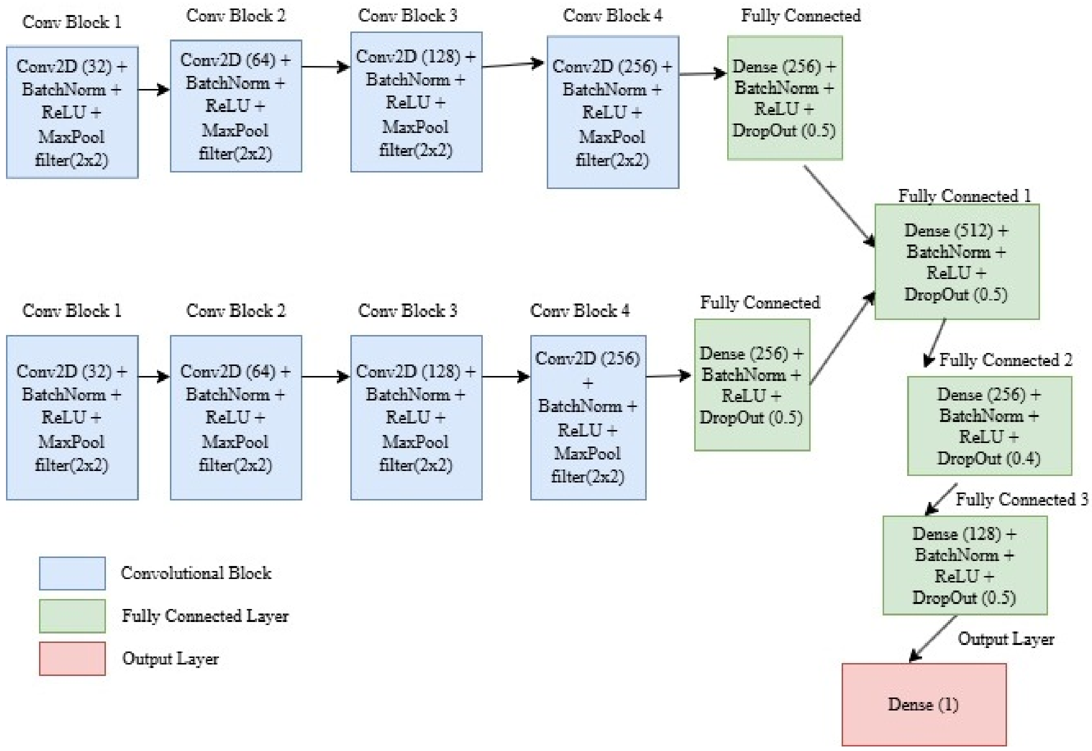

Fig. 2. Dual-branch CNN architecture for image similarity.

TABLE I. DUAL-BRANCH CNN ARCHITECTURE DETAILS

Layer	Components of a Convolutional Neural Network					
	Feature Map	**Size**	**Kernel size**	**Stride**	**Activation**	**Additional Info**
Input	Image	32×32	–	–	–	–
Conv Block 1	32	28×28	3x3	1	ReLU	BatchNorm, MaxPool 2x2
Conv Block 2	64	14v14	3x3	1	ReLU	BatchNorm, MaxPool 2x2
Conv Block 3	128	7×7	3x3	1	ReLU	BatchNorm, MaxPool 2x2
Conv Block 4	256	3x3	3x3	1	ReLU	BatchNorm, MaxPool 2x2
Fully Connected	256	–	–	–	ReLU	Dense layer, BatchNorm, Dropout (0.5)
Fully Connected 1	512	–	–	–	ReLU	Dense layer, BatchNorm, Dropout (0.5)
Fully Connected 2	256	–	–	–	ReLU	Dense layer, BatchNorm, Dropout (0.4)
Fully Connected 3	128	–	–	–	ReLU	Dense layer, BatchNorm, Dropout (0.5)
Output	1	–	–	–	Sigmoid	Dense layer

The traditional Siamese network with identical branches uses a distance metric to compare embeddings after processing to determine the similarity between inputs. In contrast, the proposed dual-branch model concatenates the feature embeddings from separate branches and passes them through fully connected layers for final classification, allowing for more specialized feature extraction. Each of the two parallel branches processes input through four convolutional blocks which extract features using Conv2D layers with 32, 64, 128, and 256 filters, combined with BatchNorm, ReLU activation, and MaxPooling. After feature extraction, each branch outputs to a fully connected layer with 256 units, BatchNorm, and Dropout for regularization. The outputs from both branches are then concatenated, merging their feature representations. The combined features are further processed through two fully connected layers with 512, 256, and 128 units, each incorporating BatchNorm and Dropout. Finally, a single output layer produces a value for binary classification.

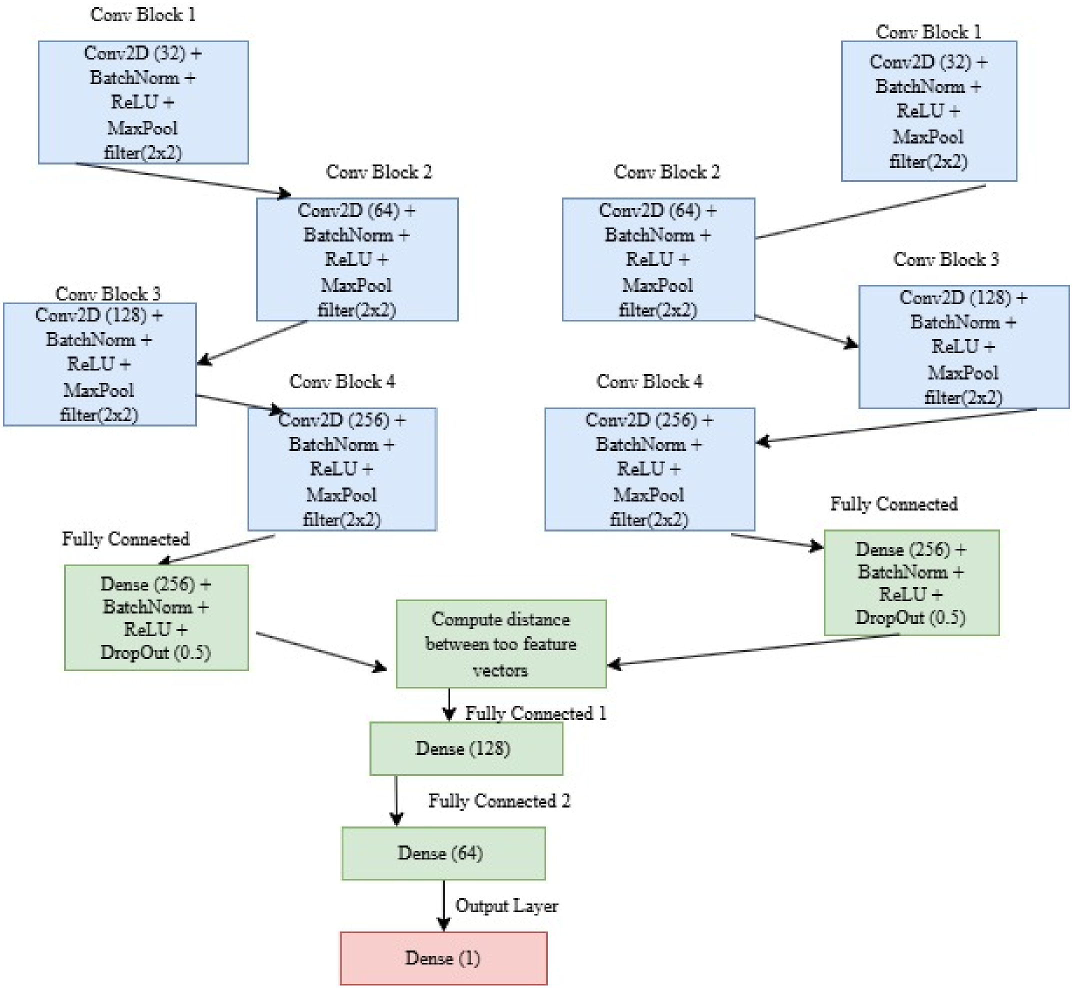

Fig. 3. Siamese architecture for image similarity.

E. Model Training

The proposed model was implemented in Python and TensorFlow and trained on the generated dataset. The dataset was split into training and test datasets using an 80% to 20% ratio, thus the training dataset comprised 4800 images and the test dataset had 1200 images. The model training lasted 100 epochs and used 5-fold cross-validation for hyperparameter tuning, with a batch size of 16 and a learning rate set to 3×10^{-4}. Early stopping was also used as regularization to prevent overfitting. The model training was performed on the following experimental platform: Intel Core i7-8700K CPU @ 3.70GHz, 64 GB RAM.

IV. RESULTS AND DISCUSSION

A. Evaluation Metrics and Results

The evaluation of the model was based on a comparison of the proposed architecture with a Siamese network as a baseline. To evaluate the effectiveness of the suggested model, a range of metrics were calculated that capture various dimensions of image similarity and classification accuracy. Table II presents the results obtained, detailing performance indicators such as accuracy, precision, recall, F1-score, Jaccard similarity, Kappa coefficients, and ROC AUC [18]. Figure 4 presents the model's accuracy and loss for both the training and the validation data, for each of the five folds used in the cross-validation of model training.

TABLE II. EVALUATION METRICS

Metric Name	**Metric Value**		
	Expected values for moderate-stakes applications	**Siamese network**	**Suggested Dual-Branch CNN**
Dataset size	5 000 – 10 000	6000	6000
Accuracy	Above 0.9	0.81	0.85
Precision	Above 0.7	0.73	0.77
Recall	Above 0.9	1	0.99
F1-Score	0.7-0.8	0.84	0.87
Jaccard similarity	0.5-0.8	0.72	0.76
Kappa coefficient	0.4-0.6	0.62	0.69
ROC AUC	0.8-0.9	0.81	0.87
Training time		12 min	20 min

B. Performance Analysis

The proposed Dual-Branch CNN model is compared against a traditional Siamese network. The baseline model featured a shared CNN feature extractor, which extracted deep visual features from both images using four convolutional blocks with convolutional layers, batch normalization, max-pooling and global average pooling, followed by a fully connected layer with dropout for feature refinement. Both input images were identically processed and the corresponding feature vectors were outputted. These feature vectors were then compared using absolute difference, enabling the model to focus on variations between the two slides. The computed difference was subsequently passed through fully connected layers refining the feature representation before generating the final output. To ensure a fair comparison, both models followed the same training process. The base model design was influenced by [19], which explored Siamese Multi-Task CNNs for learning preference-based similarities from face images. However, in this baseline model, some modifications were introduced to maintain a consistent feature extraction structure between the base and proposed models. Figure 3 illustrates the full architecture of the baseline Siamese network model used in this study.

The proposed Dual-Branch CNN model outperformed the baseline Siamese model across several key metrics including accuracy (0.85 vs 0.81), precision (0.77 vs 0.73), recall (0.99 vs 1), F1-score (0.87 vs 0.84), Jaccard similarity (0.76 vs 0.72), Kappa coefficient (0.69 vs 0.62), and ROC AUC (0.87 vs 0.81), reflecting its improved ability to capture relevant patterns and produce more accurate predictions.

Improved precision suggests fewer false positives and a more reliable identification process, while the increase in F1-score indicates a better balance between precision and recall. Some additional metrics also reflect the superior performance of the proposed dual-branch CNN compared to the Siamese model. The Jaccard similarity and the Kappa coefficient improvements highlight its ability to provide consistent predictions. Moreover, ROC-AUC increased from 0.81 to 0.87 emphasizing the model's capacity to distinguish between classes. Although training time was slightly increased, the stronger performance justifies the trade-off, making the dual-branch CNN a better choice for moderate-stakes applications.

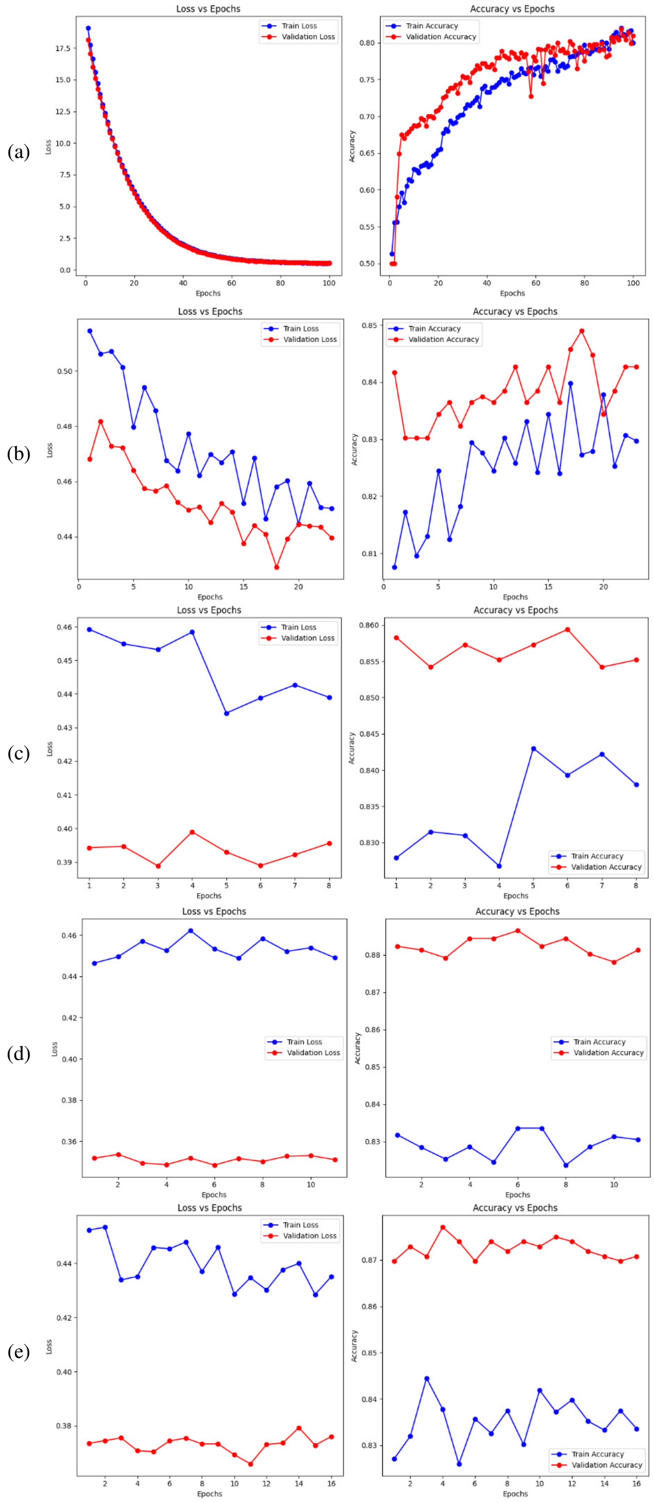

Fig. 4. Model's accuracy and loss during training: (a) fold 1, (b) fold 2, (c) fold 3, (d) fold 4, (e) fold 5.

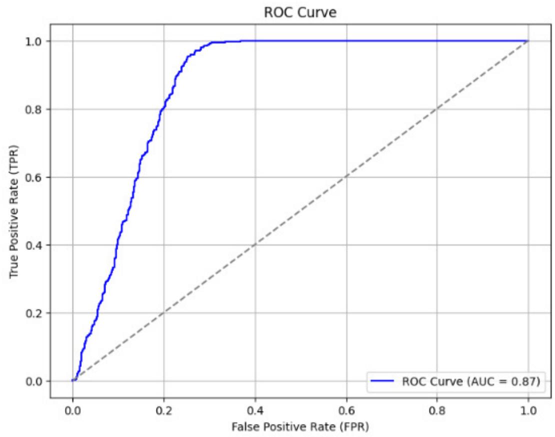

Fig. 5. ROC curve of the Dual-Branch CNN model.

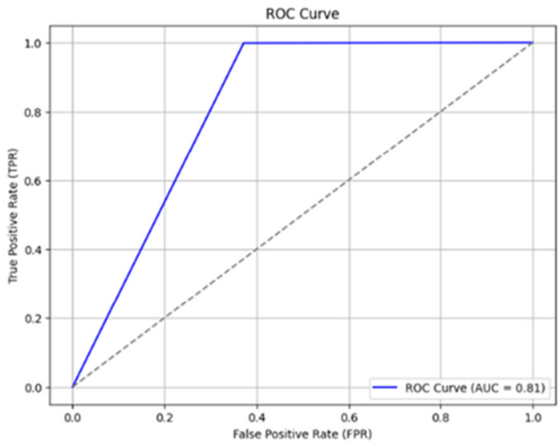

Fig. 6. ROC curve of the Siamese model.

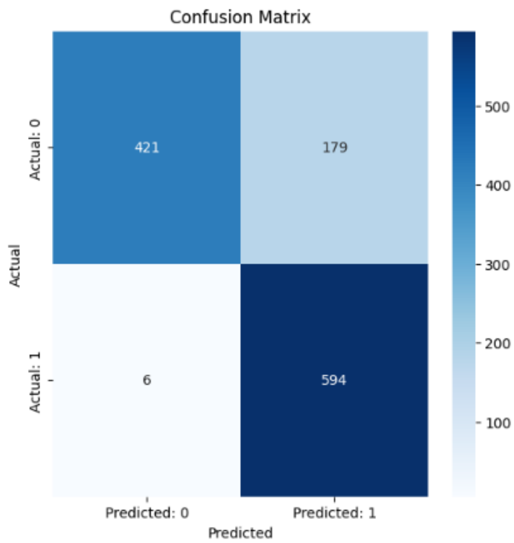

Fig. 7. Confusion matrix of the dual-branch CNN model predictions on the test dataset.

Figures 5 and 6 show the ROC curves of the dual-branch CNN and the Siamese models on the test dataset. Figures 7 and 8 show the confusion matrices of the models on the test dataset. These results show that both models performed very well with FN and not so well with FP, due to the high similarity of the images in the dataset, as most images have similarities of more than 80% or 15 of 19 matching attributes.

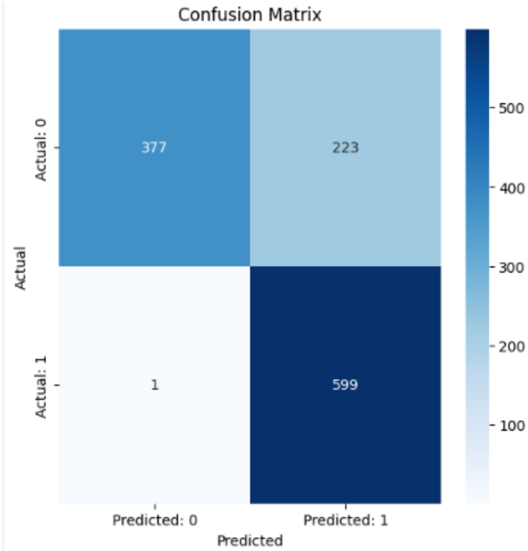

Fig. 8. Confusion matrix of the Siamese model predictions on the test dataset.

C. Discussion

Based on the experimental results, the key strengths of the proposed dual-branch CNN model for image comparison and similarity evaluation for presentation style coherence can be summarized as follows:

- Parallel feature extraction: Each branch independently extracts unique features from the input, enhancing the model's ability to capture varied aspects of image slides, such as stylistic or structural elements. This is beneficial for identifying even subtle differences, such as logo positioning, that contribute to overall coherence.
- Modularity and flexibility: The dual-branch structure supports various tasks by tailoring each branch for different inputs. This versatility is advantageous for comparing styles, text-image relationships, and performing multiview analysis.
- Effective feature combination: Concatenating outputs from both branches enriches feature representation by combining diverse features. This approach is beneficial for tasks that rely on insights from two distinct inputs, as is the use case presented for style comparison in images.
- Regularization techniques: The use of BatchNormalization and Dropout mitigates overfitting, promoting robust learning, even with limited datasets, and enhances generalizability.

- Scalability: The model can accommodate larger datasets and more complex tasks by adding layers or branches, offering scalability for future needs.

The weaknesses of the proposed model are:

- Computational complexity: Utilizing four convolutional blocks per branch increases the parameter count and computational load, requiring substantial memory and processing resources.
- Potential overfitting: Despite the regularization utilized, the model may still overfit with insufficient or homogeneous data. Careful training and data augmentation are crucial to maintaining generalizability.
- Training time: The dual-branch depth of the proposed model results in extended training times, especially with large datasets or high-resolution images.
- Interpretability challenges: Deep CNNs and particularly multi-branch architectures are often "black boxes," making it challenging to interpret specific learned features and the reasoning behind predictions.
- Task-specific nature: While effective for dual-stream comparisons such as style coherence, the dual-branch CNN may not be necessary for simpler tasks where a single-branch CNN would suffice.

In future work, the proposed method can be further optimized using advanced metaheuristic-based optimization techniques to improve its efficiency and robustness. Particle Swarm Optimization (PSO) and Pelican Optimization Algorithm (POA) are optimization techniques that can handle complex multi-dimensional optimization problems efficiently and have demonstrated their effectiveness in various applications [20, 21]. PSO and POA can be used to refine the embedding space of the dual-branch CNN to enhance similarity detection. Both metaheuristics can also be used for hyperparameter tuning of the proposed dual-branch CNN architecture, providing a comprehensive search of the hyperparameter space and considering the optimization of the learning rate, batch size, and embedding dimensionality.

V. CONCLUSION

Automating the validation process of presentation style coherence enhances the efficiency of content creation, ensuring adherence to design standards in corporate branding, professional presentations, and design audits. The proposed approach aims to simplify the creation of polished presentations that meet design standards and foster engagement. This study introduced a novel dual-branch CNN model that advances the field by offering a more effective approach for recognizing style similarities in presentation slides. Unlike traditional methods, such as the baseline Siamese network, the proposed model captures complex visual patterns and subtle discrepancies, such as logo positioning or text alignment, which are critical to maintaining design coherence.

The evaluation results demonstrate the superiority of the proposed dual-branch CNN model, outperforming the baseline Siamese model in several key metrics and underscoring its effectiveness. It achieved higher accuracy (0.85 vs 0.81), improved Jaccard similarity (0.76 vs 0.72), and a better Kappa coefficient (0.69 vs 0.62), demonstrating its enhanced ability to capture relevant patterns and provide more accurate predictions. The ROC-AUC also shows a significant improvement (0.87 vs 0.81), reflecting better discrimination between similar and dissimilar styles. Precision increased from 0.73 to 0.77, reducing false positives and ensuring a more reliable positive identification process, while the F1-score increased to 0.87, indicating a better balance between precision and recall. These metrics highlight the ability of the proposed model to capture detailed design elements while balancing the identification of relevant positives and minimizing false detections. This performance gain stems from the novel dual-branch architecture, where parallel feature extraction in each branch captures distinct stylistic and structural details. The integration of these outputs results in a comprehensive feature representation, making the model highly effective for nuanced style comparisons and text-image relationships.

Compared to other works, the novelty of this approach lies in its modular and flexible design, which supports multiview analysis and adapts to detect more intricate patterns. This flexibility positions the model as a significant advancement over simpler architectures that struggle to capture subtle stylistic variations. Additionally, the use of dedicated regularization layers mitigates overfitting and ensures robust learning even with limited datasets, addressing a common limitation in similar works. The proposed architecture contributes to current research in the field by offering a practical tool for style validation with applications in presentation advisors aimed at assisting users in creating visually consistent and professional content. Although the increased complexity of the model demands more training time, the substantial performance improvements justify the trade-off, marking a significant step forward in the automation of design quality assessment.

ACKNOWLEDGMENT

This study was funded by European Union-Next Generation EU through the National Recovery and Resilience Plan of the Republic of Bulgaria under the project BG-RRP 2.004-0005 "Improving the research capacity and quality to achieve international recognition and resilience of TU-Sofia (IDEAS)".

REFERENCES

[1] S. M. Kosslyn, R. A. Kievit, A. G. Russell, and J. M. Shephard, "PowerPoint® Presentation Flaws and Failures: A Psychological Analysis," *Frontiers in Psychology*, vol. 3, Jul. 2012, https://doi.org/10.3389/fpsyg.2012.00230.

[2] D. A. Makandar, A. Patrot, and B. Halalli, "Color Image Analysis and Contrast Stretching using Histogram Equalization," *International Journal of Advanced Information Science and Technology*, vol. 3, no. 7, pp. 27–33, 2014.

[3] F. S. Mohamad, A. A. Manaf, and S. Chuprat, "Histogram matching for color detection: A preliminary study," in *2010 International Symposium on Information Technology*, Kuala Lumpur, Malaysia, Jun. 2010, pp. 1679–1684, https://doi.org/10.1109/ITSIM.2010.5561637.

[4] N. Sahu and M. Sonkusare, "A Study on Optical Character Recognition Techniques," *International Journal of Computational Science,*

Information Technology and Control Engineering, vol. 4, no. 1, pp. 01–15, Jan. 2017, https://doi.org/10.5121/ijcsitce.2017.4101.

[5] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large Scale Online Learning of Image Similarity through Ranking," *Journal of Machine Learning Research*, vol. 11, pp. 1109–1135, 2009.

[6] H. Hsu, X. He, Y. Peng, H. Kong, and Q. Zhang, "PosterLayout: A New Benchmark and Approach for Content-Aware Visual-Textual Presentation Layout," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, Jun. 2023, pp. 6018–6026, https://doi.org/10.1109/CVPR52729.2023.00583.

[7] J. Gartner-Schmidt, "The New Normal – Virtual and Hybrid Presentations: Developing Content, Designing Slides, and Delivery Guidelines," *Ear, Nose & Throat Journal*, vol. 101, no. 9_suppl, pp. 20S-28S, Nov. 2022, https://doi.org/10.1177/01455613231158799.

[8] P. Bhargava and O. A. Awan, "Mastering Presentation Zen: How to Improve Presentation Design and Educational Impact," *Current Problems in Diagnostic Radiology*, vol. 50, no. 5, pp. 571–573, Sep. 2021, https://doi.org/10.1067/j.cpradiol.2020.11.008.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International Conference on Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 2564–2571, https://doi.org/10.1109/ICCV.2011.6126544.

[10] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Bombay, India, 1998, pp. 59–66, https://doi.org/10.1109/ICCV.1998.710701.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004, https://doi.org/10.1109/TIP.2003.819861.

[12] V. Appana, T. M. Guttikonda, D. Shree, S. Bano, and H. Kurra, "Similarity Score of Two Images using Different Measures," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, Jan. 2021, pp. 741–746, https://doi.org/10.1109/ICICT50816.2021.9358789.

[13] J. Wang *et al.*, "Learning Fine-Grained Image Similarity with Deep Ranking," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 1386–1393, https://doi.org/10.1109/CVPR.2014.180.

[14] V. Lytvyn, R. Peleshchak, I. Rishnyak, B. Kopach, and Y. Gal, "Detection of Similarity Between Images Based on Contrastive Language-Image Pre-Training Neural Network," in *Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Systems. Volume I: Machine Learning Workshop*, Apr. 2024, https://doi.org/10.31110/COLINS/2024-1/008.

[15] S. Merugu, R. Yadav, V. Pathi, and H. R. Perianayagam, "Identification and Improvement of Image Similarity using Autoencoder," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15541–15546, Aug. 2024, https://doi.org/10.48084/etasr.7548.

[16] D. Chicco, "Siamese Neural Networks: An Overview," in *Artificial Neural Networks*, H. Cartwright, Ed. Springer US, 2021, pp. 73–94.

[17] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, Dec. 2016, pp. 378–383, https://doi.org/10.1109/ICPR.2016.7899663.

[18] Ž. Đ. Vujovic, "Classification Model Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021, https://doi.org/10.14569/IJACSA.2021.0120670.

[19] N. Gessert and A. Schlaefer, "Learning Preference-Based Similarities from Face Images using Siamese Multi-Task CNNs." arXiv, Jan. 25, 2020, https://doi.org/10.48550/arXiv.2001.09371.

[20] R. Z. Khaleel, H. Z. Khaleel, A. A. A. Al-Hareeri, A. Sh. M. Al-Obaidi, and A. J. Humaidi, "Improved Trajectory Planning of Mobile Robot Based on Pelican Optimization Algorithm," *Journal Européen des Systèmes Automatisés*, vol. 57, no. 4, pp. 1005–1013, Aug. 2024, https://doi.org/10.18280/jesa.570408.

[21] H. Z. Khaleel and A. J. Humaidi, "Towards accuracy improvement in solution of inverse kinematic problem in redundant robot: A comparative analysis," *International Review of Applied Sciences and Engineering*, vol. 15, no. 2, pp. 242–251, Jan. 2024, https://doi.org/10.1556/1848.2023.00722.