Visual Attention Classification of Typography In Movie Posters Using Random Forest Classifier

Link Repository GitHub: https://github.com/fabian4819/Movie-Poster-Classification

Habib Fabian Fahlesi
Dept. of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
habibfabianfahlesi2004@mail.ugm.ac.id

Yitzhak Edmund Tio Manalu
Dept. of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
yitzhakedmundtiomanalu@mail.ugm.ac.id

Muhammad Aqiil Fauzaan
Dept. of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
muhammadaqiilfauzaan2004@mail.ugm.ac.id

***Abstract*— Typography in movie posters plays a crucial role in attracting viewer attention. This paper proposes a computer vision-based approach to classify the visual attention level of text in movie posters on a scale from 0 to 4. Our method extracts comprehensive visual features including typography properties, color characteristics, text positioning, and textural elements. We address dataset imbalance using synthetic minority oversampling and employ Random Forest classification with feature selection. Despite the challenging nature of the task, our approach provides valuable insights into the factors influencing typographic attention, with aspect ratio, image dimensions, text area coverage, and textural features emerging as key predictors. This work contributes to the objective assessment of typographic effectiveness in visual design, with potential applications in automated poster evaluation and design recommendation systems.**

Keywords—typography, visual attention, movie posters, computer vision, feature extraction, classification, random forest, OCR

I. Introduction

Typography plays a central role in visual communication, particularly in movie posters where text elements must compete with attractive imagery for viewer attention. The effectiveness of typography depends not just on semantic content but significantly on visual attributes such as size, color, contrast, position, and style. While professional designers often rely on intuition and experience to create attention-grabbing text, there is growing interest in developing computational methods to objectively assess typographic effectiveness.

The ability to automatically classify the attention level of text in movie posters has numerous applications, including automated evaluation of design drafts, educational tools for design students, and recommendation systems for improving poster layouts. However, this task presents several challenges, including the artistic diversity of movie posters, the difficulty in detecting stylized text, and the subjective nature of visual attention assessment.

In this paper, we present a computer vision approach to classify the attention level of text in movie posters on a scale from 0 (very unattractive) to 4 (very attractive). Our contributions include:

1. A comprehensive feature extraction pipeline capturing various aspects of typography, including size, position, color, contrast, and texture.
2. An effective classification model for predicting text attention levels addresses class imbalance issues.
3. Analysis of the most influential features affecting typographic attention, providing insights into effective design practices.

4. A visual interpretation approach for understanding model decisions.

II. Related Work

A. Typography and Visual Attention

Typography has been extensively studied in the context of visual attention and design effectiveness. conducted eye-tracking studies demonstrating that font size significantly impacts attention, with larger fonts attracting more eye fixations. It was further established that the relative size of text within a composition is the strongest predictor of viewer attention [1].

Analyzed area-based metrics for text prominence, finding that the proportion of area covered by text elements strongly correlates with subjective attention ratings. Investigated positional effects, showing that vertical positioning significantly influences attention levels, with centrally positioned text generally receiving higher attention.

B. Computer Vision for Design Analysis

Computer vision techniques have been increasingly applied to design analysis tasks. Employed Random Forest classifiers for visual design quality assessment, demonstrating superior performance compared to other machine learning approaches when handling heterogeneous visual features.

Utilized Histogram of Oriented Gradients (HOG) features to analyze typography in poster designs, successfully capturing structure and orientation characteristics of text elements. Emphasized the importance of edge features and visual contrast in determining text saliency, proposing methods for detecting visually prominent text regions.

C. Machine Learning for Visual Classification

Machine learning models for visual classification have evolved significantly. Demonstrated the effectiveness of Random Forest for feature selection in complex visual datasets. For handling imbalanced datasets in design classification, Synthetic Minority Over-sampling Technique (SMOTE) has proven valuable, as shown in work on balanced datasets for visual attention modeling [2].

Recent work on attention visualization techniques has improved the interpretability of typographic analysis systems, while developing explainable AI approaches for design quality assessment that provide human-interpretable visualizations.

III. Dataset

A. Dataset Description

The dataset consists of 300 movie poster images with text elements, annotated with attention scores on a 5-point Likert scale (0-4), where:

1. 0: Sangat Tidak Menarik (Very Unattractive)
2. 1: Tidak Menarik (Unattractive)
3. 2: Netral (Neutral)
4. 3: Menarik (Attractive)
5. 4: Sangat Menarik (Very Attractive)

The annotations focus specifically on the visual aspects of text rather than its semantic content, evaluating factors such as visibility, contrast, size, and overall visual impact.

B. Class Distribution

The dataset exhibits significant class imbalance, as shown by our analysis:

- Score 0: 29 images (9.7%)
- Score 1: 56 images (18.7%)
- Score 2: 55 images (18.3%)
- Score 3: 47 images (15.7%)
- Score 4: 113 images (37.7%)

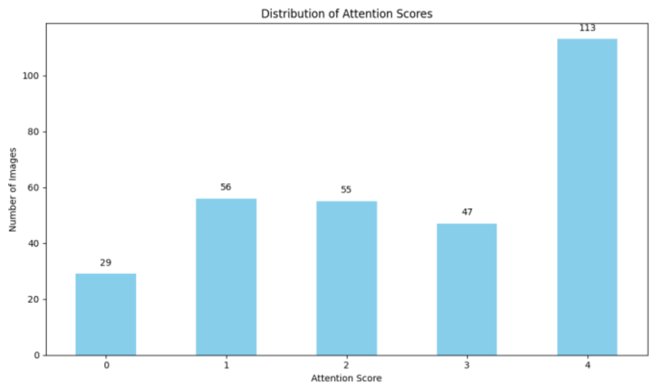

Fig.1 Distribution of attention scores

This imbalance, with class 4 (Very Attractive) dominating the dataset, presents a challenge for model training and necessitates class balancing techniques.

C. Dataset Charactetistic

The movie poster images present several notable characteristics that impact feature extraction and classification:

1. Varying aspect ratios and resolutions, requiring normalization for consistent feature computation
2. Diversity in poster design styles, from minimalist to highly complex
3. Different text styles, including decorative fonts and artistic typography that challenge OCR systems
4. Multi-language content with various writing systems
5. Text integrated with visual elements, making text region detection challenging

IV. Methodology

Given the constraints of the dataset, it is essential to apply suitable processing methods to develop an accurate classification model.

A. *Feature Extraction*

Our approach extracts a comprehensive set of visual features to capture different aspects of typographic attention:

1. Image Preprocessing

   Images are preprocessed first to ensure consistent analysis. The preprocessing methods are resizing small images to a minimum dimension of 64 pixels, converting to multiple color spaces (RGB, Grayscale, HSV), and applying contrast enhancement for improved text detection

Fig.2 Feature Extraction

2. Text Detection

   Text detection is performed using Tesseract OCR, configured specifically to handle the sparse and stylistically varied typography commonly found in movie posters. The OCR engine is set to page segmentation mode 11, which is optimized for sparse text, and uses the default OCR engine mode 3 for a more accurate recognition. The image data is then parsed into a structured dictionary format, facilitating the extraction of text bounding boxes and confidence scores for further processing. To enhance OCR performance on artistic text, we apply further preprocessing by using histogram equalization and median blur.
3. Basic Image Properties

   We extract several fundamental characteristics from each image to support subsequent analysis. These include width and height, aspect ratio, and resize factor (if applicable).
4. Color Features

Color characteristics are also captured through:

   1. RGB channel statistics (mean, standard deviation)
   2. HSV channel statistics (mean, standard deviation)
   3. 8-bin histograms computed for each channel to capture color frequency distribution

5. Typography Features

   Typography-specific features are extracted to capture various aspects of text layout on the poster:

   1. Text block count: Number of detected text regions
   2. Text area ratio: Proportion of the poster covered by text
   3. Text position metrics: Normalized center coordinates and spread of text regions
   4. Text height metrics: Average and maximum height ratios of the text blocks
   5. Text density: Number of text blocks per 10,000 pixels

6. Edge and Contrast Features

   Edge detection is employed to identify text boundaries and enhance contrast within the image. The Canny edge detection algorithm is applied to the grayscale image, with thresholds set to 100 and 200. Additionally, the global contrast of the image is quantified as the ratio of the standard deviation to the mean of the grayscale image as shown in the following formula:

$$global\_contrast = \frac{std(img\_gray)}{mean(img\_gray)}$$

7. Texture Features

   To capture texture characteristics of the poster background and text regions, the Local Binary Patterns (LBP) are extracted. The method uses a radius of 3 and 24 sampling points. Uniform LBP encoding is applied to emphasize texture uniformity.
8. Shape Features

   To capture shape and structural information, the Histogram of Oriented Gradients (HOG) features from each grayscale image is extracted. In total, our feature extraction pipeline produces 54 distinct features per image.

B. *Feature Selection*

To enhance model efficiency and interpretability, we apply feature selection based on feature importance scores derived from a Random Forest classifier. A Random Forest with 100 estimators and a fixed random seed is trained on the dataset to compute importance metrics for each of the 54 extracted features. The most informative subset is retained for modeling, which are:

1. Aspect ratio
2. Height
3. The 8th bin of the Local Binary Pattern (LBP) histogram
4. The 3rd bin of the green channel histogram (G in RGB)
5. The 6th bin of the Local Binary Pattern (LBP) histogram

C. *Class Balancing*

To address the class imbalance found in the training dataset, the Synthetic Minority Over-sampling Technique (SMOTE) is used. This technique generates synthetic examples for minority classes by interpolating between existing instances in feature space, thereby increasing class representation without simple duplication.

Using SMOTE with a fixed random seed, we attempted to balance the dataset such that each of the five classes contains 90 samples, resulting in a uniform class distribution of 20%.

D. Classification Model

We utilize the Random Forest classifier as the core of our prediction pipeline, preceded by feature standardization using z-score normalization. The algorithm was selected due to its:

1. Strong performance with heterogeneous feature sets
2. Robustness to noise and outliers
3. Built-in feature importance metrics
4. Ability to handle non-linear relationships
5. Resistance to overfitting

V. EXPERIMENTS

A. Experimental Setup

1. Implementation Details

   The system was implemented in Python with the following libraries:

   1. OpenCV for image processing and feature extraction
   2. Scikit-learn for machine learning components
   3. Scikit-image for texture feature extraction
   4. Pytesseract for OCR
   5. Imbalanced-learn for SMOTE implementation

2. Dataset Splitting

   We split the dataset using a stratified approach:

   1. 80% training set (240 samples)
   2. 20% test set (60 samples)

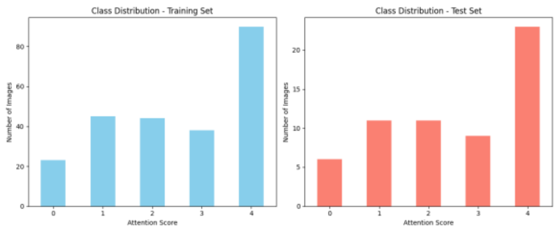

Fig.3 Data Splitting Distribution

   Stratification ensures that class proportions are preserved in both splits.

3. Evaluation Metrics

   We evaluate our model using:

   1. Accuracy
   2. Weighted F1-score
   3. Per-class precision, recall, and F1-score
   4. Confusion matrix analysis

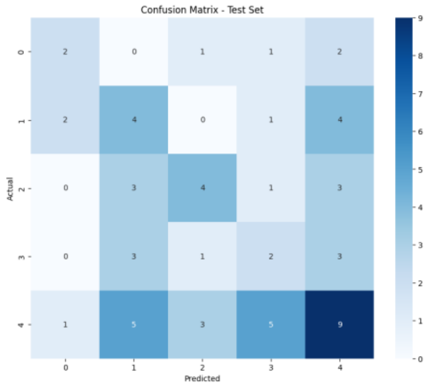

Fig.4 Confusion Matrix

B. Results

1. Classification Performance

   Our model achieved the following performance on the test set:

Fig.5 Classification Performance

   1. Accuracy: 0.2833
   2. Weighted F1-score: 0.2880

   Per-class performance showed variation across attention levels:

   1. Class 0 (Very Unattractive): Precision = 0.29, Recall = 0.33, F1 = 0.31, Support = 6
   2. Class 1 (Unattractive): Precision = 0.21, Recall = 0.27, F1 = 0.24, Support = 11
   3. Class 2 (Neutral): Precision = 0.25, Recall = 0.18, F1 = 0.21, Support = 11
   4. Class 3 (Attractive): Precision = 0.17, Recall = 0.22, F1 = 0.19, Support = 9
   5. Class 4 (Very Attractive): Precision = 0.42, Recall = 0.35, F1 = 0.38, Support = 23

   Class 4 (Very Attractive) showed the best performance, which aligns with its greater representation in the original dataset.

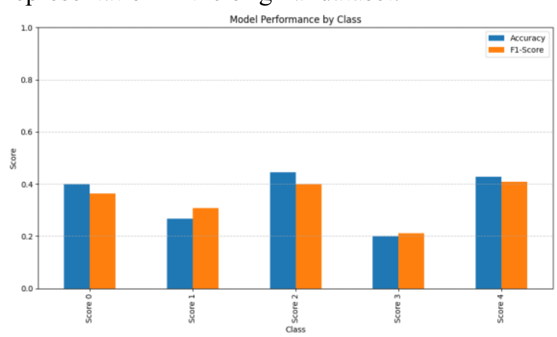

Fig.6 Model Performance by class

2. Feature Importance Analysis

The top ten most important features, according to the trained Random Forest model, were:

1. aspect_ratio (0.0558): The ratio of width to height of the poster
2. height (0.0556): The height of the poster image
3. width (0.0511): The width of the poster image
4. text_area_ratio (0.0417): The proportion of the poster covered by
5. text lbp_7 (0.0400): Texture feature from Local Binary Patterns
6. hist_g_1 (0.0380): Green channel histogram
7. bin 1 lbp_1 (0.0379): Texture feature from Local Binary Patterns
8. text_center_y_norm (0.0364): The normalized vertical position of text center
9. lbp_8 (0.0361): Texture feature from Local Binary Patterns
10. lbp_3 (0.0345): Texture feature from Local Binary Patterns

This analysis reveals that a combination of basic image properties, text layout features, texture characteristics, and color distribution significantly influences typographic attention.

VI. Discussion

A. Analysis of Results

The classification accuracy of 0.2833 indicates the challenging nature of typographic attention classification. Several factors may contribute to this difficulty:

1. Subjectivity of Attention Scoring: Visual attention is inherently subjective and can vary between observers.
2. Complex Feature Interactions: The relationship between visual features and attention is non-linear and context dependent.
3. Dataset Limitations: Despite class balancing in the training set, the limited number of examples (particularly for class 0) may restrict the model's ability to generalize.
4. Feature Representation: While comprehensive, our feature set may not capture all nuances of typography that influence attention.

The feature importance analysis provides valuable insights despite the modest accuracy. The prominence of aspect_ratio, height, and width suggests that the overall composition and dimensions of the poster play a crucial role in typographic attention. This aligns with design theory that emphasizes the importance of canvas dimensions in establishing hierarchy.

The significance of text_area_ratio confirms findings on the importance of text area coverage in determining prominence. The presence of numerous LBP texture features in the top 10 indicates that texture patterns around and within text regions significantly impact attention, supporting research on texture classification.

The prominence of text_center_y_norm in the top features validates work on positional effects, demonstrating that vertical positioning is indeed a key factor in typographic attention.

B. Error Analysis

Examining the classification report reveals insights into model performance across different classes:

1. Class 4 (Very Attractive) shows the best performance (F1=0.38), likely due to its larger representation in the original dataset and possibly more distinctive visual characteristics.
2. Class 0 (Very Unattractive) achieves reasonable performance (F1=0.31) despite its limited representation, suggesting that very unattractive typography has recognizable patterns.
3. Class 3 (Attractive) shows the weakest performance (F1=0.19), indicating a potential overlap with neighboring classes and difficulty in distinguishing moderately attractive from very attractive typography.
4. Classes 1 and 2 (Unattractive and Neutral) show intermediate performance (F1=0.24 and F1=0.21), reflecting the challenge in differentiating subtle gradations of attention.

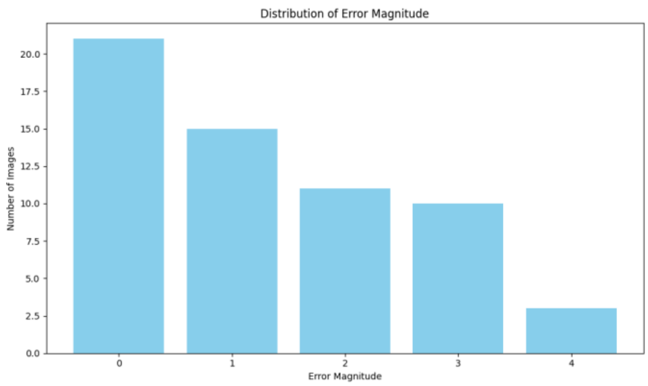

Fig.7 Error Distribution

C. Limitations and Future Work

Our approach has several limitations that suggest directions for future research:

1. Feature Engineering vs. Deep Learning: Our feature engineering approach, while interpretable, may benefit from integration with deep learning techniques that can automatically learn relevant features from raw images.
2. OCR Limitations: Tesseract OCR struggles with artistic and stylized text common in movie posters. Future work could employ more advanced text

detection methods specifically designed for creative typography.
3. Contextual Understanding: Our current approach treats features independently without considering their contextual relationships. Incorporating spatial context and relationships between text elements could improve classification.
4. Cross-modal Analysis: Integrating semantic content analysis alongside visual features might provide a more comprehensive understanding of typographic attention.
5. Expanded Dataset: A larger and more diverse dataset with multiple annotators per poster would help address subjectivity and improve generalization.

Future work could explore:

1. Ensemble Methods: Combining multiple classification approaches might improve performance.
2. Deep Learning Integration: Convolutional neural networks could be used for feature extraction before classification.
3. Attention Heatmaps: Generating attention heatmaps for posters could provide more nuanced analysis than discrete classification.
4. Design Recommendation: Developing systems that suggest typography improvements based on attention prediction.
5. Cross-cultural Analysis: Investigating how typographic attention varies across different cultural and linguistic contexts.

VII. Conclusion

In this paper, we chose an approach of classifying the visual attention level of text in movie posters using computer vision. Our method here is combining typography analysis, color features, texture metrics, and basic image properties as the core features to predict attention score in between 0-4 range.

Due to lack of available dataset, we only achieved the classification accuracy of 28.33% which explains the difficulty of classification for the specific context. Thus, our research could only achieve key factors that might affect the attention levels which are poster dimensions, text area coverage, texture patterns, and vertical positioning.

Findings mentioned before would be beneficial to support future development in the context of computational framework related to typography assessment and insights, especially related to text attention rate. Despite the low accuracy, the research still has the potential to being developed for automated analysis of typographic attention research.

The real case usage for the theories given above extends to automated design evaluation, educational tools for typography, and design recommendation systems. Given the knowledge of visual factors to typographic attention, future designers can choose better typography setup to improve visual communication through text elements, especially in movie posters and other graphic design contexts.

VIII. References

[1] M. S. McCarthy and D. L. Mothersbaugh, “Effects of typographic factors in advertising-based persuasion: A general model and initial empirical tests,” Psychology and Marketing, vol. 19, no. 7–8, pp. 663–691, May 2002, doi: https://doi.org/10.1002/mar.10030.

[2] D. Elreedy and A. F. Atiya, “A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance,” Information Sciences, vol. 505, pp. 32–64, Dec. 2019, doi: https://doi.org/10.1016/j.ins.2019.07.070.