BSc Hons in Computing

2023.1

Semantic Font Recommendation System Using Deep Learning

Coventry Index	:	14382400
NIBM Registered Name	:	D.R.H.Dunumala
NIBM Index	:	COBSCCOMPY4231P018

Faculty of Engineering, Environment and Computing
School of Computing, Electronics and Mathematics
Coventry University

School of Computing and Engineering
National Institute of Business Management

Table of Content

Chapter 1

Introduction

(1.1) Background

Typography is a critical element in digital and print design. The choice of font impacts readability, user experience, and brand identity. Designers and developers often face(2021, February 26). the challenge of selecting a suitable font from thousands of available options. This process is time-consuming and requires significant expertise.

Current solutions, such as font libraries like Google Fonts and Adobe Fonts, assist users by organizing fonts with manual tags (serif, modern, handwriting). While useful, these systems are limited. Users must search using pre-defined keywords, and the tags do not capture the semantic or emotional context of a design brief(2019, August 27). . There is a need for a more intuitive system that understands natural language descriptions to suggest appropriate fonts.

(1.2) Research Problem

The primary problem is the inefficiency and difficulty in selecting fonts based on abstract or stylistic goals. Designers and non-designers lack an effective tool that can translate a high-level creative description, such as "a font for a trustworthy financial institution," into a ranked list of suitable typefaces. The reliance on manual searching and pre-defined tags creates a gap between the user's creative intent and the final font selection.

(1.3) Objectives of the Project

The main objectives of this research are:

1. To develop a system that uses a computer vision model to analyze and extract the visual characteristics of different fonts.
2. To implement a method that converts natural language text prompts describing font styles into quantitative vector representations.

3. To build and train a deep learning model that learns the relationship between the semantic meaning of text prompts and the visual features of fonts.
4. To develop a working prototype application that takes a user's text description and suggests a list of relevant fonts.

(1.4) Research Questions

1. How can a Convolutional Neural Network (CNN) be utilized to extract meaningful visual feature vectors from font images?
2. How effectively can a pre-trained language model capture the semantic intent from a user's textual description of a desired font?
3. Can a mapping be learned between the text embedding space and the font visual embedding space to enable accurate, context-aware font recommendations?

(1.5) Scope of the research

This research will focus on developing a proof-of-concept font recommendation system. The dataset will consist of publicly available, open-source fonts, primarily from the Google Fonts library, and will be limited to the Latin alphabet. The final deliverable will be a prototype application with a simple user interface to demonstrate the system's functionality. The project will not involve the creation or generation of new fonts. The evaluation will consist of quantitative performance metrics and a small-scale qualitative user study.

Chapter 2

Literature Review

(2.1) Findings by other researchers

Existing research provides a foundation for this project in three main areas. First, in font recognition, studies have successfully used Convolutional Neural Networks (CNNs) to classify font families and styles from images, demonstrating that deep learning models can effectively extract typographic features.

Second, in the broader field of computer vision, significant work has been done on style analysis and transfer. These studies focus on teaching models to recognize the artistic style of images, which is analogous to recognizing the stylistic "feel" of a font beyond simple classification.

Third, and most relevant, is the field of cross-modal retrieval, particularly text-to-image systems. Research by Radford et al. (2021) on the CLIP model showed that it is possible to train a model to learn a shared embedding space for both text and images. This allows for powerful zero-shot image classification based on natural language descriptions. This technology forms the conceptual basis for linking text prompts to font images.

(2.2) The research gap

While the component technologies exist, a specific research gap is present in their application to the domain of typography. Current font suggestion tools are based on manual, curated tags, not a learned semantic understanding. There is limited research focused on building a cross-modal system that directly maps free-form natural language descriptions to the visual aesthetics of fonts. This project aims to address this gap by creating a model that learns this semantic relationship directly from the data.

(2.3) Chapter conclusion

The literature confirms that deep learning is effective for analyzing both images and text. The underlying technologies for creating a text-to-image retrieval system are well-established. However, the novel application of these techniques to build a semantic font recommendation engine remains a distinct and valuable area for research.

Chapter 3

Methodology

This project will follow a quantitative, experimental methodology structured in four stages.

1. Data Collection and Preprocessing

The dataset will be constructed using open-source fonts from the Google Fonts repository. A Python script using the Pillow library will be developed to automate the preprocessing stage. This script will render each font file into a standardized 256x256 pixel PNG image containing a consistent string of characters (e.g., "A-Z, a-z"). Corresponding metadata, including font category and tags, will be extracted to create a mapping file.

2. Model Architecture and Training:

The core of the system will be a deep learning model with two main components.

- **Visual Feature Extraction:** A pre-trained CNN (e.g., ResNet50) will be used via transfer learning. Font images will be fed into this model to produce high-dimensional feature vectors (embeddings) for each font.
- **Text Feature Extraction:** A pre-trained Sentence Transformer model (e.g., 'all-MiniLM-L6-v2') will be used to convert user text prompts and font tags into text embeddings of a compatible dimension.
- **Mapping Model:** A small neural network will be trained using a contrastive learning approach. This network will learn to align the visual font embeddings and the semantic text embeddings into a shared space, so that a text prompt and its corresponding font style are mathematically close.

3. Implementation and Prototyping

The system will be implemented in Python using the TensorFlow or PyTorch framework. The final prototype will be a simple web application built with the Streamlit library. The application will allow a user to enter a text prompt and will display the top-ranked font image suggestions.

4. Evaluation

The model's performance will be evaluated using a two-fold approach. First, a quantitative evaluation will measure the model's Top-K accuracy on a held-out test set. Second, a qualitative evaluation will be conducted through a user study, where participants will rate the relevance and quality of the font suggestions for a given set of prompts.

Refferance

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). *Deep residual learning for image recognition*. arXiv.org. https://arxiv.org/abs/1512.03385

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021, February 26). *Learning transferable visual models from natural language supervision*. arXiv.org. https://arxiv.org/abs/2103.00020

Reimers, N., & Gurevych, I. (2019, August 27). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. arXiv.org. https://arxiv.org/abs/1908.10084