Generating a Typeface

Background

According to Google, there have been over 130,000 typefaces published (fonts) for the Roman alphabet. Each, at a minimum, consists of 26 characters, 10 digits and a dozen or so punctuation marks. Creating custom typefaces is something commonly done for brands in the marketing world, with many desiring something unique to their brand or corporate identity. One major reason that brand design projects can garner 7-figure price tags is the development of a custom font.[1] The typeface design process can quickly consume hundreds of creative hours and many rounds of iterations making it an expensive activity to undergo using traditional methods.

The Data

Google Fonts is a robust catalog of open-source fonts that's striving to make it as easy as possible to integrate expressive type design on the web.[2] The project maintains an open repository on github that includes over 1,100 fonts. In addition to these publicly available fonts, my client and employer, Landor & Fitch, provides licensed access to over 2,500 fonts through a private cloud. Both Google Fonts and the font cloud provide fonts in the form of TrueType Font (.ttf) files. The .ttf file format contains mathematical representations of each character in a font.

Project Scope

For this project, the question I'll seek to answer is: can machine learning make the design of new/custom typefaces more accessible?

Given all of the available knowledge about fonts and typefaces, would it be less time intensive to generate font designs through a generative model? To answer this question, I'll attempt to generate a new to the world typeface using generative methods such as via an Autoencoder or Generative Adversarial Network (GAN). The ultimate goal being to produce a font file (.ttf)—or set of high-quality images that could be used to produce a font file—that can be installed on a machine and used in a standard word processor. I will explore various techniques and model architectures to arrive at as limited of an input as possible.

As no official dataset exists, part of the scope of this project will be to construct an image dataset from the available 3,500+ .ttf files. Using image processing methods, I'll create uniform images out of the mathematical representations of each character, so that each image in the dataset represents a single, grayscale image of a typeface character. The scope will be limited to 62 characters of the Roman alphabet—uppercase and lowercase letters (52) and digits (10).

The resulting deliverables will include:

---

[1] https://www.businessinsider.com/heres-how-much-the-worlds-most-iconic-logos-cost-companies-2013-3
[2] https://fonts.google.com/about

- An image dataset of ~225,000 images—62 character classes for ~3,650 fonts
- A generative model capable of creating a wholistic set of character images (62)
- Code, in the form of Jupyter Notebooks
- A written, technical report of the findings, methodology, and model performance
- A slide deck of the findings designed for stakeholders & creatives of Landor & Fitch:
    - Executive Creative Director, Landor & Fitch Chicago
    - Managing Director, Landor & Fitch Chicago
    - Design Directors
    - Graphic Designers
    - Copywriters