Lang Resources & Evaluation
https://doi.org/10.1007/s10579-020-09499-0

ORIGINAL PAPER

Semantics-aware typographical choices via affective associations

Tugba Kulahcioglu[1] · Gerard de Melo[1]

Abstract With the tens of thousands of fonts that are now readily available, it is non-trivial to select the most suitable font for a given use case. Considering the impact of the choice of font on human perception of the text, there is a strong need for semantic font search and recommendation. Aiming to fulfill this need, we induce a typographical lexicon providing associations between words and fonts. For this purpose, we determine font vectors for basic and complex emotions, based on word similarities, antonymy information, and Plutchik's Wheel of Emotions. We create a large font lexicon, named FontLex, relying on emotion associations between the words and the fonts. We evaluate our results through user studies and find that for the majority of the evaluated words, the fonts recommended by FontLex are preferred. We also further extend the dataset using synonyms of font attributes and emotion names. Finally, using CNN embeddings of the fonts, we expand our attribute score assignment to new fonts. The resulting FontLex resource provides mappings between 6.7K words and 2K fonts. Our proof of concept application demonstrates how FontLex can be invoked to obtain semantic font recommendation for poster design.

Keywords Typography · Font · Emotion

✉ Tugba Kulahcioglu
tugba.kulahcioglu@rutgers.edu

1 Rutgers, The State University of New Jersey, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA

Published online: 08 July 2020

1 Introduction

Human studies have shown that the choice of fonts has a strong influence on brand and product perception, and price expectation (Fligner 2013; Childers and Jass 2002; Van Rompay and Pruyn 2011). Similar ties have been observed in other tools and products, such as in email perception (Shaikh et al. 2007). Typographic choices have also been considered as emotional cues by text-to-speech systems, and used to determine the prosodic features of speech (Tsonos and Kouroupetroglou 2016).

Among the tens of thousands of fonts available today, it is not easily possible for users to go through each font to pick the one that serves their use case best. It is not uncommon for people to spend several minutes looking for the right font, but finally ending up using the default one (Fox 2010). And as the number of available fonts keeps increasing, the task of font selection is becoming more challenging every day. Google Fonts[1] as of June 2019 provides a catalog of 916 font families, while broader font sharing websites[2] typically serve several thousands. This makes the font selection process particularly burdensome for graphic designers, as their profession calls for such decisions to be made on a regular basis.

Despite the obvious need, the assistance offered by current tools remains very limited. Some websites and word processing tools provide a categorized presentation of fonts for the users to explore, based on a limited number of visual (e.g., *monospace*) and semantic (e.g., *fun*) categories. Suggesting more advanced support, O'Donovan et al. (2014) present a method to recommend fonts that are similar to the current font selection. The same study also provides a relatively wider range of attributes (37 such attributes) to search fonts. FontJoy (Qiao 2017) relies on vector representations to generate font pairs. The visualization by Data Scope Analytics (2017) aims to help users in discovering aesthetically pleasing font pairs.

Previous work, however, neglects the content of the text to be formatted, and in particular neglects the affective dimension of human perception. A font that is *incongruent* with the underlying meaning or theme may adversely affect how the content is perceived.

Towards the aim of supporting the development of font recommendation tools based on the *textual content* and the associated *affect* of the message, in this study, we develop methods to induce associations between words and fonts. We rely on word–emotion and font–emotion associations to connect words with fonts via their affective associations. With these techniques, we induce a *typographical lexicon* called *FontLex*, which maps 6.7K words to a set of around 2K fonts. This article extends a previous conference publication (Kulahcioglu and De Melo 2018) with the following new contributions: (1) Using the dyads from Plutchik's Wheel of Emotion, in Sect. 5.1 we present a method to infer font vectors for complex emotions. (2) In Sect. 6, we extend the font set from 200 to around 2K using font embeddings following the approach from Kulahcioglu and De Melo (2018) (3) In Sect. 5.3, we demonstrate a sample poster design application which makes use of FontLex to achieve semantic font recommendation.

[1] https://fonts.google.com.

[2] For instance, https://www.dafont.com/ and http://www.1001fonts.com/.

The rest of this paper is organized as follows. First of all, Sect. 2 reviews related work on semantic attributes of fonts and on font recommendation techniques. Section 3 presents our method to predict emotion–font scores and evaluates it through a user study.[3] Section 4 presents our method to predict word–font scores using the previously obtained emotion–font scores, and evaluates it through a further user study. Subsequently, Sect. 5 describes the extensions we propose for the dataset to increase its accuracy and to expand its coverage to more words and fonts. In Sect. 7, we provide discussions on FontLex and its potential applications. Finally, Sect. 8 concludes the paper and outlines plans for future work.

2 Related work

We begin with a review of previous studies and tools that analyze font semantics, recommend fonts, or explore the impact of font choices.

2.1 Semantic attributes of fonts

In an online survey conducted by Shaikh et al. (2006), the characteristics of 20 fonts are assessed with respect to 15 adjective pairs (e.g., stable–unstable). The fonts are presented using alphabetic, numeral, and common symbols.

Through a crowdsourced study, O'Donovan et al. (2014) associate 200 fonts with 37 semantic attributes (e.g., happy). They ask users to pick one of two presented fonts for a given attribute, and then based on these selections assign scores between 0 and 100 for each font–attribute combination.[4] The fonts are presented using the sentence "The quick brown fox jumps over the lazy dog.", which is a sentence that contains all letters of the English alphabet and is commonly used to test font graphics by the designers. The aforementioned study will be discussed further in Sect. 3.

Kulahcioglu and de Melo (2018) extend the above crowdsourced dataset using deep Convolutional Neural Network (CNN) embeddings as a means of obtaining a similarity measure between fonts. To predict semantic attribute scores for a font outside the dataset, the authors take weighted averages of the nearest four font scores, as determined by the embeddings. Based on leave-one-out cross validation test results, the method is able to predict scores with around 9% mean absolute error.

Further studies (Velasco et al. 2014, 2015) analyze the relationship between visual font characteristics and attributes that are associated with taste (sweet, sour, etc.). Based on their user study, they discover connections such as that round fonts exhibit an association with sweet taste.

Finally, many font-focused websites (Sam Berlow and Sherman 2017; dafont.com 2017; Bloch 2017) allow contributors to tag fonts with attributes, some of which are more semantic than visual.

[3] All studies in this paper received IRB approval.

[4] http://www.dgp.toronto.edu/~donovan/font/.

2.2 Font recommendation

Font recommendation studies mainly focus on two aspects: recommending fonts that are similar to a target font, and recommending fonts that would make a good pair.[5]

Font Similarity O'Donovan et al. (2014) present a method of proposing fonts that are similar to a given font. In their experiments, they find that semantic attributes are more conducive to predicting the similarity of fonts than geometrical features. Thus, making use of a set of semantic attributes, they learn a font similarity metric based on crowdsourced comparisons, in which users need to assess which of two presented fonts is more similar to a provided reference font. Wang et al. (2015) rely on a deep learning approach to find similar fonts. It is claimed that a qualitative comparison of both methods reveals this approach as producing better results than the former one by O'Donovan et al. (2014). For those users that are interested in using font embeddings as similarity measure, the visualization[6] in Ho (2017) displays around 800 font embeddings mapped into a 2D space, which could support exploration of similar fonts.

Font Pairing Several websites, including those of Sam Berlow and Sherman (2017), Canva.com (2017) and Mills (2017), provide font pair suggestions gathered from users or from other web sources. Making use of the data from Sam Berlow and Sherman (2017), the force-directed graph visualization[7] developed by Data Scope Analytics (2017) displays 458 fonts and 1807 co-usages of fonts, with the goal of facilitating the font pairing process. In an attempt to recommend font pairs without relying on existing data, Qiao (2017) identify fonts that are both contrasting and complementary using vector representations that are provided online.[8] The system can either propose a novel pair of fonts, or suggest a second font for an already specified one.

2.3 Impact of font choices

A number of Stroop-style studies have been conducted to investigate the effect of font characteristics on perception. Hazlett et al. (2013) asked users to judge whether a displayed word is positive or negative, comparing 5 fonts and 25 words that are all strongly associated with positive or negative emotion. The results indicate that congruent typefaces yield faster responses. Lewis and Walker (1989) ask users to press a left hand key if the words *slow* or *heavy* appear, versus a right hand one if *fast* or *light* appears. In a second experiment, they display related words (e.g., *fox*) instead of the original words (e.g., *fast*) to ensure that the user needs to grasp the meaning of the displayed word. In both experiments, they repeat the tasks with

[5] Typically, multiple fonts are used in a single document, and it is a common task to obtain a pair of fonts that both contrast and complement each other.

[6] http://fontmap.ideo.com/.

[7] https://datascopeanalytics.com/fontstellations/.

[8] https://github.com/Jack000/fontjoy.

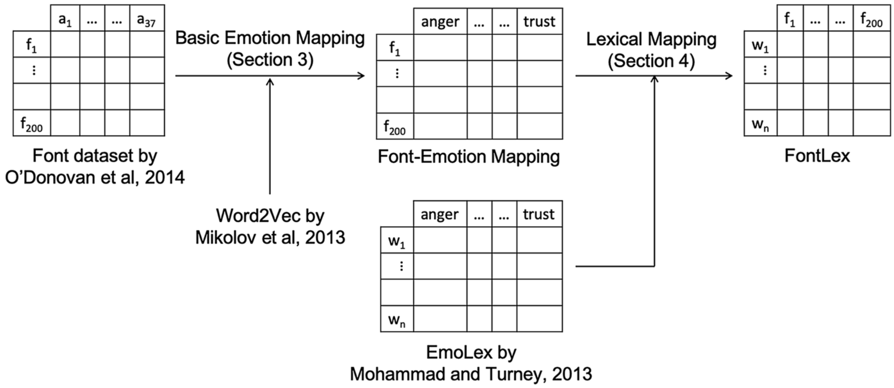

Fig. 1 Overview of our approach to obtain FontLex, where f_i are fonts, a_i are font attributes, and w_i are words

congruent and incongruent fonts, finding that the former significantly reduce the response time.

In terms of survey-style studies, Juni and Gross (2008) present newspaper articles using two different fonts. Their survey reveals that the same text is perceived as more humorous or angry when read in a certain font compared to another. Shaikh (2007b) presents documents to participants using congruent, incongruent, and neutral fonts, while soliciting ratings to assess the perception of the document (e.g., as *exciting*) as well as the perceived personality of the author (e.g., in terms of trustworthiness). The findings show strong effects across the assessed font types with respect to the perception of documents, whereas congruent and neutral fonts appear to evoke similar perceptions of an author's personality.

Shaikh et al. (2007) study the effect of the choice of font on email perception. Their results suggest that fonts with low congruency may result in different perceptions of an email than fonts with medium to high congruency. A similar study on the perception of a company website (Shaikh 2007a) demonstrates that neutral and low congruency fonts can negatively affect a company's perception in terms of professionalism, believability, trust, and intent to act on the site.

Many studies in marketing analyze font effects, especially in packaging design. For instance, Fligner (2013) shows that fonts associated with the attribute *natural* increase the perceived *healthfulness* of products when used in their packaging, particularly if the products' intrinsic cues (e.g., being fat-free) and extrinsic ones (e.g., being sold at Whole Foods Market) also concur. Childers and Jass (2002) establish that the semantic attributes of a font bear an impact on user perception for both high and low engagement levels. Through experiments using bottled water of a fictional brand, Van Rompay and Pruyn (2011) find additional evidence that the congruence between fonts and other design elements influence the perception of brand credibility, aesthetics, and value.

Table 1 Top three closest attributes for the basic emotions, where ¬ indicates attributes that are negated

Index	Emotion	1	2	3
1	Anger	¬Calm	Clumsy	Capitals
2	Anticipation	Fresh	Formal	Dramatic
3	Disgust	Clumsy	Bad	Sloppy
4	Fear	Bad	Capitals	¬Calm
5	Joy	Happy	Playful	Graceful
6	Negative	Bad	Strong	Sharp
7	Positive	Strong	¬Bad	Happy
8	Sadness	¬Happy	Gentle	¬Graceful
9	Surprise	Dramatic	Happy	¬Sharp
10	Trust	Strong	Calm	¬Bad

3 Basic emotion mapping

In this section, we describe the *Basic Emotion Mapping*, i.e., our method to obtain associations between fonts and emotional attributes. These will later, in the following section on the *Lexical Mapping* process, be used to obtain associations between fonts and words via their respective emotional associations. Figure 1 provides an overview of this process.

3.1 Method

Our method assumes as input a set of fonts $\mathcal{F}$ that are described in terms of a set of font attributes $\mathcal{A}$. For this, we rely on the crowdsourced data from (O'Donovan et al. 2014), which for a given font $f \in \mathcal{F}$ provides scores in [0, 100] for each attribute $a \in \mathcal{A}$. From this data, we derive $|\mathcal{F}|$-dimensional vectors $\mathbf{a} \in [0, 1]^{|\mathcal{F}|}$ for each font attribute $a \in \mathcal{A}$. For this, we simply transform the dataset to consider the fonts for a given font attribute, normalizing scores to [0, 1].

Then, to induce FontLex, we first generate $|\mathcal{F}|$-dimensional font vectors for a set of emotion attributes $\mathcal{E}$. Subsequently, using existing word–emotion associations, we will infer $|\mathcal{F}|$-dimensional font vectors for words such that each component of such a vector quantifies the strength of the association between a word and a font.

As the set of emotions $\mathcal{E}$, we consider the ten emotion attributes used in EmoLex (Mohammad and Turney 2013). Our first step is to map these $e \in \mathcal{E}$ to vectors $\mathbf{e} \in \mathbb{R}^{|\mathcal{F}|}$ that characterize their association with fonts $f \in \mathcal{F}$ in our data.

To achieve this, we proceed as follows. For each emotion $e \in \mathcal{E}$, we determine the $k = 3$ most similar font attributes $a \in \mathcal{A}$, as shown in Table 1. To decide on this value, we have carried out leave-one-out tests on the crowdsourced seed dataset (O'Donovan et al. 2014). Although the average overall success of the method in terms of the mean error was slightly higher for higher k than 3, we found that for $k = 3$ the most attributes attained their highest scores. Also considering the complexity of the negation decisions as will be described shortly, we opted to use the closest $k = 3$ neighbors.

We rely on word2vec (Mikolov et al. 2013) distances $d(e, a)$, using cosine distances on the standard word2vec Google News pretrained model,[9] to determine similarity scores $\mathrm{sim}(e, a)$ between emotion names and font attribute names as below:

$$\mathrm{sim}(e, a_i) = \frac{1}{k-1} \frac{\sum_{\substack{j=1 \\ i \neq j}}^{k} d(e, a_j)}{\sum_{j=1}^{k} d(e, a_j)} \tag{1}$$

One aspect that needs to be addressed, however, is the widely known fact that distributional models of semantics tend to conflate synonyms with antonyms. Hence, we first define

$$\boldsymbol{\mu}(e, a) = \begin{cases} \mathbf{1} - \mathbf{a} & \text{if } a \text{ is assessed as an antonym of } e \\ \mathbf{a} & \text{otherwise,} \end{cases} \tag{2}$$

where $\mathbf{1}$ is an $|\mathcal{F}|$-dimensional vector of ones. Thus, for those words that are assessed as antonyms, we do not use the regular font vector $\mathbf{a}$, but instead consider an inverted vector, in which we subtract each value from the maximum value of 1. The assessment is performed manually. For relationships such as between *anger* and *calm*, determining antonym relationships was straightforward. However, for some more challenging decisions, such as *negative* and *sharp*, we evaluated both options and discussed the obtained results with a graphic designer before making the final decision. In Table 1, attributes labelled as antonyms are marked with a "$\neg$" symbol.

To obtain font vectors $\mathbf{e}$ for emotions $e \in \mathcal{E}$, we compute

$$\mathbf{e} = \sum_{i=1}^{k} \mathrm{sim}(e, a_i)\, \boldsymbol{\mu}(e, a_i) \tag{3}$$

where the a_i are the k most similar attributes, as described above. Thus, the font vectors are a weighted average of the vectors for related attributes, after possibly inverting their respective vectors.

3.2 Results

Figure 2 depicts the top 3 fonts that are most strongly associated with the ten emotion attributes, whereas Fig. 3 shows sample fonts that are predicted to be neutral in terms of the respective emotion. Figure 4 shows the three fonts for each emotion that are found to have the weakest associations. More specifically, the neutral fonts for emotion e are defined as those that are in the middle of the ranked font vector $\mathbf{e}^{\mathbf{r}}$ of size n, namely $\mathbf{e}^{\mathbf{r}}_i$ for $i \in \{\lfloor \frac{n}{2} \rfloor - 1, \lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor + 1\}$, where $\mathbf{e}^{\mathbf{r}}_1$ has the strongest association with the emotion, and $\mathbf{e}^{\mathbf{r}}_n$ has the weakest association. In all figures, the emotion names are rendered using the corresponding fonts.

[9] https://code.google.com/archive/p/word2vec/.

anger	anticipation	disgust	fear	joy	NEGATIVE	positive	sadness	surprise	trust
ANGER	anticipation	disgust	FEAR	joy	NEGATIVE	positive	sadness	surprise	trust
anger	anticipation	disgust	FEAR	joy	negative	positive	sadness	surprise	trust

Fig. 2 Emotion attributes rendered using the three most congruent fonts as predicted by our method. The renderings on the first line uses the fonts ranked 1st, the second line uses fonts ranked 2nd, and the third line uses fonts ranked 3rd

anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust
anger	anticipation	disgust	fear	JOY	negative	positive	sadness	surprise	trust
anger	anticipation	disgust	fear	JOY	negative	positive	sadness	surprise	trust

Fig. 3 Emotion attributes rendered using the neutral fonts as predicted by our method. The renderings on the first line use the fonts ranked 99th, the second line uses fonts ranked 100th, and the third line uses fonts ranked 101st

anger	ANTICIPATION	disgust	fear	joy	negative	positive	sadness	surprise	trust
anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	TRUST
anger	anticipation	disgust	fear	joy	negative	POSITIVE	sadness	surprise	trust

Fig. 4 Emotion attributes using the three most incongruent fonts as predicted by our method. The renderings on the first line use the fonts ranked 198th, the second line uses fonts ranked 199th, and the third line uses fonts ranked 200th

The fonts that are strongly associated with emotions share some special characteristics. For instance, for *joy*, we encounter handwriting-style typefaces, whereas for *disgust*, we find display fonts with salient stylization. It should also be noted that not all fonts that share these characteristics are strongly associated with these emotions, since the relationships between emotion attributes and font characteristics are not straightforward (Kulahcioglu and de Melo 2018).

3.3 Evaluation

To assess the quality of the obtained emotion font score predictions, we carry out a user study.

3.3.1 User study

For each of the ten emotion attributes, we generated four tasks with different random font choices. An example is given in Fig. 5. Each task includes 5 fonts, two congruent fonts selected randomly among the top-scoring 10 fonts for that emotion, two incongruent fonts selected randomly among the lowest-scoring 10 fonts for that

Fig. 5 An example task for *positive*. The second and fifth fonts are congruent, the third and fourth is incongruent and the first is neutral

Table 2 Evaluation results (in %) for emotions

Expected value	Congruent 40.00	Neutral 20.00	Incongruent 40.00
Anger	74.04	14.42	11.54
Anticipation	28.85	34.62	36.54
Disgust	70.19	10.58	19.23
Fear	78.85	5.77	15.38
Joy	91.35	4.81	3.85
Negative	59.80	15.69	24.51
Positive	60.00	23.81	16.19
Sadness	46.15	16.35	37.50
Surprise	72.12	8.65	19.23
Trust	62.50	20.19	17.31
Average	64.38	15.49	20.13

emotion, and one neutral font selected randomly among the ten fonts that are in the middle of the ranked list of fonts. In each task, the user is requested to select a single image that best reflects the semantics of the word. As described above, the available options include the same word presented using five different fonts.

Each task is carried out by 30 participants via Mechanical Turk, all from the United States, with at least 5000 approved hits and an overall approval rating of 97% or more. We used counterbalancing, i.e., half of the users received the tasks in the reverse order from the other half. We also used three validation tasks, and eliminated results of three participants who incorrectly answered all three of them.

3.3.2 Evaluation results

Table 2 summarizes the results of this user study. The *congruent* column lists the percentages of selections in which the congruent fonts (those in the top 10 for that word) are preferred. Similarly, the *neutral* and *incongruent* columns list the percentages of choices of neutral and incongruent fonts, respectively. The first row lists the expected value assuming the null hypothesis of a uniform distribution over the five choices, of which 2 are congruent, 1 neutral, and 2 incongruent.

The average is 64.38% for congruent font preferences. Compared to the expected value of 40%, this shows a strong trend toward the fonts predicted to be congruent,

Fig. 6 Selected words rendered using the three most congruent fonts as predicted by our method. The renderings on the first line uses the fonts ranked 1st, the second line uses fonts ranked 2nd and the third line uses fonts ranked 3rd

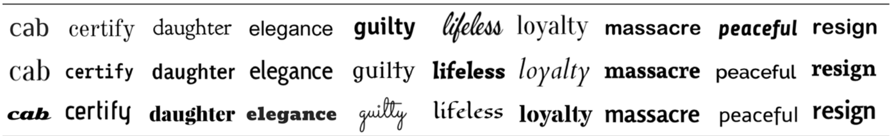

Fig. 7 Selected words rendered using the three fonts from the middle of the ranked list as predicted by our method. The renderings on the first line uses the fonts ranked 99th, the second line uses fonts ranked 100th and the third line uses fonts ranked 101st

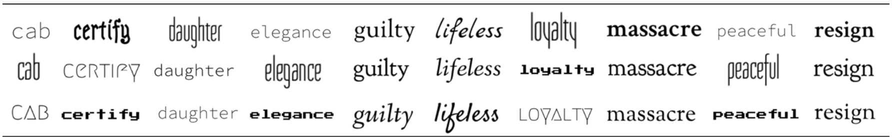

Fig. 8 Selected words rendered using the three most incongruent fonts as predicted by our method. The renderings on the first line uses the fonts ranked 198th, the second line uses fonts ranked 199th and the third line uses fonts ranked 200t

hence validating our results in general. Similarly, the preferences for the fonts that are found to be incongruent by our method was much lower than the expected value, with an average of only 20.13%.

However, a detailed look at the values for individual emotion attributes reveal that the performance differs between them. The strongest preference is obtained for *joy*, with a value of 91.35%, whereas the lowest is for *anticipation* with 28.85%. Another comparably low value is obtained for *sadness*, with a congruency of 46.15%. This suggests that different emotions may differ in how saliently and uniquely they are associated with visual font characteristics (cf. Sect. 7).

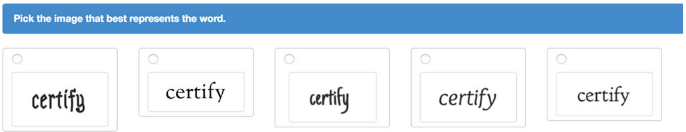

Fig. 9 An example task for the word *certify*. The second and fifth fonts are congruent, the first and third is incongruent, and the fourth is neutral

4 Lexical mapping

The next phase involves computing font vectors for words that reflect the degree of association between words and potential fonts. As shown earlier in Fig. 1, we rely on the results of the Basic Emotion Mapping from Sect. 3 as our input, along with data from a word–emotion lexicon, to induce our FontLex resource.

4.1 Method

EmoLex (Mohammad and Turney 2013) provides binary emotion association indicators between words and the emotion attributes $e \in \mathcal{E}$ listed in Table 1. There are 6468 words with at least one emotion association in their data. For words w in this set, we consider their data as providing vectors $\mathbf{w}_{\mathrm{E}} \in [0,1]^{|\mathcal{E}|}$.

To generate a font vector $\mathbf{w}_{\mathrm{F}}$ for a word w, we compute

$$\mathbf{w}_{\mathrm{F}} = \frac{1}{\|\mathbf{w}_{\mathrm{E}}\|_1} \mathbf{M}_{\mathrm{E}} \mathbf{w}_{\mathrm{E}} \tag{4}$$

where $\|\mathbf{w}_{\mathrm{E}}\|_1$ denotes the ℓ_1 norm of $\mathbf{w}_{\mathrm{E}}$ and $\mathbf{M}_{\mathrm{E}} = [\mathbf{e}_1 \ldots \mathbf{e}_{|E|}]$, i.e., a matrix with columns that capture the font vectors for the emotions $e \in \mathcal{E}$ (in the same order as captured in $\mathbf{w}_{\mathrm{E}}$).

4.2 Results

Figure 6 shows the top three congruent fonts associated with ten sample words, Fig. 7 shows sample fonts that are predicted to be neutral for the respective words, and Fig. 8 shows the most incongruent three fonts for the same words. In all images, the words are rendered using the corresponding fonts. These words are among those used in the evaluation user study in the following section.

4.3 Evaluation

We evaluate the dataset through a user study. In the following, we provide details on the design and the results of this study.

Table 3 Evaluation results (in %) and emotion associations for the words in the user study

Word	C 40.00	N 20.00	I 40.00	Corresponding emotions by EmoLex
Appreciation	70.59	15.69	13.73	Joy, positive, trust
Cab	53.85	11.54	34.62	Positive
Certify	79.59	6.12	14.29	Trust
Conformance	61.54	19.23	19.23	positive
Congenial	42.86	20.41	36.73	Positive
Daughter	70.59	13.73	15.69	Joy, positive
Elegance	76.00	16.00	8.00	Anticipation, joy, positive, trust
Guilty	49.02	21.57	29.41	Anger, negative, sadness
Instruct	55.77	28.85	15.38	Positive, trust
Kill	75.00	5.77	19.23	Fear, negative, sadness
Lifeless	32.00	26.00	42.00	Fear, negative, sadness
Loyalty	76.92	9.62	13.46	Positive, trust
Massacre	56.00	16.00	28.00	Anger, disgust, fear, negative, sadness
Medley	40.38	36.54	23.08	Positive
Murky	82.35	5.88	11.76	Disgust, negative, sadness
Noble	72.55	15.69	11.76	Positive, trust
Oracle	52.00	16.00	32.00	Anticipation, positive, trust
Outcome	64.71	17.65	17.65	Positive
Peaceful	64.00	12.00	24.00	Anticipation, joy, positive, surprise, trust
Persistent	65.38	9.62	25.00	Positive
Precedence	56.86	15.69	27.45	Positive, trust
Resign	20.00	18.00	62.00	Anger, disgust, fear, negative, sadness
Shameful	50.00	28.85	21.15	Negative, sadness
Tickle	63.46	9.62	26.92	Anticipation, joy, positive, surprise, trust
Verified	64.71	23.53	11.76	Positive, trust
Average	59.85	16.78	23.37	

C congruent, *N* neutral, *I* incongruent

4.3.1 User study

For our study, we consider 25 words randomly selected from the set of words with at least one salient font association. For this purpose, we consider any of the 3882 words that have a score of 0.75 or higher in any of the components of their respective font vectors. For each of the random 25 words, we generated two tasks with different random font choices. We have reduced the number of tasks to two, compared to the four tasks used in the previous section, to keep the total number of tasks reasonable for each participant.

An example task for the word *certify* is given in Fig. 9. Each task includes 5 fonts, two congruent fonts selected randomly among the top-scoring 5 fonts for that

anticipation	+	joy	→	optimisim	surprise	+	sadness	→	DISAPPROVAL
anticipation	+	fear	→	ANXIETY	surprise	+	disgust	→	unbelief
anticipation	+	trust	→	hope	surprise	+	anger	→	OUTRAGE
joy	+	trust	→	love	sadness	+	disgust	→	remorse
joy	+	fear	→	GUILT	sadness	+	anger	→	ENVY
joy	+	surprise	→	delight	sadness	+	anticipation	→	pessimism
trust	+	fear	→	SUBMISSION	disgust	+	anger	→	contempt
trust	+	surprise	→	curiosity	disgust	+	anticipation	→	cynic
trust	+	sadness	→	sentimentality	disgust	+	joy	→	morbid
fear	+	surprise	→	AWE	anger	+	anticipation	→	aggressive
fear	+	sadness	→	DESPAIR	anger	+	joy	→	PRIDE
fear	+	disgust	→	shame	anger	+	trust	→	DOMINANCE

Fig. 10 Emotion combinations (dyads) from Plutchik's Wheel of Emotion, rendered using the congruent fonts as determined by our study

word, two incongruent fonts selected randomly among the lowest-scoring 5 fonts for that word, and one neutral font selected randomly among the three fonts that are in the middle of the ranked list of fonts for the word. The decision to use 5 fonts as opposed to 10 is again based on considerations regarding the workload per user.

Each task involves a user being requested to select the image that best represents the word. As described above, the available options include the same word presented using five different fonts. Each task is carried out by 30 participants in Mechanical Turk, all from the United States, with at least 5000 approved hits and an overall approval rating of 97% or more. We used counterbalancing and eliminated results of one participant that accidentally completed both of the original and reversed task sessions. We have also used three validation tasks, and eliminated results of one participant that incorrectly answered both of the two validation tasks.

4.3.2 Evaluation results

Table 3 summarizes the evaluation results for the 25 randomly selected words as described above. The *congruent* column lists the percentages of selections in which the congruent fonts (those in the top 5 for that word) are preferred. Similarly, the

Table 4 Basic emotion and sentiment associations of complex emotions as suggested by EmoLex

	Basic emotions	Sentiment
Aggressive	Anger, fear	Negative
Anxiety	Anger, anticipation, fear, sadness	Negative
Awe	–	–
Contempt	Anger, disgust, fear	Negative
Curiosity	Anticipation, surprise	Positive
Cynic	–	–
Delight	Anticipation, joy	Positive
Despair	Anger, disgust, fear, sadness	Negative
Disapproval	Sadness	Negative
Dominance	–	–
Envious	–	Negative
Guilt	Disgust, sadness	Negative
Hope	Anticipation, joy, surprise, trust	Positive
Love	Joy	Positive
Morbid	Sadness	Negative
Optimisim	Anticipation, joy, surprise, trust	Positive
Outrage	Anger, disgust	Negative
Pessimism	Anger, fear, sadness	Negative
Pride	Joy	Positive
Remorse	Sadness	Negative
Sentimentality	–	Positive
Shame	Disgust, fear, sadness	Negative
Submission	–	–
Unbelief	–	Negative

neutral and *incongruent* columns list the percentages of choices of neutral and incongruent fonts, respectively.

The average is 59.85% for congruent font preferences, which shows that the consensus between our data and the users were strong. The strongest preference is obtained for the word *murky*, with a value of 82.35%, whereas the lowest is for the word *resign* with 20.00%. Similarly, the average for the incongruent preferences was only 23.37%, bearing further witness to the quality of the results. Only two out of twenty-five words, namely *lifeless* and *resign*, received congruent preferences that are less than the expected value of 40%. Such results are expected, given that different words may differ in the strength and uniqueness of their associations (cf. Sect. 7).

Table 3 also displays the corresponding emotions for the words used in the evaluation, allowing us to analyze the relationship between the success of the two datasets. In some cases, words associated with the same set of emotions obtained similar user ratings, such as *instruct*, *noble*, *precedence*, and *verified*. Whereas in some cases, words with the same emotion set obtained quite divergent ratings: *massacre* and *resign*.

boring	DOMINANCE	joy	WIDE	sharp
deadening	ascendance	pleasure	broad	crisp
slow	say-so	gladden	extensive	acute
irksome	AUTHORITY	rejoice	FULL	needlelike
dull	CONTROL	joyousness	spacious	knife

Fig. 11 Examples of synonyms retrieved form WordNet for the attributes *boring*, *dominance*, *joy*, *wide* and *sharp*; rendered using congruent fonts

5 Extensions to FontLex

In this section, we present methods to extend the dataset and increase its accuracy.

5.1 Extension 1: complex emotion mapping

We propose using Plutchik's Wheel of Emotion (Plutchik 2001) to infer font scores for complex emotions (e.g., hope). Plutchik's theory suggests that complex emotions are indeed combinations of basic ones, referred to as *dyads*. For instance, the theory posits that *hope* is a superposition of the more basic emotions *anticipation* and *trust*. Relying on the font vectors **e** obtained in Sect. 3, we compute font vectors **c** for complex emotions $c \in \mathcal{C}$ as

$$\mathbf{c} = \frac{1}{2}(\mathbf{e}_i + \mathbf{e}_j),$$

where the e_i and e_j are the underlying basic emotions for c, and i and j are their indices from Table 1.

Figure 10 provides dyads for all $c \in \mathcal{C}$, while rendering the words for each basic and complex emotion with the most congruent font as inferred by our study. As an example, the first entry suggests that *anticipation* and *joy* together evoke the feeling of *optimism*. The font determined as most congruent for *optimism* appears to combine visual characteristics of both its underlying basic emotions, namely *anticipation* and *joy*.

The obtained font scores for complex emotions serve two purposes. The first is that, similar to the basic emotions, they could be used as seed information to infer higher-quality font vectors for arbitrary words. A second purpose is to override font vectors for the complex emotion words in FontLex, potentially improving its accuracy.

To explore this more, we assess how the complex emotions described by Plutchik's Theory are annotated in EmoLex (Mohammad and Turney 2013). For these complex emotion words, Table 4 lists the corresponding basic emotions as given by EmoLex. The most notable problem is that 7 of the 24 complex emotions are not associated with any basic emotions. This might stem from issues of ambiguity, e.g., for *submission*. However, 3 of these 7 words with missing emotions are actually assigned a sentiment, which reduces the likelihood of such issues for these entries. Indeed, none of these associations are exact matches to their corresponding dyads. In total, 11 complex emotions have emotions assigned but miss at least one of the dyad emotions (e.g., love, guilt). Out of these, five complex emotions have only one basic emotion associated. In addition, 6 complex emotions have the two emotions from the dyad defined by Plutchik, but also additional ones (e.g., optimism, hope).

Overall, we conclude that EmoLex is incomplete in its description of complex emotion words, and that relying on Plutchik's theory can yield better font associations.

5.2 Extension 2: semantic relationships

We extend the dataset and increase its accuracy by accounting for semantic relationships given by WordNet (Fellbaum 1998). For all attribute words in $\mathcal{A} \cup \mathcal{E} \cup \mathcal{C}$, in total 71 attributes (37 original font attributes from (O'Donovan et al. 2014), 10 basic emotion attributes as computed in Sect. 3, and 24 complex emotion attributes as computed in Sect. 5.1, we gather synonyms. For the original font attributes, we gather the set of words that share a common synset with the attribute names (such as the words *deadening*, *dull*, *ho-hum*, *irksome*, *slow*, *tedious*, *tiresome* and *wearisome* for the font attribute *boring*). We then go through this list manually to exclude any synonyms with an irrelevant meaning (such as the word *building complex* for the font attribute *complex*). For the basic and complex emotion attributes, we pick the sense describing an emotion, and then use the synonyms from these synsets. These synonyms are assigned the font vectors of the corresponding words in $\mathcal{A} \cup \mathcal{E} \cup \mathcal{C}$. This results in 464 additional word-font assignments, 166 of which override the ones from the methods in Sect.s 3 and 4 While small in number, these provide for particularly salient associations (examples provided in Fig. 11).

5.3 Extension 3: more fonts

Our study relies on the data from (O'Donovan et al. 2014), which connects 200 fonts with 37 attributes. In the previous sections, we extend its attribute set and connect it with the words from EmoLex, keeping the font set the same. We now proceed to extend our lexicon to use further fonts following the method proposed by Kulahcioglu and de Melo (2018).

Our goal is to predict font vectors $\mathbf{f}'$ for fonts $f' \notin \mathcal{F}$. To achieve this, we use weighted k-nearest neighbors (k-NN) regression using 4 neighbors. The weighted k-NN approach generates weights using the following equation (Kulahcioglu and de Melo 2018).

$$w_i = \frac{1}{3} \frac{\sum_{j=1,\, i \neq j}^{4} d(f', f_j)}{\sum_{j=1}^{4} d(f', f_j)} \tag{5}$$

The distance between two fonts, denoted as $d(f_i, f_j)$ is computed using Convolutional Neural Network (CNN) embeddings from Qiao (2017). For each font, an

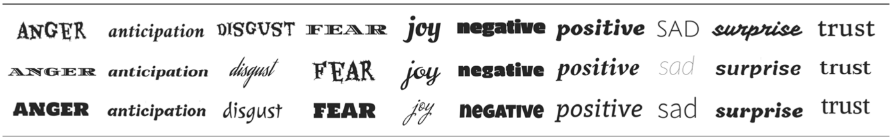

Fig. 12 Basic emotions rendered using the three most congruent fonts from the extended font set (excluding 200 fonts from the original dataset) as predicted by our method

Fig. 13 A word cloud of complex emotions rendered using fonts from the extended font set (excluding 200 fonts from the original dataset) that are inferred to be congruent by our method

image is generated rendering the letters (L, a, s, e, g, d, h, u, m, H, l, o, i, v) on a grid. These images are processed by the CNN and the obtained representations can be regarded as visual font embeddings. The visual distance between two fonts can then be computed as the Euclidean distance of their visual font embeddings.

Subsequently, the weighted values are generated as follows:

$$\mathbf{f} = \sum_{i=1}^{4} w_i \mathbf{f}_i \tag{6}$$

Using the above approach and aforementioned embeddings, we extend our dataset from 200 to 1922 fonts, while each font vector include scores for every word in $\mathcal{A} \cup \mathcal{E} \cup \mathcal{C}$. Figure 12 presents basic emotions using the most congruent three attributes from the extended dataset. We exclude the fonts from the original dataset to be able to portray results from just the extension.

Finally, the word cloud in Fig. 13 provides complex emotions rendered with corresponding high-scoring fonts from the extended font set.

6 Application example

In this section, we introduce a proof of concept *Poster Design* application for which FontLex could prove useful. In this example, the tool provides two types of support. In the first scenario, the tool recommends a font for a poster based on the words it includes. We compute a font vector $\mathbf{p}$ for a poster P as follows:

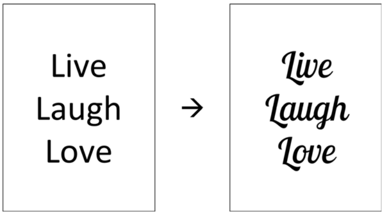

Fig. 14 A poster design example where the congruent font recommendations for the poster is generated using FontLex

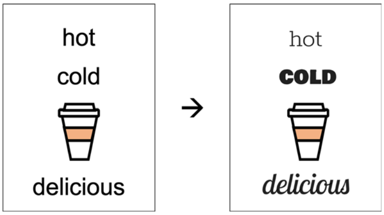

Fig. 15 A poster design example where a congruent font is individually recommended for each word using FontLex

$$\mathbf{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \tag{7}$$

where $\mathbf{w}_i$ are the font vectors for the words w_i in poster P. We omit the words for which no font vector is found in FontLex. A sample is provided in Fig. 14. In this example, the image on the left shows the poster with the default font, whereas the image on the right makes use of the recommended font (the font with the highest score in $\mathbf{p}$).

In the second scenario, each word is assigned a different font. For proof of concept purposes, we focus on semantic congruence, and ignore other important design concerns such as the harmony of different fonts. For a word w_i in P, the fonts with the highest values in $\mathbf{w}_i$ are considered as candidate fonts and one of them is selected randomly. Figure 15 provides an example, in which different fonts are assigned to the words in the poster. The poster on the left uses the default font, whereas the poster on the right makes use of the recommended fonts.

7 Discussion

In this section, we discuss our results and potential applications of FontLex and of dyads as a means of inferring complex emotions.

7.1 Results

We have introduced two datasets that connect emotions and words with fonts in terms of real-valued scores. Besides showing strong support for the datasets, the user evaluations also revealed that the performance varies for different emotions and words. Below, we discuss the potential sources for these differences.

For the emotion–font dataset, one reason for the differences between results could be the varying potential of fonts to represent or evoke different emotions (Kulahcioglu and de Melo 2018). This could be observed in the results for *anticipation*, for which determining a font type may prove difficult even for an experienced graphic designer. It is also observed that emotions with higher arousal, namely *anger*, *disgust*, *fear*, *joy*, and *surprise*, received higher congruent user preferences compared to other emotions, which may be a direction that merits further analysis.

The second reason may be a lack of appropriate similar attributes in the crowdsourced seed dataset. Looking at Table 1, it could be argued that *joy* has semantically close neighbors in the dataset, whereas this is not the case for *anticipation*.

For the word–font dataset, assessing the underlying emotion connections in Table 3 may shed some light on the differences. Recalling that the lowest performing emotion–font scores are for *anticipation* and *sadness*, one might expect that words associated with these emotions are prone to showing fewer user preferences that are congruent. The words associated with *anticipation*, namely *elegance*, *oracle*, *peaceful*, and *tickle*, do not seem to possess the same difficulty, as the lowest preference for these words is 52% (for *outcome*), which shows a strong preference.

On the other hand, among the words associated with *sadness*, the words *lifeless* and *resign* do not show such strong preferences. One might conjecture that this stems from low-performing emotion–font associations. However, looking at this in more detail, we find that *kill* and *massacre* have the same underlying emotion associations as *lifeless* and *resign*, respectively. The fact that the fonts for *kill* and *massacre* received strong support from users suggests that the word–emotion associations might have played a role. Some words may have inaccurate or missing emotion associations, while other words may have weaker emotional associations than others, which is not reflected in the binary scheme used by EmoLex. Using a dataset with real-valued scores instead of binary associations might help to capture the latter case.

Fortunately, overall, both datasets have received strong support from users, with around 60% and 64% of the average user preferences towards the fonts found to be congruent by our datasets. Only for two words out of 25, incongruent fonts are preferred more frequently than chance would predict, i.e., $\frac{2}{5} = 40\%$. In contrast, for 23 words, congruent fonts are preferred more frequently than chance would predict. Despite the subjective nature of font preferences and associations, we observe that there is a clear correspondence between the fonts chosen by our method and those assessed as appropriate by the human participants.

7.2 Application areas

The main use cases we foresee for FontLex are font search and font recommendation.

Semantic Font Search. Currently, content creation tools that heavily rely on text, such as word processors or graphic design tools, use traditional search methods for font search. O'Donovan et al. (2014) propose semantic attribute based font search as a step towards sufficient user support. We believe FontLex can help taking semantic search one step further by providing search using any keyword instead of a predefined small set of attributes. This could help users make use of a large number of fonts which is otherwise hard to achieve. Its flexibility would also allow users to be more creative.

Font recommendation. We are not aware of any applications providing semantics-aware font recommendation support. We believe font–emotion mappings and FontLex could be utilized to enable such recommendation, and we demonstrate such usage in Sect. 6. In addition to the support described in the example, FontLex could be utilized to provide more advanced support using some of its attributes (e.g., *legible* for readability, *artistic* for aesthetics) as filtering options. For instance, in our poster design application example, fonts could be filtered to pick only the *display* ones. Recently, we proposed a word cloud tool that provides typographical recommendations based on user input determining the intended affect of the word cloud (Kulahcioglu and de Melo 2019).

7.3 Emotion combinations (Dyads)

In this study, we use combinations of basic emotions to calculate scores for the complex emotions based on the dyads provided by (Plutchik 2001). Based on our qualitative analysis, it is a powerful method to infer complex emotions, which is otherwise a challenging task. To the best of our knowledge, dyads have not been utilized before to infer complex emotions of content. Thus, a similar approach could be applied to other domains, such as for text and image.

7.4 Personal and demographic differences

Fonts are strongly tied to cultural elements, and hence may bear associations with various concepts, such as historical epochs, brands, or even music genres. Although we do not explicitly explore these connections, we believe the seed dataset that we rely on (O'Donovan et al. 2014) accounts for such associations implicity, as it is a crowdsourced dataset and the emotional ratings that users provide are affected by such connections. At the same time, there are also differences between users, especially based on their demographics, such as culture, gender, and age. These personal and demographic differences of font semantics remain to be explored.

8 Conclusions and future work

Currently, no existing tool or resource provides a comprehensive semantic font recommendation support, in which the meaning of the text is computationally matched with the semantic attributes of the fonts. Our study aims to support the development of such font recommendation tools.

Following this aim, we have created FontLex , a dataset that maps 6.7K words to 1922 fonts. These derive mainly from the affective associations between words and fonts. As part of the future work, we plan to further expand the dataset by making use of font attributes such as *thin*, *wide*, and *angular*, and their connections with objects, as opposed to the more abstract focus in this paper.

References

Bloch, M. (2017). 1001 Fonts. Retrived 29 July 2017 from, http://www.1001fonts.com.

Canvacom. (2017). Canva font combinations. Retrived 29 July 2017 from, https://www.canva.com/font-combinations/.

Childers, T. L., & Jass, J. (2002). All dressed up with something to say: Effects of typeface semantic associations on brand perceptions and consumer memory. *Journal of Consumer Psychology*, *12*(2), 93–106.

dafontcom. (2017). Retrived 29 July 2017 from, http://www.dafont.com.

Data Scope Analytics (2017). Retrived 29 July 2017 from, Fontstellations. https://datascopeanalytics.com/blog/fontstellations/.

Fellbaum, C. (1998). Wordnet: an eletronic lexical database. cambridge, massachusetts, eua.

Fligner, A. (2013). The effect of packaging typeface on product perception and evaluation. PhD thesis, The Ohio State University.

Fox, D. (2010). Free will for typeface selection: Myth or reality? *Usability News*, *12*, 1.

Hazlett, R. L., Larson, K., Shaikh, A. D., & Chaparo, B. S. (2013). Two studies on how a typeface congruent with content can enhance onscreen communication. *Information Design Journal*, *20*, 3.

Ho, K. (2017). Organizing the world of fonts with ai. Retrived 29 July 2017 from,https://medium.com/@kho98.

Juni, S., & Gross, J. S. (2008). Emotional and persuasive perception of fonts. *Perceptual and Motor Skills*, *106*(1), 35–42.

Kulahcioglu, T., & De Melo, G. (2018). Fontlex: A typographical lexicon based on affective associations. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).

Kulahcioglu, T., & de Melo, G. (2018). Predicting semantic signatures of fonts. In: Proceedings of the 12th IEEE International Conference on Semantic Computing, IEEE.

Kulahcioglu, T., & de Melo, G. (2019). Paralinguistic recommendations for affective word clouds. In: Proceedings of the 24th International Conference on Intelligent User Interfaces, ACM, New York, NY, USA, IUI '19, pp 132–143, https://doi.org/10.1145/3301275.3302327

Lewis, C., & Walker, P. (1989). Typographic influences on reading. *British Journal of Psychology*, *80*(2), 241–257.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119.

Mills, H. (2017). Font pair. Retrived 29 July 2017 from,http://fontpair.co/.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *29*(3), 436–465.

O'Donovan, P., Libeks, J., Agarwala, A., & Hertzmann, A. (2014). Exploratory font selection using crowdsourced attributes. *ACM Transaction of Graph*, *33*(4), 92:1–92:9.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*(4), 344–350.

Qiao, J. (2017). Fontjoy—font pairing made simple. Retrived 29 July 2017 from, http://fontjoy.com/.

Sam Berlow, S. C., & Sherman, N. (2017). Fonts in use. Retrived 29 July 2017 from, https://fontsinuse.com/.

Shaikh, A. D. (2007a). The effect of website typeface appropriateness on the perception of a company's ethos. *Usability News*, *9*, 2.

Shaikh, A. D. (2007b). Psychology of onscreen type: Investigations regarding typeface personality, appropriateness, and impact on document perception. PhD thesis.

Shaikh, A. D., Chaparro, B. S., & Fox, D. (2006). Perception of fonts: Perceived personality traits and uses. *Usability News*, *8*(1), 1–6.

Shaikh, A. D., Fox, D., & Chaparro, B. S. (2007). The effect of typeface on the perception of email. *Usability News*, *9*(1), 1–7.

Tsonos, D., & Kouroupetroglou, G. (2016). Prosodic mapping of text font based on the dimensional theory of emotions: a case study on style and size. *EURASIP Journal on Audio, Speech, and Music Processing*, *1*, 8. https://doi.org/10.1186/s13636-016-0087-8.

Van Rompay, T. J., & Pruyn, A. T. (2011). When visual product features speak the same language: Effects of shape-typeface congruence on brand perception and price expectations. *Journal of Product Innovation Management*, *28*(4), 599–610.

Velasco, C., Salgado-Montejo, A., Marmolejo-Ramos, F., & Spence, C. (2014). Predictive packaging design: Tasting shapes, typefaces, names, and sounds. *Food Quality and Preference*, *34*, 88–95.

Velasco, C., Woods, A. T., Hyndman, S., & Spence, C. (2015). The taste of typeface. *I-Perception*, *6*, 4.

Wang, Z., Yang, J., Jin, H., Shechtman, E., Agarwala, A., Brandt, J., et al. (2015). Deepfont: Identify your font from an image. *Proc* (pp. 451–459). ACM Multimedia: ACM.