Leveraging End-User Data for Enhanced Design Concept Evaluation: A Multimodal Deep Regression Model

Chenxi Yuan
Department of Mechanical and Industrial Engineering,
Northeastern University,
Boston, MA 02115
e-mail: yuan.chenx@northeastern.edu

Tucker Marion
D'Amore-McKim School of Business,
Northeastern University,
Boston, MA 02115
e-mail: t.marion@northeastern.edu

Mohsen Moghaddam[1]
Department of Mechanical and Industrial Engineering,
Northeastern University,
Boston, MA 02115
e-mail: m.moghaddam@northeastern.edu

Design concept evaluation is a key process in the new product development process with a significant impact on the product's success and total cost over its life cycle. This paper is motivated by two limitations of the state-of-the-art in concept evaluation: (1) the amount and diversity of user feedback and insights utilized by existing concept evaluation methods such as quality function deployment are limited. (2) Subjective concept evaluation methods require significant manual effort which in turn may limit the number of concepts considered for evaluation. A deep multimodal design evaluation (DMDE) model is proposed in this paper to bridge these gaps by providing designers with an accurate and scalable prediction of new concepts' overall and attribute-level desirability based on large-scale user reviews on existing designs. The attribute-level sentiment intensities of users are first extracted and aggregated from online reviews. A multimodal deep regression model is then developed to predict the overall and attribute-level sentiment values based on the features extracted from orthographic product images via a fine-tuned ResNet-50 model and from product descriptions via a fine-tuned bidirectional encoder representations from transformer model and aggregated using a novel self-attention-based fusion model. The DMDE model adds a data-driven, user-centered loop within the concept development process to better inform the concept evaluation process. Numerical experiments on a large dataset from an online footwear store indicate a promising performance by the DMDE model with 0.001 MSE loss and over 99.1% accuracy. [DOI: 10.1115/1.4052366]

Keywords: design automation, new product development, design evaluation, image processing, natural language processing, deep regression

1 Introduction

Innovative design processes in early-stage product development typically involve generating and evaluating numerous alternative concepts through a variety of methods and heuristics [1]. This is a crucial requirement for a successful design process for several reasons, such as increasing the quantity of generated concepts—following Osborn's rules for brainstorming [2]—to inspire the designer's exploration and creativity [3–5], preventing the designer's fixation on a few ideas by exposing her to various concepts [6–8], and enhancing the quality of design by incorporating more creative ideas and concepts in the ideation and prototyping processes [9–11]. Evidence suggests that early-stage concept generation contributes to only about 8% of product development costs. In comparison, the decisions made during this phase can determine up to 70% of the total cost over the entire product life cycle [12]. Nevertheless, although generating a large number of novel concepts is necessary for successful innovative design in new product development processes [13,14], it is not sufficient without rigorous evaluation against a set of performance metrics that reflect users' needs.

Evaluating concepts is often difficult due to imprecise, incomplete, or subjective data [15]. As such, much research has been undertaken to develop methods to better inform the evaluation and ultimate selection of promising design concepts [16–18]. Design concept evaluation—a process in which the design team evaluates alternative design concepts and refines/narrows down a set of concepts based on their anticipated success—is therefore an essential step to take once a multitude of design concepts is generated [19]. Various normative decision-making tools and methods have been proposed in the literature for design concept evaluation. Examples include concept selection [16], concept screening [20], pairwise comparison charts [21], concept scoring matrices [22], multiattribute utility analysis [23], decision matrices [24], utility function analysis [25], fuzzy sets [26], and analytic hierarchy process (AHP) [27]. Fuzzy sets and AHP are proven effective for strategic [28] and multi-criteria [29] decision-making in design concept evaluation processes. A systematic decision process via the fuzzy sets method was presented in Ref. [30] for identifying and choosing the best design concept based on expert knowledge combined with optimization-based methodology. Integrated fuzzy sets with genetic algorithms and neural networks were proposed in Ref. [18] for obtaining an optimal concept from a group of satisfactory concepts. Furthermore, an AHP-based method combined with fuzzy set theory was presented in Ref. [24] for evaluating the alternatives of conceptual design through a score-ranking mechanism. In a related study [31], an analytic network process, a more generic form of AHP, was used to determine the most satisfactory conceptual design by considering the variety of interactions and dependencies between higher and lower level elements. A different evaluation process based on fuzzy reasoning and neural networks was discussed in Ref. [32] for evaluating design concepts based on a set of user requirements. A detailed review of the state-of-the-art in design concept evaluation is provided in Sec. 2.

Systematic and technology-focused methodologies for evaluating concepts and informing their selection have been a fertile and impactful area of research in design methods and tools for over two decades. However, irrespective of the systematic methodology used, a significant factor in developing and evaluating concepts is

[1]Corresponding author.

Contributed by the Design Theory and Methodology Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received June 29, 2021; final manuscript received September 3, 2021; published online September 21, 2021. Assoc. Editor: Christopher McComb.

the users' insights [33]. Eliciting insights from users in the upfront of the development process has been long established as a best practice in new product development [34–37]. Multi-criteria decision-making (MCDM) methods that integrate user feedback into concept evaluation and selection include quality function deployment (QFD) [38] and experiments using data envelopment analysis [39]. Without proper user validation, establishing design prototypes and allocating resources to those potential products may lead to excessive costs in later stages due to revisions or potential failures [40].

This paper is motivated by a lack of rigorous data-driven methods for *user-centered* evaluation of design concepts, which can better inform the design team's concept evaluation and selection process. The state-of-the-art in design concept evaluation predominantly relies on the judgment and expertise of the design team—either through subjective concept rating and selection [41–43] or using the aforementioned rule-based quantitative methods (e.g., fuzzy sets, AHP). Yet, user needs and opinions are proven to play a critical role in successful concept evaluation and selection [19]. The growing popularity of online reviews on e-commerce platforms as a medium for users to express their sentiments and feedback about their experience with previous products provides an unprecedented opportunity to rethink design concept evaluation and actively engage users in the creative design process. For a new product to be a success, it must resonate with the needs and desires of large and diverse populations of users. Thus, there is a need for devising new methods that capture user sentiments and feedback on a large scale and leverage that information to project the success of new designs from the perspectives of potential users. This, in turn, would augment the ability of designers to make informed judgments and decisions during concept evaluation processes. This process has the potential to increase the quality, quantity, and diversity of user feedback versus traditional methods such as interviews, focus groups, and surveys, which are often used to inform MCDM methods such as QFD [44]. Traditionally, customer sentiments have been integrated into the new product development process in a sequential fashion before the concept development process begins or after concepts have been developed [45,46].

This paper develops and validates a deep multimodal design evaluation (DMDE) model that learns the complex relationships between the visual and functional characteristics of past designs and the attribute-level sentiments of users on a large scale and utilizes the learned patterns to measure the expected desirability of design concepts and their constituent attributes. DMDE draws on the existing literature on data-driven sentiment analysis from online reviews (e.g., Refs. [47–51]) for incorporating large-scale user feedback in concept evaluation. The proposed process allows the development of an iterative cycle where design concepts and user feedback are automated and integrated in parallel with the concept development process, thereby providing the design team with more data to better assess and select valuable concepts. From the perspective of the concept development process, the DMDE method adds a data-driven user-centered loop during the user insight and ideation process, which allows for more expansive user sentiments to be integrated within the design process and presented during concept evaluation. The DMDE loop added to the concept development process is shown in Fig. 1.

An overview of the DMDE model is shown in Fig. 2. The DMDE model processes the orthographic views of existing products using a state-of-the-art deep neural network-based model, ResNet-50 [52], which is pretrained on Image-Net [53] and fine-tuned on the collected product dataset. The DMDE model also processes textual product descriptions using a state-of-the-art deep language model, bidirectional encoder representations from transformer (BERT) [54], which is fine-tuned on a large product description dataset. To integrate these two modalities (i.e., image and text), a self-attention mechanism is developed and tested, which is proven to outperform baseline multimodal fusion architectures in jointly learning representations from multiple modalities. Comprehensive experiments are conducted on a large-scale dataset scraped from a major online footwear store to evaluate the effectiveness of the DMDE model in predicting the desirability of a concept using its orthographic renderings and textual descriptions (see, e.g., Fig. 3). Comparative analyses are performed on the training loss and regression accuracy of the DMDE model against two single modality deep neural networks and three multimodal fusion architectures, which indicate superior performance by a wide margin in terms of both mean squared loss and prediction accuracy rate (PAR). In sum, the main contributions of this paper are as follows:

(1) This paper tackles the challenging problem of user-centered design concept evaluation by devising a novel data-driven model that accurately predicts user sentiments and feedback for a new design concept based only upon its orthographic images and a brief description of characteristics (e.g., Fig. 3). The authors believe this model to be instrumental in providing better user data for the design team during the concept evaluation process.
(2) This paper proposes a novel multimodal deep regression architecture based on a self-attention mechanism that seamlessly integrates two different modalities in an end-to-end fashion to learn representations from both visual and textual features simultaneously.
(3) This paper conducts comprehensive experiments on a large-scale, real dataset to demonstrate the feasibility and performance of the proposed multimodal architecture compared to two sets of baselines: (a) single modality deep neural networks for image processing networks and natural language processing networks on a deep regression task

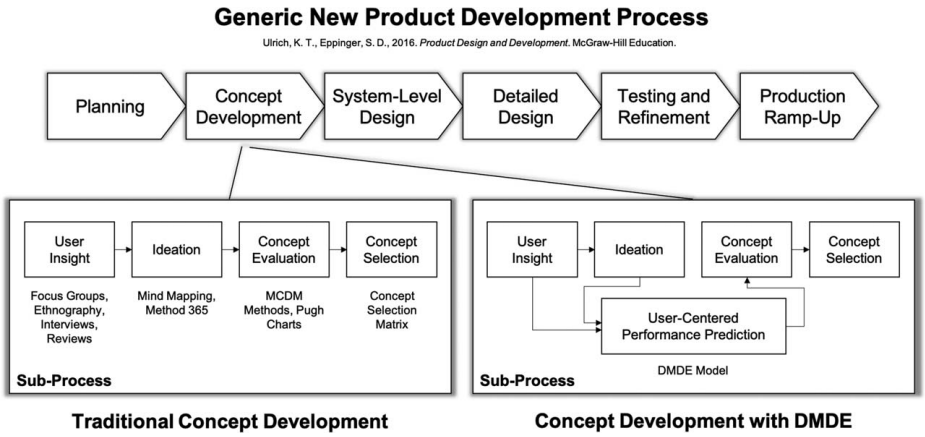

Fig. 1 The application of the DMDE model in the concept development process

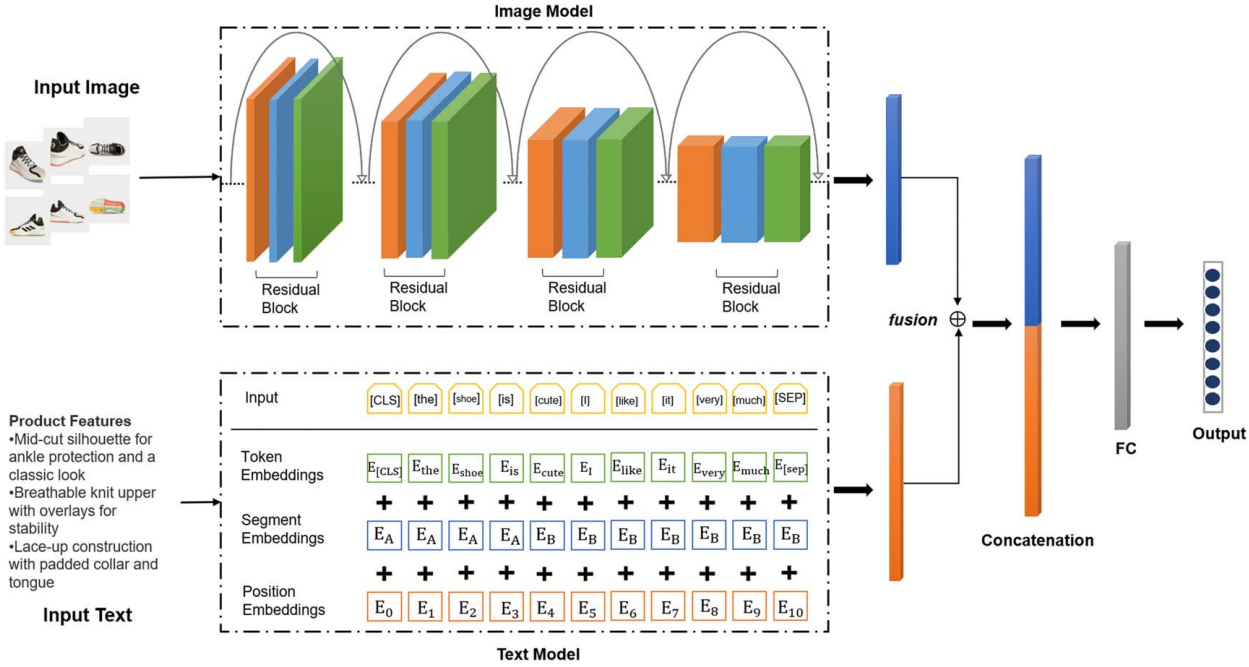

Fig. 2 Overview of the proposed DMDE model

Product Features
- •Mid-cut silhouette for ankle protection and a classic look
- •Breathable knit upper with overlays for stability
- •Lace-up construction with padded collar and tongue
- •Cushiony Lightstrike midsole for comfort
- •Durable rubber traction outsole
- •The adidas D Rose 11 is imported.

Fig. 3 Example of input data for the proposed DMDE model: orthographic product images (top) and textual product descriptions (bottom) used for predicting the overall and attribute-level user ratings

and (b) three state-of-the-art multimodal fusion methods. To the authors' knowledge, this is the first work that examines multimodal deep neural network architectures on a complex regression problem.

It must be noted, however, that the DMDE model is limited in its current form as it is most relevant to incremental innovations or improvements to existing products in the market. This model has a limited ability to predict more radical innovations in which the user has limited or no knowledge. Radical innovation is more related to technology-push innovation, in which a market is partially developed or not fully developed at all [55]. To address this limitation, we plan to investigate concept development and evaluation approaches using more latent needs and behavioral-based approaches, which are not tied explicitly to attributes or features that are common among a base of users.

The remainder of this paper is organized as follows. Section 2 provides a detailed overview of related work and topics in design concept evaluation as well as in multimodal network architectures and models. Section 3 discusses the details of the proposed DMDE model. Section 4 presents the experimental results, analyses, and performance evaluation. Finally, Sec. 5 provides concluding remarks and directions for future research.

2 Related Work

This section presents a detailed review of the related work on various design concept evaluation approaches and multimodal networks. Readers familiar with these topics may skip this section.

2.1 Design Concept Evaluation. Efficient evaluation of design concepts is a key requirement to facilitate new product development by ensuring design creativity and quality and preventing potential failures in later stages of the product development cycle [56,57]. Various evaluation approaches have been proposed and investigated in the design literature. AHP [58] was illustrated as a decision support model to aid designers in selecting new product ideas to pursue by helping identify the relationship between various design concepts in the evaluation process. AHP was adopted by Ayag [27] to select the best concept to satisfy the expectations of both the company and its customers. With the aid of the consultative AHP for computing the concept weighting values, the technique for order preference by similarity to an ideal solution [59] was integrated and proposed to assist designers in determining the optimal conceptual alternatives for further detailed development. By integrating perception-based concept evaluation and target costing of complex and large-scale systems, a system design methodology [60] decomposes a system into modules and evaluates each module concept with its target requirements and cost. A generalized purchase modeling approach [61] that considers generic factors such as anticipated market demand for the design, designers' preferences, and uncertainty in achieving predicted design attribute levels under different usage conditions and situations was proposed to develop a user-based expected utility metric.

To make the evaluation decisions more effective and to avoid the vagueness and uncertainty of experts' subjective judgments in conventional ways, various fuzzy set-based decision-making methods and algorithms have been proposed in the literature. Ayağ [62] employed a fuzzy AHP to reduce candidate concepts and exploited simulation analysis to improve the concept evaluation and selection. Further research [31,63] was conducted on the analytical network process (ANP), a more general form of AHP, to address the problem of accommodating the dependencies between higher and lower level elements. Fuzzy logic has also been proposed in conjunction with ANP to evaluate a set of conceptual design alternatives. The fuzzy-weighted average [64] method was developed to calculate desirability levels in engineering design evaluation, which suggests a new method of measuring design candidates by computing an aggregate fuzzy set [65]. A systematic decision process via the fuzzy set method [30] was also proposed in the literature to identify and choose the best design concept based upon expert knowledge and experience combined with optimization-based methodologies. Fuzzy analysis-based multi-criteria group decision-making methods [25,38] have also been employed for evaluating the performance of design alternatives, where all design alternatives are ranked and then selected according to the multiplied evaluation scores of concepts along with their weights.

To improve the evaluation process based on the fuzzy set method, interval arithmetic, rough sets, ranking design alternatives, and new methods were developed and integrated with other methods. An interval-based method [66] was proposed to effectively address uncertain and incomplete data and information in various instances of product design evaluation. Owing to the strength of rough sets in handling vagueness, a gray relation analysis integrated multi-criteria decision-making method [67] was proposed to evaluate design concepts to improve the effectiveness and objectivity of the design concept evaluation process. Other rough sets-based methods [68–70] were also developed to reduce evaluation bias in the pairwise comparison process in criteria weighting or rule mining. Integrated fuzzy sets [18] with genetic algorithms and neural networks were developed to identify the optimal concepts from a group of satisfactory concepts. Many experts apply methods of evaluating design concepts by ranking design alternatives in a qualitative fashion, such as multiattribute utility theory [71,72], preference ranking organization method for enrichment evaluations [73,74], and technique for order preference by similarity to an ideal solution [75,76].

2.2 Data-Driven Evaluation Methods. The approaches described above are based on subjective, insufficient, and ambiguous information for design concept evaluation [77]. Most information used in existing design concept evaluation practices comes from the subjective judgments of experts, which may be biased, vague, or even inconsistent. Therefore, data-driven methods have been proposed in the literature to achieve more reliable, quantitative evaluation results. For example, support vector machine (SVM) [78] has been proposed for predicting the design concept performance. SVM-based approaches [79,80] were introduced to develop a model that predicts the users' affective responses for product form design with satisfactory predictive performance. Other studies [80,81] have also reported comparable promising evaluation results. In addition, neural network-based models [82,83] have been recently proposed for design concept evaluation. Non-parametric models exploiting artificial neural networks [84] can predict software reliability based on a single and unified modality (e.g., fault history data) without any assumptions.

An automatic process [85] was recently proposed to extract the subjective knowledge of users and represent it using a fuzzy ontology, where the inherent user information is stored as a knowledge database and can be easily accessed by others. User preferences are then extracted for group decision-making. Various group decision-making methods [86–89] have been introduced in the literature to deal with heterogeneous information in a dynamic environment and measure the consistency of preferences provided by experts. Fuzzy morphological matrix-based systematic decision-making approaches [90,91] have also been studied for validating conceptual product design by employing the knowledge and preferences of designers and users with subjective uncertainties in function solution principles to evaluate design concepts quantitatively. In recent years, radicality computing formulas [92] have been proposed to regress through a statistical analysis of known design cases for additive manufacturing. According to their degree of radicality at the very beginning of the new product development process, the approach can rank potentially radical ideas. A new metric for evaluating creativity [41] was developed utilizing adjective selection and semantic similarity to minimize the designers' biases during the evaluating process. In light of these works that improve the efficiency and effectiveness of design concept evaluation, this paper proposes a novel data-driven method for design concept evaluation. To the authors' knowledge, no prior work has addressed the problem of multimodal design concept evaluation based upon hundreds of past designs as well as large-scale user sentiments and feedback extracted from myriad reviews available on e-commerce and social media platforms.

2.3 Multimodal Networks. In new product development, image processing-based and natural language processing-based approaches for need finding and design ideation have been well-studied independently in recent years. For example, image processing allows for the use of generative adversarial networks (GANs) to edit design concepts (e.g., apparel) at the attribute level automatically [93,94]. Convolutional neural network (CNN)-based architectures have also been utilized to predict the future desirability of styles discovered from fashion images in an unsupervised manner [95]. Furthermore, BERTs-based approaches have been adopted in recent years to extract information about the needs of users from online reviews [96,97]. However, these approaches base their evaluations on only a single modality (e.g., product image, review), which may naturally exclude other aspects of the design concept or product that are not represented by that single modality. In the design concept evaluation processes specifically, multimodal methods can provide more comprehensive and accurate information about the expected performance of a concept based on various metrics.

Recent research in machine learning has reported promising results in combining textual and visual data to learn multiple levels of representations through hierarchy network architectures. A deep CNN model [98] was trained to detect words in images, compose words into sentences, and map them onto the image features. A generative model-based method [99] was developed to generate natural sentences describing an image in an end-to-end manner using an encoder–decoder architecture. A deep learning-based text-to-image generation method [100] was proposed which uses the long short-term memory (LSTM) architecture for iterative handwriting-based control of image generation. Another study [101] employs deep learning methods to extract audio and visual features for noise removal in speech recognition. A recurrent CNN method [102] was also proposed to capture contextual information and extract features of images for text classification without human-designed features. Another work [103] proposed a joint feature learning approach that combines image features and text embeddings to classify document images. Similarly, Yang et al. [104] developed a fully CNN model for extracting semantic structures from document images, and Xu et al. [105] proposed an architecture that jointly learns text and layout in a single framework for document classification.

These multimodality methods mainly project language and image features into a shared representation and infer a single-modal feature from another feature, like inferring image features from a linguistic feature. However, this approach is most likely to cause information loss inevitably during the feature projection process. To avoid this issue, the model proposed in this paper addresses

such problems by capturing the single modality features (i.e., textual product descriptions and orthographic images) independently and then integrating them in an optimized fusion. Furthermore, it is observed that a large body of past research has leveraged multiple modalities to solve classification problems [106–109], while the multimodal deep learning architecture proposed in this paper solves a regression task; i.e., predicting the overall and attribute-level ratings of a design concept based on past user sentiments and feedback.

3 Methodology

This section presents the proposed DMDE model (Fig. 2). This paper builds on an attribute-level analysis approach for generating user-centered ratings for products based on online reviews without loss of generality, as described next. First, a sentiment analysis method is discussed, which is utilized to quantify the attribute-level sentiment intensity of users for each product based on online reviews. Next, the deep neural network architectures for image processing and natural language processing are discussed independently, followed by a description of the multimodal fusion method to integrate image and text features and relative performance metrics used to validate the models. Finally, it is worth noting that the DMDE model is domain-agnostic and can be generalized to any type of end-user product so long as the following data are available: product images, textual descriptions of product features, and overall and/or attribute-level ratings of the product. This paper builds on an attribute-level analysis approach for generating user-centered ratings for products based on online reviews, as described next.

3.1 Attribute-Level Sentiment Analysis. A given product typically receives tens, hundreds, or even thousands of user reviews presented in the form of unstructured natural language on an e-commerce platform. To inform the design process based on users' sentiments and feedback, advanced computational methods are required to translate large-scale, unstructured natural language data into valuable design knowledge and insights. In this paper, the first step of the proposed methodology is to process individual reviews to extract the attribute-level sentiment intensity of the users (i.e., the positivity or negativity of their emotions) associated with different attributes of the product. To this end, the analysis of sentiment expressions (ASEs) approach presented in Ref. [96] is adopted to measure the attribute-level sentiment intensity of users in four steps:

Step 1: A *product attribute lexicon* is created based on existing online product catalogs and attribute dictionaries.[2] In the case of footwear, for example, various synonyms of the main attributes collected from 10,000 reviews of our scraped dataset of footwear are selected as a total of 500 attribute words and grouped into 23 main attributes (e.g., color, energy return, permeability, weight, stability, durability). Then, the attribute lexicon of the products is used to extract product attributes from user reviews.

Step 2: The descriptions of the attributes are then tracked using natural language toolkit post tagger, which translates phrases or sentences into part-of-speech tags. Then the syntactic context of each sentence is derived. For example, a review sentence "I love the classic style" is chopped and translated into multiple pieces ("I," "PRP"), ("love," "VBP"), ("the," "DT"), ("classic," "JJ"), ("style," "NN").

Step 3: A *sentiment lexicon* is built on an enhanced state-of-the-art sentiment lexicon [51], which includes manually picked sentiment words from a dictionary with a vocabulary size of over 6000 words. In this paper, the sentiment lexicon is adapted by enriching it with domain-specific sentiment expressions related to the target product (e.g., footwear).

Step 4: Word embedding, a language modeling method that transfers words into high-dimensional vectors, is conducted to encode each word into a unique real vector so the computer can comprehend and operate on them. Word2Vec [110], one of the prominent pretrained models for word embedding, is utilized in this paper to learn word associations from a large corpus of text and translate each distinct word into a particular list of number vectors. Its simplicity drove the choice of Word2Vec for embedding; however, future studies may utilize more advanced context-aware embedding methods such as BERT [54].

Step 5: The ASE approach utilizes the product attribute and sentiment lexicons and word embeddings to identify and map sentiment expressions to the differentiated product attributes. The sentiment expressions of users are then converted into sentiment intensity values in $[-1, 1]$ using SenticNet, with -1 and 1 representing extremely negative sentiment and extremely positive sentiment, respectively.

The extracted attribute groups along with the user sentiment intensity values extracted from the ASE approach are then utilized as labels for the training data, as described in Sec. 4.

3.2 Image Processing. Images are an essential part of a design concept, representing the visual aspects of a conceptual design. The image features must be processed and extracted to estimate the expected user-centered desirability of a concept based on its orthographic renderings. Deep convolutional neural networks (CNNs) have led to a series of breakthroughs in image classification. The deep CNN-based model ResNet-50 [52] is a neural network used as a backbone for many computer vision tasks and has the strong ability to learn rich feature representations from a wide range of images. In this paper, Image-Net [53] pretrained ResNet-50 model is fine-tuned based on the scraped product image dataset to extract visual features from orthographic images.

To train the model, six orthographic images of each product serve as inputs of the network. The sentiment intensity values of users on ten product attributes "Traction," "Shape," "Heel," "Cushion," "Color," "Fit,""Impact absorption," "Durability," "Permeability," and "Stability," as well as on the overall rating are served as labels of training data in $[-1, 1]$. Images from the dataset are first resized to a batch of $224 \times 244 \times 3$ RGB images to fit the network. The ResNet-50 model consists of four stage residual blocks, each with a convolution and identity block. Each convolution block has three convolution, batch normalization, and ReLU layers, and each identity block also has 1 Conv 1×1 and a batch normalization layer to downsample the features. Finally, an average pool and a fully connected layer followed by a tanh function transfer features to the desired dimensional vector $X_1 \in \mathbb{R}^{d_1}$ at the end of the architecture. The ResNet-50 model has over 23 million trainable parameters in total. The main benefit of using such a deep network is that it can represent complex functions and learn features at many different levels of abstraction, from edges (at the lower layers) to very complex features (at the deeper layers) to better understand the dependency between the orthographic images of the design concepts (inputs) and the user sentiment intensity values (outputs).

3.3 Natural Language Processing. Online product catalogs typically comprise brief textual descriptions of the product features (e.g., Fig. 3). To identify the relationship between the technical descriptions of the products and the sentiment intensity of the users, BERTs [54] are utilized to train deep bidirectional representations from unlabeled text. BERT is the encoder stack of the transformer architecture [111]. A transformer architecture is an encoder–decoder network that uses self-attention on the encoder side and attention on the decoder side. In the proposed DMDE model, a

[2]https://github.com/hanyidaxia/NER_BERT

pretrained BERT by Wolf et al. [112] is adapted to learn and extract useful information from descriptive sentences in product descriptions and transform the textual content into a feature vector.

The pretrained BERT is applied on a regression task, estimating the relationship between inputs and multiple independent variables. The inputs are sentences describing the product and the multiple independent variables are sentiment intensity values in [−1, 1] associated with ten product attributes and the overall product rating. In the training process, most hyperparameters remain the same as the original BERT training. The BERT model size is $L = 12$, $H = 768$, and $A = 12$, where L, H, and A denote the number of layers, a hidden layer of size, and the number of self-attention heads, respectively. The model fits the input sequence and delivers the labels of text within the sequence as an output, where the sequence of inputs starts from [*CLS*] containing special embeddings and finish with the token [*SEP*] at the end of the sequence. The model performs tokenization by splitting the input text into a 128 sequence list of tokens. The input embeddings are then passed to the attention-based bidirectional transformer. The fully connected layer is revised at the end of the BERT model to ensure the desired dimension of the output feature $X_2 \in \mathbb{R}^{d_2}$ and followed by a tanh function to predict the labels (i.e., the sentiment intensity values).

3.4 Multimodal Architecture. The image processing model and the natural language processing model extract and represent the features from images and text, respectively, in an independent fashion. Each model has the capability to map the single modality feature (i.e., orthographic product images or textual product descriptions) into the extracted overall and attribute-level product ratings in [−1, 1]. Therefore, to model the connection between the orthographic images and descriptive language as input and the extracted overall and attribute-level product ratings as output, the DMDE model is enhanced with a novel fusion model to integrate the features associated with different modalities. The goal of this paper, however, is to evaluate a design concept based on both visual and textual information for more accurate and comprehensive evaluation. This section first describes two baseline multimodal fusion methods, followed by a novel self-attention-based multimodal fusion model with demonstrated improved performance (see Sec. 4) for integrating visual features from the ResNet-50 model and textual features from the BERT model for design concept evaluation.

3.4.1 Naïve Fusion. This approach integrates vectorized features from different information modes through naïve concatenation. The obtained image features $X_1 \in \mathbb{R}^{d_1}$ and text features $X_2 \in \mathbb{R}^{d_2}$ with their original dimensions are integrated. The generated multimodal features X_m are given by

$$X_m = X_1 \oplus X_2, \quad X_m \in \mathbb{R}^{d_1+d_2} \tag{1}$$

where $\oplus$ represents the concatenation operator of vectors.

3.4.2 Weighted Fusion. The linear weighted combination provides more flexibility for networks to assemble textual and visual representations. This approach integrates the multimodal features as follows:

$$X_m = w_1 \times X_1 \oplus w_2 \times X_2, \quad X_m \in \mathbb{R}^{d_1+d_2} \tag{2}$$

where $\oplus$ is the vector concatenate operator, and w_1 and w_2 denote the weighting parameters of image features and text features, respectively. The weighted parameters are tuned over the entire training process of the DMDE model.

3.4.3 Self-Attention Fusion. The attention mechanism is a powerful and widely used approach to integrate multiple modalities [113]. In this paper, a novel self-attention-based module inspired by Vaswani et al. [111] is developed to capture the representation and connection across the complementary information of multimodal features. Figure 4 illustrates the generation process of the multimodal self-attention features. Naïve concatenation features $X_m = X_1 \oplus X_2$ are initiated as the input of the self-attention-based module, and then the inputs X_m are projected onto a set of subspaces: query Q, key K, and value pair V. The multimodal self-attentive features S are formulated as

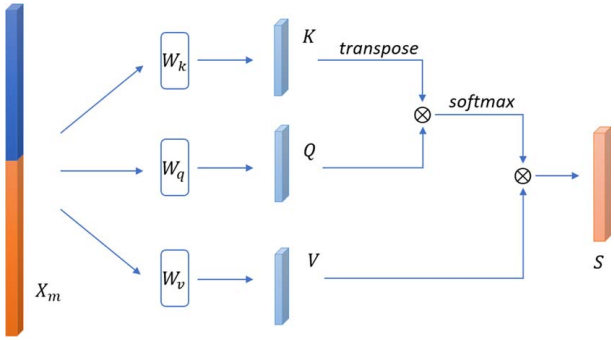

Fig. 4 Overview of the self-attention-based fusion module procedure. X_m denotes naïve concatenation features $X_1 \oplus X_2$ and $\otimes$ denotes matrix multiplication.

$$S = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d}}\right)V$$
$$S = [S_1, S_2, \ldots, S_n] \in \mathbb{R}^{d_{i\times n}} \tag{3}$$

where $Q = W_q \times S$; $K = W_k \times S$; $V = W_v \times S$; W_q, W_k, and W_k are three learned parameter matrices within the self-attention-based module; and d is the dimension of q, k. A softmax function is used to ensure attentions across each visual and textual cell.

3.4.4 Multimodal Fusion Process. The orthographic product images and textual product descriptions are used as inputs for the DMDE model (Fig. 2). The text features and image features are extracted simultaneously by the fine-tuned ResNet-50 model and BERT model, respectively. Once the two modality features are identified, they are integrated using the multimodal fusion layer constructed by the multimodal fusion methods described above. To ensure that the concatenated features have the desired dimension, a fully connected layer is added as follows to generate the final output:

$$Y' = W'^{T} \times S + b \tag{4}$$

where b is a bias vector and W' is the weight matrix. The entire procedure can be trained on the product images and descriptions scraped from an e-commerce platform to optimize the performance metrics described next.

3.5 Performance Metrics. The training procedure of the DMDE model is conducted using a loss function based on mean squared error (MSE). MSE is calculated as the mean or average of the squared differences between predicted and expected target values in a dataset, presented as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{5}$$

where Y_i is the expected value in the dataset and $\hat{Y}_i$ is the predicted value. The MSE loss can reflect the actual situation of regression error and evaluate the performance of the proposed multimodal networks.

To further investigate the effectiveness of the DMDE models in integrating the multimodality information and predicting the overall and attribute-level desirability of the design concepts, a PAR metric, inspired by the field accuracy rate [114], is utilized. PAR counts the

number of exact matches between ground-truth and predicted results as follows:

$$\text{PAR} = \frac{\#|Y_i - \hat{Y}_i| < \eta}{N} \tag{6}$$

where $\#|Y_i - \hat{Y}_i|$ is the number of products with an absolute value of the ground-truth and the predicted values below a threshold η, and N is the total number of products. η is set to 0.1 based on empirical knowledge; however, the sensitivity of PAR results to this value is analyzed in Sec. 4.4.

The MSE loss metric represents the squared difference between the predicted and actual performance ratings of products. Thus, a smaller loss represents the ability of the model to correctly map the images and textual descriptions of a product to its overall and attribute-level desirability. PAR is another metric that measures the rate of product performance predictions that fall within a pre-specified acceptable threshold. These metrics can help gauge the accuracy of the DMDE model and other baseline models in predicting the overall and attribute-level desirability of a new concept.

4 Experiments and Results

In this section, the dataset and implementation details of the proposed DMDE model are first described, followed by an explanatory analysis of the results of multimodal networks and the comparison with single-modal networks to demonstrate the accuracy and effectiveness of DMDE in predicting the expected desirability of design concepts and their attributes.

4.1 Dataset and Implementation Details. To test and validate the performance of the proposed multimodal networks in the evaluation of the newly designed concepts, a large-scale dataset was scraped from a major online footwear store to conduct numerical experiments. In the dataset, each product has four types of information: six orthographic images, one numerical rating score, a list of textual product descriptions, and real textual customer reviews from an e-commerce platform, where images and feature description are the inputs to the model and the numerical rating score and sentiment intensity values from customer reviews are the outputs. A total number of 8706 images and 113,391 reviews for 1452 identified shoes were collected from an online retail store. One example from the dataset is shown in Fig. 3. The experiments are conducted with k-fold [115], with $k = 10$, to randomly split the dataset into the train, validate and test sets with the ratio of 7 : 1 : 2. All experimental results are conducted five times and reported as mean ± std to alleviate the randomness effect. All neural networks are trained on PyTorch [116]. Adam [117] optimizer with $\beta = (0.9, 0.999)$ and learning rate = 0.01 is used to train the model parameters and save the model with the best loss in the validation dataset. To avoid overfitting, a dropout layer is added to the self-attention fusion model with the dropout rate of $P_{\text{drop}} = 0.1$. The DMDE model was trained over 40 epochs. The training time cost per epoch was 5–7 min, which added up to 3–4 h.

4.2 Data Processing. The proposed DMDE model is designed to predict the overall and attribute-level desirability of a product regarding the frequently mentioned attributes. The dataset provides a large-scale product image and description with labels of rating score and customer reviews. However, textual customer reviews are required to be represented as readable information for networks. Therefore, 113,391 reviews for 1452 identified shoes are analyzed by the natural language processing approach ASE (Sec. 3.1) and ten frequently mentioned product attributes were selected: "Traction," "Shape," "Heel," "Cushion," "Color," "Fit," "Impact Absorption," "Durability," "Permeability," and "Stability." Nevertheless, the attribute list can be expanded beyond the above list pending the availability of a sufficiently large number of examples for training the data-driven models. The sufficiency of attribute-specific data for training the neural networks is judged based on empirical knowledge. The sentiment expressions of users for each attribute were also extracted using the ASE method and converted into numerical values ranging from −1 to 1, representing both the polarity and intensity of user sentiments. Table 1 shows an example of the sentiment analysis of one product analyzed based on multiple user reviews, where positive/negative values represent the positivity/negativity of user sentiments. The closer the value to the endpoints of [−1, 1], the higher the sentiment intensity. The ten identified attribute values and the overall rating serve as labels in the training process of the deep regression model.

Table 1 Example of extracted attribute-level sentiment polarity and intensity from user reviews on a footwear product

Attributes	Sentiment value
Traction	−0.051
Shape	0.376
Heel	0.479
Cushion	0.449
Color	0.188
Fit	0.227
Impact absorption	0.302
Durability	0.356
Permeability	−0.326
Stability	0.202

4.3 Results and Analyses. To test and validate the performance of the proposed DMDE model for design concept evaluation, an ablation study was conducted to examine two unique aspects of the DMDE model: the self-attention-based fusion model and multimodal regression. First, the performance of the proposed self-attention-based fusion model is compared to the baseline models, naïve fusion and weighted fusion. Next, two single modality models (i.e., image-based and text-based) are used as two baselines and compared with the DMDE model. Finally, partial testing samples are presented to further demonstrate the effectiveness of the proposed DMDE model.

4.3.1 Multimodal Fusion Evaluation. The experiments on the three fusion models for integrating image and text modalities are shown in the last three rows of Table 2. Mean squared error (MSE) and PAR are used to measure and compare the performance of the three fusion models. Results show that naïve concatenation provides the DMDE model with the training loss of 0.0016, validation loss of 0.0019, and testing loss of 0.0020, while weighted concatenation achieves a lower loss at about 0.0013. However, the self-attention-based fusion model significantly outperforms both the naïve fusion model and the weighted fusion model by reducing the loss by 30–50%. The low loss values for training, validation, and testing indicate that the self-attention-based fusion module is capable to integrate and extract features from multiple modalities with very few errors.

The results of the experiments also indicate that the self-attention-based fusion model outperforms both the naïve fusion model and the weighted fusion model with the highest PAR of 99.14% and 99.10% in predicting the overall rating and the attribute-level rating, respectively. Additionally, the statistical p-values corresponding to a t-test on PAR are computed to demonstrate the significance of the difference between the multimodal models in Table 2 at the significance level of 0.05. The p-value is calculated as 0.03 for the self-attention fusion model versus the naïve fusion model, and the p-value associated with the comparison between the self-attention fusion model and the weighted fusion model is 0.02. These results of this statistical hypothesis testing indicate that the self-attention fusion model significantly outperform the other multimodal models in terms of PAR. To sum up, out of the three proposed models for integrating textual descriptions

Table 2 Comparison of the proposed DMDE model with baselines on MSE loss and PAR

Evaluated models		MSE loss			PAR (%)	
Modality	Model	Train	Validate	Test	Overall rating	Attribute rating
Single modality	Image model (ResNet-50)	0.0345 ± 0.015	0.0368 ± 0.0018	0.0408 ± 0.0022	76.54 ± 5.1	46.76 ± 4.5
	Text model (BERT)	0.0025 ± 0.007	0.0025 ± 0.008	0.0025 ± 0.0011	91.43 ± 2.6	95.46 ± 1.2
Multiple modalities	Naïve fusion	0.0016 ± 0.0003	0.0019 ± 0.0003	0.0020 ± 0.0004	98.27 ± 1.4	96.44 ± 1.1
	Weighted fusion	0.0013 ± 0.0002	0.0013 ± 0.0002	0.0014 ± 0.0005	98.87 ± 1.2	98.46 ± 0.9
	Self-attention fusion	0.0010 ± 0.0002	0.0010 ± 0.0003	0.0010 ± 0.0004	99.14 ± 0.8	99.10 ± 0.6

Note: Columns (1–2) are the five models including DMDE and baselines; columns (3–5) are MSE loss values for training, validating, and testing the models, respectively; columns (6–7) are the PAR of the models for overall rating and attribute-level rating of concepts, respectively.

and orthographic image features, the self-attention-based fusion method is proven to yield the best performance with the lowest MSE loss and the highest PAR for the DMDE model.

The limitation of naïve and weighted fusion for the DMDE model stems from the fact that the feature extraction modules for different modalities hardly interact with each other, which in turn limits their semantic relatedness and inevitably leads to information loss. Using the self-attention-based fusion method, the dependencies between the features of different modalities are not restricted by the in-between distance between them, unlike the naïve and weighted fusion methods which concatenate multimodal features. Consider a simple hypothetical example with image features $[a^1, a^2]$ and text features $[a^3, a^4]$. Using the naïve and weighted fusion methods, the combined features will be $[a^1, a^2, a^3, a^4]$ and $[w^1 a^1, w^2 a^2, w^3 a^3, w^4 a^4]$, respectively, where w denotes weight. The self-attention method, however, will combine these modalities as depicted in Fig. 5. The self-attention mechanism learns the relationships between the modalities using three query, key, and value matrices. Query and key construct the relationships, while value summarizes an output which comprises the relationships among all elements. The self-attention mechanism allows the inputs to interact with each other (i.e., "self") and determine what to pay more attention to (i.e., "attention"). The outputs are aggregates of these interactions and attention scores. Therefore, self-attention offers a larger and more optimal parameter space. Self-attention allows for identifying the latent relationships between visual features, semantic features, and product ratings. The proposed DMDE model with a self-attention mechanism has a promising ability to effectively capture essential features by combining the orthographic images and descriptions of products in predicting their overall and attribute-level ratings and desirability.

4.3.2 Single Modality Versus Multiple Modalities. The self-attention-based DMDE model was shown to deliver superior performance in solving the design concepts rating regression task. To further demonstrate the effectiveness of incorporating multiple modalities in the regression process as opposed to a single modality, this section presents the results of comparisons between the proposed self-attention-based DMDE model and single modality-based regression networks (i.e., image or text only) as baselines (Table 2).

First, single modality regression experiments with images were conducted using the ResNet-50 model. Given the orthographic images of products as input, the deep regression network predicts the overall and attribute-level ratings. As shown in Table 2, the image-only regression model achieves MSE loss values ranging from 0.0345 to 0.0408, with about a 30% higher margin than the self-attention-based model. The PAR of the image-only regression model is 76.54% for overall ratings, over 20% lower than the PAR of the DMDE model with self-attention-based fusion. The PAR of this model for attribute-level ratings is even lower, at 46.76%, almost half of the PAR of the DMDE model. Second, single modality regression experiments with textual descriptions were conducted using the BERT model. Results shown in Table 2 indicate that the text-only regression model outperforms the image-only regression

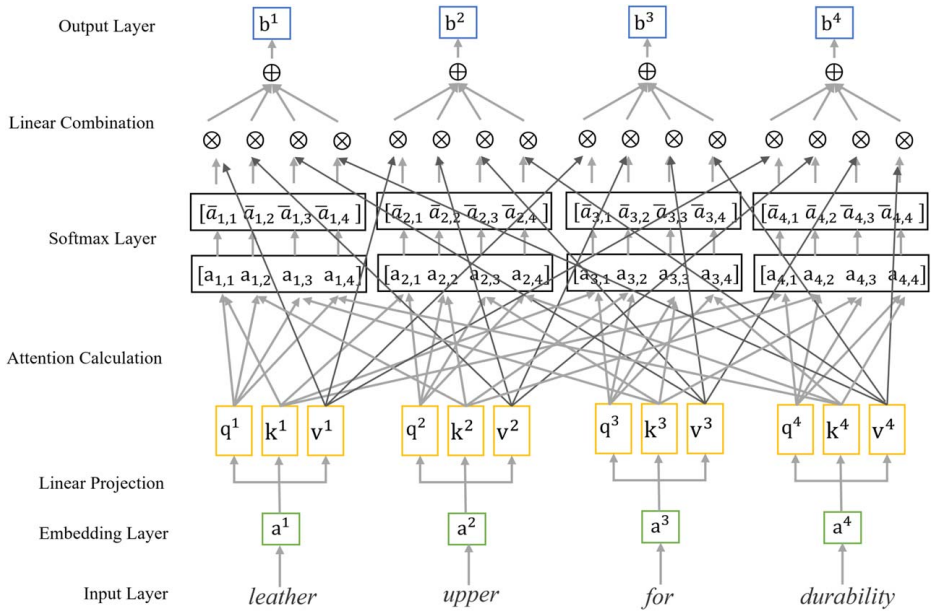

Fig. 5 Example of the self-attention mechanism: the fusion of image features $[a^1, a^2]$ and text features $[a^3, a^4]$. $q^i = w_q a^i$, $k^i = w_k a^i$, $v^i = w_v a^i$, $a_{i,j} = q^{i,*} k / \sqrt{d}$, $b^i = \sum_j \bar{\alpha}_{i,j} v^j$, where q is query, k is key, v is value pair, w_q, w_k, w_v are the target to train in a layer, d is the dimension of q, k.

Table 3 Comparison of PAR for overall rating with different η

Evaluated models		PAR (%) for overall rating		
Modality	Model	$\eta = 0.1$	$\eta = 0.075$	$\eta = 0.05$
Single modality	Image model (ResNet-50)	76.54	73.71	68.54
	Text model (BERT)	91.43	90.68	85.23
Multiple modalities	Naïve fusion	98.27	97.09	93.41
	Weighted fusion	98.87	97.26	94.41
	Self-attention fusion	99.14	98.34	97.91

model, with the MSE loss of 0.0025 and PAR of 91.43% (overall) and 95.46% (attribute level), which was anticipated because textual descriptions naturally contain more information about the product than images. However, the results of the DMDE model are still far better than the text-only regression model, which points to the importance of incorporating multiple modalities in the regression task for more accurate and representative results.

It is observed that the image-only regression model has the capability to predict the overall rating with higher PAR than the attribute-level ratings. On the contrary, the text-only regression model performs better in predicting attribute-level ratings than the overall rating. Two speculations can be drawn from these observations. First, only three out of the ten attributes can be rated based on the visual aspects of footwear (i.e., shape, heel, color). The remainder of the attributes cannot be judged only based on the orthographic images of the footwear, even by a human. That may be one of the reasons behind the poor performance of the image-only regression model in predicting the attribute-level ratings. Although its performance in predicting the overall rating is significantly better, it is still much worse than the text-only regression model and the DMDE model due to the same reason. Second, the text-only regression model demonstrates the opposite behavior with significantly improved attribute-level rating PAR. This may be partly due to the nature of the textual product description data (see Fig. 3) which contain structure information about different attributes of the footwear. Nevertheless, the multiple-modality models are shown to outperform both models by leveraging and integrating their features.

The experimental results demonstrate that the proposed deep multimodal networks for design concept evaluation enable state-of-the-art performance in predicting the expected overall and attribute-level ratings of products with low error and over 99.10% accuracy. Thus, the DMDE model creates an unprecedented opportunity for designers to accurately predict the expected success of their design concepts from the perspective of their end users, based only on their orthographic image renderings and standard textual descriptions.

4.4 Sensitivity to Prediction Accuracy Rate Threshold (η). In the PAR formula (Eq. 6), the threshold η was initially set to 0.1 based on empirical knowledge. To test the sensitivity of the prediction performance to the value of η, the experiments with the two single modality models and the three multimodality models were conducted using different values of η as shown in Table 3. The performances of the five models with respect to PAR were compared based on three values for η: 0.1, 0.075, 0.05, and 0.01. Results of the ablation study indicate that the single modality models are more sensitive to the threshold value η than the multimodality models. Specifically, reducing η from 0.1 to 0.05, the PAR of the image-only model and the text-only model decreases by 10% and 6.7%, respectively. Yet, the same change in the value of η results in a 4.9% reduction in the PAR of the naïve fusion model, a 4.5% reduction in the PAR of the weighted fusion model, and only a 1.2% reduction in the PAR of the proposed self-attention fusion model.

This ablation study shows that multimodality models outperform single modality models in preserving the robustness of the predicting accuracy; specifically, the PAR of the self-attention fusion model is the least sensitive to η because of good performance in learning the relationship among large datasets.The self-attention mechanism achieves construct relationships and extracts information by constructing the relationships, summarizing all relations within inputs, and concludes an output that contains relations among input vector elements via the three subspaces and optimal parameters. The self-attention allows the image features and textual features of the input to interact with each other and find out the high correspondence between them, which explains the strong ability of the model to obtain latent representation. Therefore, self-attention ensures a better fit to the deep multimodal evaluating procedure for product design than other models. The larger η provides higher accuracy of the model's testing performance and $\eta = 0.1$ allows self-attention fusion model a 99.14% accuracy which is the rationale of parameter setting for PAR metric.

4.5 Ablation Study on Multimodal Inputs. For training the DMDE model, six orthographic images of the product along with standard textual descriptions of product features serve as input. Once the model is trained, it must be able to evaluate the overall and attribute-level desirability of a new concept given its orthographic renderings and textual description data. This section presents the results of an ablation study on the input data after the DMDE model has been fully trained. The experiments are conducted to demonstrate the performance of the DMDE model during a test with different subsets of input data: a combination of two, four, or six images with full-text description or half-text description. Full-text description means the description of product features collected in the dataset is completely served as input. In the dataset, the product descriptions are itemized in several lists. To conduct the comparison, half of the lists that contain informative texts of product features are randomly selected to use as input of the model. The subsets of images are chosen randomly as well. Table 4 presents the results of the ablation study in terms of PAR for the overall rating with $\eta = 0.1$. It is observed that half-text descriptions with six images achieve 2.32% and 9.53% higher PAR than half-text description with four images and two images, respectively. In contrast, the full-text description with six images leads to 1.53% and 0.31% improvement in PAR when compared to full-text description with four images and two images,

Table 4 Comparison of testing PAR with subsets of inputs

Evaluated models Model	Testing subsets Number of images	PAR (%) for overall rating Half-text description	Full-text description
DMDE model	2	86.61	97.65
	4	92.71	98.83
	6	94.86	99.14

respectively. These findings indicate that the proposed DMDE model can guarantee a high accuracy of predicting the desirability of a new concept on any given number of images when full textual descriptions are provided. Furthermore, providing full-text descriptions can increase PAR by 12.75%, 6.6%, and 4.51% compared to half-text description, when two, four, and six images are available, respectively. Accordingly, the DMDE model appears to be significantly sensitive to the quality of textual descriptions, as the model performance considerably degrades when only half-text descriptions are provided. When full-text description is provided, however, the prediction accuracy value remains significantly high even with a small number of images (e.g., 97.65% with only two images). It is therefore concluded that once the DMDE model is trained, it can be used for evaluating new concepts even with a small number of images per concept, although a higher number of images is shown to result in higher accuracy and smaller MSE loss (see Table 2). However, it is highly recommended to provide sufficient descriptions of product features as inputs to obtain better prediction results from the DMDE model.

5 Conclusions and Future Research Directions

A novel neural network-based DMDE model was developed in this paper, which allows for accurate prediction of the overall and attribute-level desirability of a concept with respect to concerning large-scale user sentiments and feedback on past designs. A case study on a large-scale dataset scraped from an online footwear store was conducted to test and validate the performance of the DMDE model in terms of MSE error and PAR. Ablation studies on two unique aspects of the DMDE model indicated superior performance in terms of both MSE error and PAR when (1) multiple modalities are incorporated in the regression task and (2) the modalities are integrated using the proposed self-attention-based fusion mechanism. To construct a multimodal network, the single image processing model and natural language processing model are built independently based on state-of-the-art pretrained models ResNet-50 and BERT, respectively. The main goal of the fine-tuned ResNet-50 and BERT models is to extract useful features from orthographic product images and textual product descriptions, respectively. The self-attention method was then applied to integrate the textual and visual features and capture the dependency between multiple modalities to predict product rating labels accurately. The proposed model can serve as an intelligent guidance tool for new product designers to predict how their concepts will perform from end users' perspectives regarding both overall and attribute-level desirability. Specifically, a design team can simply feed the photorealistic renderings and technical description of a new concept (e.g., a pair of sneakers; see Fig. 3) into the DMDE model to accurately predict its overall and attribute-level desirability based on large-scale user feedback on previous designs.

Another important direction for future research in this area is to couple the proposed DMDE model with generative design algorithms for automated design concept generation. Deep generative models have been recently adopted for design automation [118–120] to improve designers' performance through *co-creation with AI*. Specifically, GANs [121] have shown tremendous success in a variety of generative design tasks, from topology optimization [118] to material design [122] and shape parametrization [119]. In line with Osborn's rules for brainstorming [2], these generative models have proven effective in increasing the quantity of ideas at the designer's disposal to inspire her exploration and avoid investing too heavily in few ideas. Current approaches for assessing the quality of GAN-generated samples are limited to manual assessment and the use of various convergence criteria and distance metrics for comparing real and generated images in the feature space. Some recent studies have proposed using physics-based simulators for performance assessment of generative design with respect to form and function [119]; however, those mechanisms are domain-specific and applicable to a limited set of functional attributes (e.g., aerodynamic performance). The proposed DMDE model can potentially bridge this knowledge gap by serving as a disruptive tool for accurate, data-driven evaluation of GAN-generated design concepts.

Acknowledgment

This material is based upon work supported by the National Science Foundation under the Engineering Design and System Engineering (EDSE) Grant No. 2050052. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtained from the corresponding author upon reasonable request.

References

[1] Takeuchi, H., and Nonaka, I., 1986, "The New Product Development Game," Harvard Business Rev., **64**(1), pp. 137–146.
[2] Osborn, A., 1953, *Applied Imagination*, Scribner's, New York.
[3] Yilmaz, S., Seifert, C., Daly, S. R., and Gonzalez, R., 2016, "Design Heuristics in Innovative Products," J. Mech. Des., **138**(7), p. 071102.
[4] Crismond, D. P., and Adams, R. S., 2012, "The Informed Design Teaching and Learning Matrix," J. Eng. Educ., **101**(4), p. 738.
[5] Forbes, H., and Schaefer, D., 2018, "Crowdsourcing in Product Development: Current State and Future Research Directions," Proceedings of the DESIGN 2018 15th International Design Conference, Dubrovnik, Croatia, May 21–24, pp. 579–588.
[6] Mumford, M. D., Feldman, J. M., Hein, M. B., and Nagao, D. J., 2001, "Tradeoffs Between Ideas and Structure: Individual Versus Group Performance in Creative Problem Solving," J. Creat. Behav., **35**(1), pp. 1–23.
[7] Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., and Markman, A. B., 2011, "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods," ASME J. Mech. Des., **133**(3), p. 031008.
[8] Yilmaz, S., Daly, S. R., Seifert, C. M., and Gonzalez, R., 2016, "Evidence-Based Design Heuristics for Idea Generation," Des. Stud., **46**(C), pp. 95–124.
[9] Simonton, D. K., 1990, *Psychology, Science, and History: An Introduction to Historiometry*, Yale University Press, New Haven, CT.
[10] Daly, S. R., Seifert, C. M., Yilmaz, S., and Gonzalez, R., 2016, "Comparing Ideation Techniques for Beginning Designers," ASME J. Mech. Des., **138**(10), p. 101108.
[11] Han, J., Shi, F., Chen, L., and Childs, P. R., 2018, "The Combinator—A Computer-Based Tool for Creative Idea Generation Based on a Simulation Approach," Des. Sci., **4**(11), pp. 1–34.
[12] Gerhard, P., and Karl-Heinrich, G., 1984, *Engineering Design: A Systematic Approach*, Springer, Berlin, Germany.
[13] Howard, T. J., Culley, S., and Dekoninck, E. A., 2011, "Reuse of Ideas and Concepts for Creative Stimuli in Engineering Design," J. Eng. Des., **22**(8), pp. 565–581.
[14] Gray, C. M., McKilligan, S., Daly, S. R., Seifert, C. M., and Gonzalez, R., 2019, "Using Creative Exhaustion to Foster Idea Generation," Int. J. Technol. Des. Educ., **29**(1), pp. 177–195.
[15] Shidpour, H., Da Cunha, C., and Bernard, A., 2016, "Group Multi-Criteria Design Concept Evaluation Using Combined Rough Set Theory and Fuzzy Set Theory," Expert Syst. Appl., **64**(C), pp. 633–644.
[16] Pugh, S., and Clausing, D., 1996, *Creating Innovtive Products Using Total Design: The Living Legacy of Stuart Pugh*, Addison-Wesley Longman Publishing Co Inc., Boston, MA.
[17] Tsai, H.-C., and Hsiao, S.-W., 2004, "Evaluation of Alternatives for Product Customization Using Fuzzy Logic," Inform. Sci., **158**(10), pp. 233–262.
[18] Huang, H.-Z., Bo, R., and Chen, W., 2006, "An Integrated Computational Intelligence Approach to Product Concept Generation and Evaluation," Mech. Mach. Theory, **41**(5), pp. 567–583.
[19] Huang, H.-Z., Liu, Y., Li, Y., Xue, L., and Wang, Z., 2013, "New Evaluation Methods for Conceptual Design Selection Using Computational Intelligence Techniques," J. Mech. Sci. Technol., **27**(3), pp. 733–746.

[20] Ulrich, K. T., 2003, *Product Design and Development*, Tata McGraw-Hill Education, New York.
[21] Dym, C. L., Wood, W. H., and Scott, M. J., 2002, "Rank Ordering Engineering Designs: Pairwise Comparison Charts and Borda Counts," Res. Eng. Des., **13**(4), pp. 236–242.
[22] Frey, D. D., Herder, P. M., Wijnia, Y., Subrahmanian, E., Katsikopoulos, K., and Clausing, D. P., 2009, "The Pugh Controlled Convergence Method: Model-Based Evaluation and Implications for Design Theory," Res. Eng. Des., **20**(1), pp. 41–58.
[23] Scott, M. J., and Antonsson, E. K., 1998, "Aggregation Functions for Engineering Design Trade-offs," Fuzzy Sets Syst., **99**(3), pp. 253–264.
[24] King, A. M., and Sivaloganathan, S., 1999, "Development of a Methodology for Concept Selection in Flexible Design Strategies," J. Eng. Des., **10**(4), pp. 329–349.
[25] Thurston, D., and Carnahan, J., 1992, "Fuzzy Ratings and Utility Analysis in Preliminary Design Evaluation of Multiple Attributes," ASME J. Mech. Des., **114**(4), pp. 648–658.
[26] Wang, J., 2001, "Ranking Engineering Design Concepts Using a Fuzzy Outranking Preference Model," Fuzzy Sets Syst., **119**(1), pp. 161–170.
[27] Ayag*, Z., 2005, "An Integrated Approach to Evaluating Conceptual Design Alternatives in a New Product Development Environment," Int. J. Prod. Res., **43**(4), pp. 687–713.
[28] Papadakis, V. M., and Barwise, P., 2002, "How Much Do CEOs and Top Managers Matter in Strategic Decision-Making?," Br. J. Manage., **13**(1), pp. 83–95.
[29] Scott, M. J., 2002, "Quantifying Certainty in Design Decisions: Examining AHP," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Montreal, Quebec, Canada, Sept. 29–Oct. 2, Vol. 3624, pp. 219–229.
[30] Malekly, H., Mousavi, S. M., and Hashemi, H., 2010, "A Fuzzy Integrated Methodology for Evaluating Conceptual Bridge Design," Expert Syst. Appl., **37**(7), pp. 4910–4920.
[31] Ayağ, Z., and Özdem[idot]r, R., 2007, "An Analytic Network Process-Based Approach to Concept Evaluation in a New Product Development Environment," J. Eng. Des., **18**(3), pp. 209–226.
[32] Huang, H.-Z., Li, Y., Liu, W., Liu, Y., and Wang, Z., 2011, "Evaluation and Decision of Products Conceptual Design Schemes Based on Customer Requirements," J. Mech. Sci. Technol., **25**(9), pp. 2413–2425.
[33] Ramanujan, D., Nawal, Y., Reid, T., and Ramani, K., 2015, "Informing Early Design Via Crowd-Based Co-creation," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Boston, MA, Aug. 2–5, Vol. 57175, American Society of Mechanical Engineers, p. V007T06A043..
[34] Griffin, A., and Hauser, J. R., 1993, "The Voice of the Customer," Market. Sci., **12**(1), pp. 1–27.
[35] Cooper, R. G., 2008, "Perspective: The Stage-gate® Idea-to-Launch Process–Update, What's New, and Nexgen Systems," J. Prod. Innov. Manage., **25**(3), pp. 213–232.
[36] Zheng, J., and Jakiela, M. J., 2009, "An Investigation of the Productivity Difference in Mechanical Embodiment Design Between Face-to-Face and Threaded Online Collaboration," Proceedings of the ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, San Diego, CA, Aug. 30–Sept. 2, Vol. 48999, pp. 1173–1182.
[37] Burnap, A., Hartley, J., Pan, Y., Gonzalez, R., and Papalambros, P. Y., 2016, "Balancing Design Freedom and Brand Recognition in the Evolution of Automotive Brand Styling," Des. Sci., **2**(9), pp. 1–28.
[38] Zhang, Z., and Chu, X., 2009, "A New Integrated Decision-Making Approach for Design Alternative Selection for Supporting Complex Product Development," Int. J. Comput. Int. Manuf., **22**(3), pp. 179–198.
[39] Sa'Ed, M. S., and Al-Harris, M. Y., 2014, "New Product Concept Selection: An Integrated Approach Using Data Envelopment Analysis (DEA) and Conjoint Analysis (CA)," Int. J. Eng. Technol., **3**(1), p. 44.
[40] Feyzioglu, O., and Buyukozkan, G., 2006, "Evaluation of New Product Development Projects Using Artificial Intelligence and Fuzzy Logic," International Conference on Knowledge Mining and Computer Science, Las Vegas, NV, June 26–29, Vol. 11, pp. 183–189.
[41] Gosnell, C. A., and Miller, S. R., 2016, "But Is It Creative? Delineating the Impact of Expertise and Concept Ratings on Creative Concept Selection," ASME J. Mech. Des., **138**(2), p. 021101.
[42] Toh, C. A., and Miller, S. R., 2015, "How Engineering Teams Select Design Concepts: A View Through the Lens of Creativity," Des. Stud., **38**(C), pp. 111–138.
[43] Nikander, J. B., Liikkanen, L. A., and Laakso, M., 2014, "The Preference Effect in Design Concept Evaluation," Des. Stud., **35**(5), pp. 473–499.
[44] Hauser, J. R., and Clausing, D., 1988, *The House of Quality*, Harvard Business Review.
[45] Liedtka, J., 2015, "Perspective: Linking Design Thinking With Innovation Outcomes Through Cognitive Bias Reduction," J. Prod. Innov. Manage., **32**(6), pp. 925–938.
[46] Ulrich, K., and Eppinger, S., 2016, *Product Design and Development*, McGraw-Hill Education, New York.
[47] Suryadi, D., and Kim, H. M., 2019, "A Data-Driven Methodology to Construct Customer Choice Sets Using Online Data and Customer Reviews," ASME J. Mech. Des., **141**(11), p. 111103.
[48] Joung, J., and Kim, H. M., 2021, "Approach for Importance–Performance Analysis of Product Attributes From Online Reviews," ASME J. Mech. Des., **143**(8), p. 081705.
[49] Zhang, L., Wang, S., and Liu, B., 2018, "Deep Learning for Sentiment Analysis: A Survey," Wiley Interdiscipl. Rev.: Data Mining Knowledge Discov., **8**(4), p. e1253.
[50] Tang, H., Tan, S., and Cheng, X., 2009, "A Survey on Sentiment Detection of Reviews," Expert Syst. Appl., **36**(7), pp. 10760–10773.
[51] Liu, B., 2010, "Sentiment Analysis and Subjectivity," *Handbook of Natural Language Processing*, 2nd ed., N. Indurkhya and F. J. Damerau, eds, Vol. 2, pp. 627–666.
[52] He, K., Zhang, X., Ren, S., and Sun, J., 2016, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, June 27–30, pp. 770–778.
[53] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., 2009, "ImageNet: A Large-Scale Hierarchical Image Database," IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, June 20–25, pp. 248–255.
[54] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., 2018, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, {NAACL-HLT} 2019, Minneapolis, MN, June 2–7.
[55] Verganti, R., 2008, "Design, Meanings, and Radical Innovation: A Metamodel and a Research Agenda," J. Prod. Innov. Manage., **25**(5), pp. 436–456.
[56] Zhu, G.-N., Hu, J., Qi, J., Gu, C.-C., and Peng, Y.-H., 2015, "An Integrated AHP and Vikor for Design Concept Evaluation Based on Rough Number," Adv. Eng. Inform., **29**(3), pp. 408–418.
[57] Toh, C. A., and Miller, S. R., 2016, "Choosing Creativity: The Role of Individual Risk and Ambiguity Aversion on Creative Concept Selection in Engineering Design," Res. Eng. Des., **27**(3), pp. 195–219.
[58] Calantone, R. J., Di Benedetto, C. A., and Schmidt, J. B., 1999, "Using the Analytic Hierarchy Process in New Product Screening," J. Prod. Innov. Manage.: Inter. Public Product Dev. Manage. Assoc., **16**(1), pp. 65–76.
[59] Lin, M.-C., Wang, C.-C., Chen, M.-S., and Chang, C. A., 2008, "Using AHP and Topsis Approaches in Customer-Driven Product Design Process," Comput. Ind., **59**(1), pp. 17–31.
[60] Takai, S., and Ishii, K., 2006, "Integrating Target Costing Into Perception-Based Concept Evaluation of Complex and Large-Scale Systems Using Simultaneously Decomposed QFD," ASME J. Mech. Des, **128**(6), pp. 1186–1195.
[61] Besharati, B., Azarm, S., and Kannan, P., 2006, "A Decision Support System for Product Design Selection: A Generalized Purchase Modeling Approach," Decision Support Syst., **42**(1), pp. 333–350.
[62] Ayağ, Z., 2005, "A Fuzzy AHP-Based Simulation Approach to Concept Evaluation in a NPD Environment," IIE Trans., **37**(9), pp. 827–842.
[63] Ayağ, Z., and Özdemir, R. G., 2009, "A Hybrid Approach to Concept Selection Through Fuzzy Analytic Network Process," Comput. Ind. Eng., **56**(1), pp. 368–379.
[64] Vanegas, L., and Labib, A., 2001, "Application of New Fuzzy-Weighted Average (NFWA) Method to Engineering Design Evaluation," Int. J. Prod. Res., **39**(6), pp. 1147–1162.
[65] Vanegas, L. V., and Labib, A. W., 2005, "Fuzzy Approaches to Evaluation in Engineering Design," ASME J. Mech. Des., **127**(1), pp. 24–33.
[66] Chin, K.-S., Yang, J.-B., Guo, M., and Lam, J. P.-K., 2009, "An Evidential-Reasoning-Interval-Based Method for New Product Design Assessment," IEEE Trans. Eng. Manage., **56**(1), pp. 142–156.
[67] Zhai, L.-Y., Khoo, L.-P., and Zhong, Z.-W., 2009, "Design Concept Evaluation in Product Development Using Rough Sets and Grey Relation Analysis," Expert Syst. Appl., **36**(3), pp. 7072–7079.
[68] Li, Y., Tang, J., Luo, X., and Xu, J., 2009, "An Integrated Method of Rough Set, Kano's Model and AHP for Rating Customer Requirements' Final Importance," Expert Syst. Appl., **36**(3), pp. 7045–7053.
[69] Aydogan, E. K., 2011, "Performance Measurement Model for Turkish Aviation Firms Using the Rough-AHP and Topsis Methods Under Fuzzy Environment," Expert Syst. Appl., **38**(4), pp. 3992–3998.
[70] Zou, Z., Tseng, T.-L. B., Sohn, H., Song, G., and Gutierrez, R., 2011, "A Rough Set Based Approach to Distributor Selection in Supply Chain Management," Expert Syst. Appl., **38**(1), pp. 106–115.
[71] Ashour, O. M., and Kremer, G. E. O., 2013, "A Simulation Analysis of the Impact of Fahp–maut Triage Algorithm on the Emergency Department Performance Measures," Expert Syst. Appl., **40**(1), pp. 177–187.
[72] Jimenez, A., Mateos, A., and Sabio, P., 2013, "Dominance Intensity Measure Within Fuzzy Weight Oriented Maut: An Application," Omega, **41**(2), pp. 397–405.
[73] Kilic, H. S., Zaim, S., and Delen, D., 2015, "Selecting "the Best" ERP System for SMES Using a Combination of ANP and Promethee Methods," Expert Syst. Appl., **42**(5), pp. 2343–2352.
[74] Vetschera, R., and De Almeida, A. T., 2012, "A Promethee-Based Approach to Portfolio Selection Problems," Comput. Oper. Res., **39**(5), pp. 1010–1020.
[75] Song, W., Ming, X., and Wu, Z., 2013, "An Integrated Rough Number-Based Approach to Design Concept Evaluation Under Subjective Environments," J. Eng. Des., **24**(5), pp. 320–341.
[76] Ayağ, Z., 2016, "An Integrated Approach to Concept Evaluation in a New Product Development," J. Intell. Manuf., **27**(5), pp. 991–1005.
[77] Zhang, Z.-J., Gong, L., Jin, Y., Xie, J., and Hao, J., 2017, "A Quantitative Approach to Design Alternative Evaluation Based on Data-Driven Performance Prediction," Adv. Eng. Inform., **32**(C), pp. 52–65.

[78] Cortes, C., and Vapnik, V., 1995, "Support-Vector Networks," Mach. Learn., **20**(3), pp. 273–297.
[79] Shieh, M.-D., and Yang, C.-C., 2008, "Classification Model for Product Form Design Using Fuzzy Support Vector Machines," Comput. Ind. Eng., **55**(1), pp. 150–164.
[80] Yang, C.-C., and Shieh, M.-D., 2010, "A Support Vector Regression Based Prediction Model of Affective Responses for Product Form Design," Comput. Ind. Eng., **59**(4), pp. 682–689.
[81] Yang, C.-C., 2011, "Constructing a Hybrid Kansei Engineering System Based on Multiple Affective Responses: Application to Product Form Design," Comput. Ind. Eng., **60**(4), pp. 760–768.
[82] Hsiao, S.-W., and Huang, H.-C., 2002, "A Neural Network Based Approach for Product Form Design," Des. Stud., **23**(1), pp. 67–84.
[83] Hsiao, S.-W., and Tsai, H.-C., 2005, "Applying a Hybrid Approach Based on Fuzzy Neural Network and Genetic Algorithm to Product Form Design," Int. J. Ind. Ergon., **35**(5), pp. 411–428.
[84] Roy, P., Mahapatra, G., Rani, P., Pandey, S., and Dey, K., 2014, "Robust Feedforward and Recurrent Neural Network Based Dynamic Weighted Combination Models for Software Reliability Prediction," Appl. Soft. Comput., **22**(C), pp. 629–637.
[85] Morente-Molinera, J., Pérez, I., Ureña, M., and Herrera-Viedma, E., 2016, "Creating Knowledge Databases for Storing and Sharing People Knowledge Automatically Using Group Decision Making and Fuzzy Ontologies," Inform. Sci., **328**(C), pp. 418–434.
[86] Lourenzutti, R., and Krohling, R. A., 2016, "A Generalized Topsis Method for Group Decision Making With Heterogeneous Information in a Dynamic Environment," Inform. Sci., **330**(C), pp. 1–18.
[87] Zhang, X., Ge, B., Jiang, J., and Tan, Y., 2016, "Consensus Building in Group Decision Making Based on Multiplicative Consistency With Incomplete Reciprocal Preference Relations," Knowled.-Based Syst., **106**(5), pp. 96–104.
[88] Cabrerizo, F. J., Moreno, J. M., Pérez, I. J., and Herrera-Viedma, E., 2010, "Analyzing Consensus Approaches in Fuzzy Group Decision Making: Advantages and Drawbacks," Soft Comput., **14**(5), pp. 451–463.
[89] Pérez, I. J., Cabrerizo, F. J., Alonso, S., and Herrera-Viedma, E., 2013, "A New Consensus Model for Group Decision Making Problems With Non-homogeneous Experts," IEEE Trans. Syst. Man Cybernet.: Syst., **44**(4), pp. 494–498.
[90] Ma, H., Chu, X., Xue, D., and Chen, D., 2017, "A Systematic Decision Making Approach for Product Conceptual Design Based on Fuzzy Morphological Matrix," Expert Syst. Appl., **81**(C), pp. 444–456.
[91] Zheng, H., Feng, Y., Gao, Y., and Tan, J., 2018, "A Robust Predicted Performance Analysis Approach for Data-Driven Product Development in the Industrial Internet of Things," Sensors, **18**(9), p. 2871.
[92] Liu, W., Tan, R., Cao, G., Zhang, Z., Huang, S., and Liu, L., 2019, "A Proposed Radicality Evaluation Method for Design Ideas at Conceptual Design Stage," Comput. Ind. Eng., **132**(C), pp. 141–152.
[93] Kang, W.-C., Fang, C., Wang, Z., and McAuley, J., 2017, "Visually-Aware Fashion Recommendation and Design With Generative Image Models," 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, Nov. 18–21, IEEE, pp. 207–216.
[94] Yuan, C., and Moghaddam, M., 2020, "Attribute-Aware Generative Design With Generative Adversarial Networks," IEEE Access, **8**(C), p. 190710.
[95] Al-Halah, Z., Stiefelhagen, R., and Grauman, K., 2017, "Fashion Forward: Forecasting Visual Style in Fashion," Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, Oct. 22–29, pp. 388–397.
[96] Han, Y., and Moghaddam, M., 2021, "Analysis of Sentiment Expressions for User-Centered Design," Expert Syst. Appl., **171**(C), p. 114604.
[97] El Dehaibi, N., Goodman, N. D., and MacDonald, E. F., 2019, "Extracting Customer Perceptions of Product Sustainability From Online Reviews," ASME J. Mech. Des., **141**(12), p. 121103.
[98] Fang, H., Gupta, S., Iandola, F., Srivastava, R. K., Deng, L., Dollár, P., Gao, J., He, X., Mitchell, M., Platt, J. C., Zitnick, C. L., and Zweig, G., 2015, "From Captions to Visual Concepts and Back," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, June 7–12, pp. 1473–1482.
[99] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D., 2015, "Show and Tell: A Neural Image Caption Generator," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, June 7–12, pp. 3156–3164.
[100] Gregor, K., Danihelka, I., Graves, A., Rezende, D., and Wierstra, D., 2015, "Draw: A Recurrent Neural Network for Image Generation," Proceedings of the 32nd International Conference on Machine Learning, Lille, France, July 6–11, PMLR, pp. 1462–1471.
[101] Huang, J., and Kingsbury, B., 2013, "Audio-Visual Deep Learning for Noise Robust Speech Recognition," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, May 26–31, pp. 7596–7599.
[102] Lai, S., Xu, L., Liu, K., and Zhao, J., 2015, "Recurrent Convolutional Neural Networks for Text Classification," Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, Jan. 25–30, AAAI Press, pp. 2267–2273.
[103] Audebert, N., Herold, C., Slimani, K., and Vidal, C., 2019, "Multimodal Deep Networks for Text and Image-Based Document Classification," Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Würzburg, Germany , Sept. 16–20, Springer, pp. 427–443.
[104] Yang, X., Yumer, E., Asente, P., Kraley, M., Kifer, D., and Lee Giles, C., 2017, "Learning to Extract Semantic Structure From Documents Using Multimodal Fully Convolutional Neural Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, July 21–26, pp. 5315–5324.
[105] Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., and Zhou, M., 2020, "Layoutlm: Pre-training of Text and Layout for Document Image Understanding," Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA, July 6–10.
[106] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Y., 2011, "Multimodal Deep Learning," Proceedings of the 28th International Conference on International Conference on Machine Learning, Bellevue, WA, June 28–July 2, pp. 689–696.
[107] Lynch, C., Aryafar, K., and Attenberg, J., 2016, "Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, Aug. 13–17, pp. 541–548.
[108] Kiros, R., Salakhutdinov, R., and Zemel, R., 2014, "Multimodal Neural Language Models," International Conference on Machine Learning, Beijing, China, June 21–26, PMLR, pp. 595–603.
[109] Gong, Y., Wang, L., Hodosh, M., Hockenmaier, J., and Lazebnik, S., 2014, "Improving Image-Sentence Embeddings Using Large Weakly Annotated Photo Collections," European Conference on Computer Vision, Zurich, Switzerland, Sept. 6–12, Springer, pp. 529–545.
[110] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J., Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., 2013, "Distributed Representations of Words and Phrases and Their Compositionality," Advances in Neural Information Processing Systems, Lake Tahoe, NV, Dec. 5–10.
[111] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., 2017, "Attention Is All You Need," 31st Conference on Neural Information Processing System, Long Beach, NY, Dec. 4–9.
[112] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M., 2020, "Transformers: State-of-the-Art Natural Language Processing," Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Virtually, Nov. 16–20, pp. 38–45.
[113] Hori, C., Hori, T., Lee, T.-Y., Zhang, Z., Harsham, B., Hershey, J. R., Marks, T. K., and Sumi, K., 2017, "Attention-Based Multimodal Fusion for Video Description," Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, Oct. 22–29, pp. 4193–4202.
[114] Kerroumi, M., Sayem, O., and Shabou, A., 2021, "Visualwordgrid: Information Extraction From Scanned Documents Using a Multimodal Approach," ICDAR Workshops 2021, Lausanne, Switzerland, Sept. 5–7.
[115] McLachlan, G. J., Do, K. -A., and Ambroise, C., 2004, *Analyzing Microarray Gene Expression Data*, Wiley, Hoboken, NJ.
[116] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., and Antiga, L., 2019, "Pytorch: An Imperative Style, High-Performance Deep Learning Library," Conference on Neural Information Processing Systems, Vancouver, Canada, Dec. 8–14.
[117] Kingma, D. P., and Ba, J., 2015, "Adam: A Method for Stochastic Optimization," International Conference for Learning Representations, San Diego, CA, May 7–9.
[118] Oh, S., Jung, Y., Kim, S., Lee, I., and Kang, N., 2019, "Deep Generative Design: Integration of Topology Optimization and Generative Models," ASME J. Mech. Des., **141**(11), p. 111405.
[119] Shu, D., Cunningham, J., Stump, G., Miller, S. W., Yukish, M. A., Simpson, T. W., and Tucker, C. S., 2020, "3d Design Using Generative Adversarial Networks and Physics-Based Validation," ASME J. Mech. Des., **142**(7), p. 071701.
[120] Zhang, Z., Liu, L., Wei, W., Tao, F., Li, T., and Liu, A., 2017, "A Systematic Function Recommendation Process for Data-Driven Product and Service Design," ASME J. Mech. Des., **139**(11), p. 111404.
[121] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 2014, "Generative Adversarial Nets," Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, Dec. 8–13.
[122] Yang, Z., Li, X., Catherine Brinson, L., Choudhary, A. N., Chen, W., and Agrawal, A., 2018, "Microstructural Materials Design Via Deep Adversarial Learning Methodology," ASME J. Mech. Des., **140**(11), p. 111416.