Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 100 (2021) 31–36

www.elsevier.com/locate/procedia

31st CIRP Design Conference 2021 (CIRP Design 2021)

Extracting functional requirements from design documentation using machine learning

Haluk Akay, Sang-Gook Kim*

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

* Corresponding author. Tel.: 617-452-2472 ; *E-mail address:* sangkim@mit.edu

Abstract

Good design practice and digital tools have enabled industry to produce valuable products. Early-stage design research involves rigorous background study of large volumes of design documentation which designers must analyze manually, to extract functional requirements which are abstracted and prioritized to guide a design. Recent advances in Machine Learning, specifically Natural Language Processing (NLP), can be applied to enhance the time-consuming and difficult practice of the human designer by performing tasks such as extracting functional requirements from long-form written documentation. This work demonstrates how extractive question-answering by neural networks can be applied to design as a tool for automating this initial step in the design process. We applied the language model BERT, fine-tuned on question-answering, to identify functional requirements in written documentation. Limitations due to wording sensitivity are discussed and an outline for training a design-specific model is discussed with a MEMS product design case. This work presents how this application of AI to design could enhance the work of human designers using the power of computing, which will open the door for learning from big data of past product designs by allowing machines to "read" them.

Keywords: Natural Language Processing; Design Research; Question-Answering; Artificial Intelligence

1. Introduction

Design is a creative process, heavily dependent on the ingenuity of human designers to synthesize concepts that solve complex problems satisfying the world's needs with novel products and systems. The activity of creative ideation is a uniquely human ability, but there still exist tasks in the early-stage design process, which may benefit from human creativity working in harmony with the computational power and large memory of machines.

In the domain of solution prototyping and concept generation, designers have enjoyed the use of CAD, for decades now, to create digital models and quickly obtain performance assessments (static, dynamic, and thermal analyses) of their creations, before physical prototyping and testing, with ease. Recently, Machine Learning methods have been applied to physical object creation where a model is trained to synthesize a digital geometry with minimal designer guidance. Generative models have been demonstrated to create high-resolution 3D reconstructions of products such as furniture [17], as well as explore physical parameters to generate novel geometries of components such as airfoils based on maximizing certain performance objectives [4]. In contrast to applying Artificial Intelligence (AI) to generate design solutions, this work seeks to support human designers in the early-stage labor-intensive phase of the design research process which may benefit from the support of Machine Learning methods.

This work builds on the idea of Hybrid Intelligence in design [8], a framework for close collaboration between human designer and machine where, rather than replacing the role of the designer with generative models, the tasks which humans find difficult, time-consuming, and require considerable experience, are facilitated with Machine Learning-based methods, allowing designers to practice the creative aspects of their craft with more focus. Before novel solutions can be ideated by a designer, a problem must be identified. Although terminology may vary based on different schools of design, the importance of accurately identifying functional requirements in Axiomatic Design Thinking [12] or in the Product Design and Development process [15] is universally significant in initializing a rigorous design process. Practitioners of design research spend

10.1016/j.procir.2021.05.005

substantial resources on conducting thorough background research with the aim of surfacing functional insights which may translate to concrete requirements. Griffin et al. estimated that 98% of a product domain's functional information will only be surfaced after 25 hours of stakeholder interviews [6], generating hundreds of pages of transcripts that must be analyzed to surface functional information. Equivalent lengths of academic papers and theses in scientific domains must also be read and dissected to assemble a thorough corpus or text body of design research.

Processing massive amounts of textual information to surface functional requirements may be prone to inefficiency, inaccuracy, and bias when tasked to a human designer. These challenges increase when significant time has passed between data collection and analysis, or when design research projects are passed between human collaborators. Likewise if a design research process is paused and re-started by a new collaborator with low topic familiarity, the task becomes increasingly challenging.

This paper presents an application of Machine Learning methods, specifically in the area of Natural Language Processing (NLP), to develop a process for automating the task of extracting functional information from long-form textual documentation. This work is conducted with the aim of providing a tool for design researchers so that they may focus their craft on creative tasks currently outside the scope of AI's capabilities, and so that AI may be responsible for repetitive computationally-intensive aspects of their process which are difficult for humans.

2. Background

2.1. Language Representation

While humans are able to read words, phrases, sentences, and paragraphs to gain a contextual meaning of information being communicated, machines gain no such abstracted meaning from alphabetical characters. In 2003, Bengio et al. proposed an AI framework for representing language to machines based on a probabilistic model to essentially convert words to vectors [3]. This framework demonstrated how embedding elements of language (words) in a distributed feature vector positioned in multi-dimensional space could create a notion of semantic similarity between words.

One of the first neural network-based models to successfully execute this word vectorization task is known as *Word2vec* [10] in 2013. *Word2vec* was trained on word-prediction tasks on a large corpus (text dataset) such that the model learned a set of parameters that could represent words in high-dimensional vector space with remarkable results. Dimension-reduction techniques can be used to visualize how these language representations are spatially located, an example of which is shown in Figure 1. Vector embeddings of words closely mirrors semantic meaning; this is apparent in the visualization, where the words for numbers are closely clustered, as are geographic words, and so on. Furthermore, simple vector arithmetic similarly captured

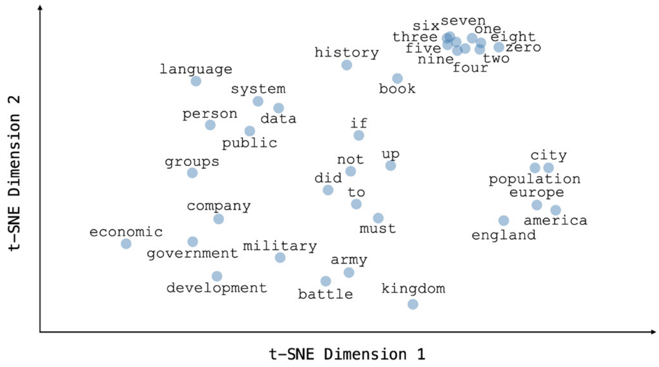

Fig. 1. The semantic meaning of language can be encoded into feature vectors where similar words occupy closer positions in multi-dimensional space. Using dimension reduction methods like t-SNE, this space may be visualized in 2D, shown above, adapted from [9]

semantics. For example, if the vector for "man" was subtracted from the vector for "king," the resultant vector was more closest to "queen" than any other word vector. The model was able to capture the contextual meaning of words such that the feature vectors for more similar words had higher cosine similarity.

As neural networks grew more sophisticated, and "deeper" with hardware able to efficiently train an increasing number of parameters on even larger datasets, the capabilities of NLP models intensified as well. In 2017, the framework for a novel neural network architecture called the *Transformer* [16] was proposed by researchers at Google. The *Transformer* architecture's specialty is its use of the *Attention* mechanism, which gives the model the ability to intelligently place greater weight on more important elements in a sequence, or rather "pay attention" to more relevant words in linguistic context.

Following the *Transformer* architecture, in 2018 Google released a novel language model *Bidirectional Encoder Representations from Transformers*, or BERT [5]. BERT was built using the *Transformer* architecture, but was also unique in that rather than processing language from left-to-right or right-to-left, it processed sequences bi-directionally to gain a comprehensive understanding of context. BERT immediately outperformed other similar language models and set new records in various NLP benchmark performance tasks, such as question-answering, which will be analyzed in detail in this paper.

Perhaps the feature of BERT which makes it most useful to the scientific community was the method in which it is trained. Training is broken into two phases: "pre-training" and "fine-tuning." The pre-training phase allows BERT to learn parameters used for producing vector representations of language through two tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). MLM involves randomly "masking" certain words in a span of text, and training BERT to correctly predict the masked word. NSP trains BERT on understanding sentence-level relationships through the binary prediction task of whether two sentences occur consecutively in a context. Pre-training is conducted over datasets totaling approximately 3.3 billion words, a computational resource-intensive feat, which Google performs itself, releasing the pre-trained model parameters to the scientific community.

2.2. NLP for Design: Task-specific Analysis and Generation

The field of Design has not benefited from developments in Machine Learning (ML) in the way NLP has been transformed. A major challenge for implementing ML methods for design tasks is the lack of sizeable design-relevant datasets available. In the face of low-resource domains (with respect to training data) such as design, alternative paths may be taken. The concept of *transfer learning* involves transferring knowledge of a learned source task to improve performance in a related but distinct target task [14]. The target task in this work has been identified as surfacing relevant functional information from long-form text documents. In the absence of many curated examples of design documentation with corresponding key functional information labeled, a more generalized language-related source task analogous to the target task may be identified for application to design.

In previous work [2], we have demonstrated how vector representations of language, obtained from neural-network based language models such as BERT, can be used to quantify metrics of system design, because the functional domain of design mirrors the semantic domain of language in simple cases. It was shown how succinct natural language descriptions of various designs of products such as faucets or refrigerators may be vectorized to accurately measure if the designs were functionally coupled or independent, purely based on the semantic similarity between descriptions of a matrix of requirements and solutions in simple systems. This was evidence of the relevancy of adapting NLP models to use in processing design documentation.

Alternatively, in the absence of available design task-specific data, synthetic data may be generated to use for the task at hand. In other previous work [1], we have demonstrated how generative text models, similar to those used when suggesting auto-complete options on social messaging apps, can be used to complete a prompting "seed" design statement with the aim of building a dataset of different types of design descriptions. This dataset was used to train a binary classifier to identify if an unlabeled statement described a problem or a solution.

Although large design-specific datasets may be rare, methods in Machine Learning and specifically NLP present an opportunity to transform the field of design. When early-stage design is considered, a significant amount of information is captured in textual documentation, before physical components and modules are prototyped. It is at this stage of design that NLP models may be applied; this paper describes how the challenge of extracting functional requirements from design texts may be approached using Machine Learning and NLP.

3. Method

3.1. Surfacing Information

The motivating problem statement can be described as: given long-form textual documentation, surface the functional domain information from within, for a human to subsequently prioritize and abstract to formal requirements of an eventual design. This involves identifying the highest-level functional requirements of a design given documentation. In order to be able to automate this process, a module with the following critical ability must be developed: given a span of text as contextual input, identify functional information, and return this as output.

The task of identifying functional information in linguistic context can be carried out *extractively* or *abstractively*. Extracting functional information from a sequence simply involves returning a sub-sequence from the original context. The subsequence is of shorter length, and has a higher "density" of information of interest. This is akin to a human analyzing a document and highlighting certain lines which are interesting to the reader. Abstracting functional information from a language sequence is a generative process, where based on information of interest in the original sequence, a novel sequence of shorter length is succinctly composed. This is akin to a human reading a document and summarizing the interesting portions in their own words.

From a feasibility standpoint, extracting function from context is the simpler task; just the start and end positions demarcating the functional information in the original context must be identified. In contrast, abstracting function from context requires a generative text module which adds unwarranted complexity to the solution, with just benefit of more succinct summaries. Futhermore, extracted functional information preserves the original ordering and identity of elements in a sub-sequence, which provides an opportunity for applying recursive methods for more structured decomposition in the future. This paper therefore focuses on an extractive solution.

With a rule-based system, an algorithm could theoretically be developed to satisfy the goal of extractively highlighting functional information. Such an algorithmic model would require a logic system (series of if – then statements) to parse the input text. The insurmountable challenge of applying such logic in practice is the diversity of style that humans use to express themselves verbally (in user interviews) or describe concepts formally (in scientific papers). The advantage of machine-learning-based NLP is that a model can be trained to extract functional information by learning from examples of context and extracted function. Because a dataset of adequate size does not exist, an NLP model trained on a source task analogous to the target task can be applied to extractively surface functional information from context. The following section describes the language task of question-answering, an analogous source task which may be applied to design.

3.2. Mechanics of Question-Answering

Question-answering is a benchmark language processing task, analogous to the target task of extracting function from context, but more generalized in nature. Given context, the answer to any question posed must be extractively returned. For application to design, this puts great importance on posing the correct question which will provoke the resultant answer to contain functional information. The task of extracting an answer from context involves identifying the sub-sequence, contained inside the context, which contains the answer to the question posed. Mathematically, this means finding the indices denoting the start position i, and the end position j, of the span of text defining the answer, where $i \leq j$. The context is broken, or tokenized, into a list of word-units which correspond to words, punctuation, or acronyms. Figure 3 in Section 4 indexes the context by word-unit.

A neural network can be trained to perform this extractive question-answering task. A measure of performance in this task is test accuracy on the Stanford Question Answering Dataset (SQuAD) [11], which is a set of 100,000 crowd-sourced examples of contexts, questions, and correct answers. The language model BERT, previously introduced, demonstrated state-of-the-art performance on this metric at the time of its release. With regard to NLP tasks, BERT is a multi-purpose tool due to the two-step process in which it is trained. An initial pre-training step provides a Transformer Neural Network with a set of learned parameters which can represent language with vector embeddings. For each word in an inputted language span, a vector containing that unit's semantic information is returned.

During fine-tuning of BERT on the SQuAD dataset, new learnable parameters are introduced which are trained to return the probability of a given word in the context being the start of the answer P_i and the probability P_j of being the end of the answer. The example questions, contexts, and answers in the SQuAD dataset are used to optimize the model's parameters to eventually perform extractive question-answering rivaling human reading comprehension. This model may be applied to design to extract functional requirements from context, if the correct inputs are used.

3.3. Application to Design

The source task of question-answering has analogous inputs and outputs to the target task of functional information extraction, except with the additional input of the question Q. Directly applying a question-answering model, such as BERT fine-tuned on SQuAD, requires a question to be worded in a way that $[i, j]$ bound the position of functional information in context.

As a basis for wording such a question, we can borrow from various schools of design theory. In Ulrich's process for user research as a step of Product Design and Development [15], the primary guideline for identifying design requirements is eliciting information that describes "*What*, not *How*." In Axiomatic Design thinking, Suh defines the functional domain as representing "What we want to achieve" [13]. This declarative statement describes the desired output from the question-answering model. It may be modified to be a question input by changing the sentence type to that of an interrogative, to read: `What do we want to achieve?` The wording of the question may be further generalized, because the tensed Verb Phrase [^{VP}`we want to achieve`] introduces specific thematic roles such as an agent "`we`" that is distracting from the desired functional information being sought. The Verb Phrase may be replaced by a more generalized semantically synonymous Noun Phrase

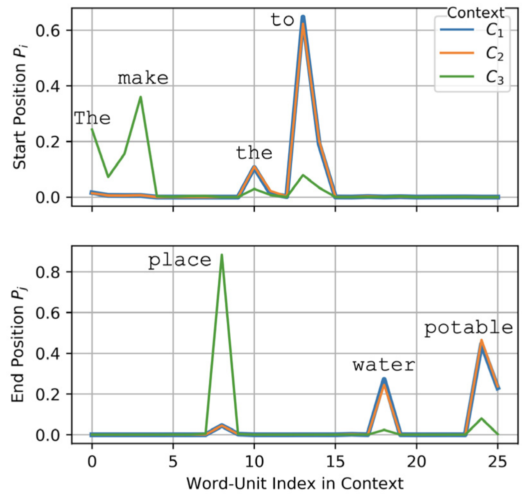

Fig. 2. Answer start and end position probability distributions across word-units in each context, to the question "`What is the goal?`"

[^{NP}`the goal`]. The question can be rephrased to read: `What is the goal?`

The wording of the question is important in that differently worded questions asking for synonymous information may yield different answers. In particular, the choice of the word for the object of the interrogative sentence, which is "`goal`" above, can trigger different responses based on context. This sensitivity is shown in the following test. Three nearly similar contexts C_{1-3} are considered. Each context contains two sentences, each of which includes different functional information. One word, boldfaced below, is synonymously interchanged among the three contexts below:

C_1 The design must make the world a better place. The **aim** is to purify dirty water so that it becomes potable.

C_2 The design must make the world a better place. The **goal** is to purify dirty water so that it becomes potable.

C_3 The design must make the world a better place. The **requirement** is to purify dirty water so that it becomes potable.

Three questions Q_{1-3}, identical except for one word, are posed. The resulting answers for each context C_{1-3}, extracted by BERT-SQuAD, are shown below.

Q_1 What is the **aim**?
$A_{1,1}$ to purify dirty water so that it becomes potable $A_{1,2}$ to purify dirty water so that it becomes potable $A_{1,3}$ make the world a better place

Q_2 What is the **goal**?
$A_{2,1}$ to purify dirty water so that it becomes potable $A_{2,2}$ to purify dirty water so that it becomes potable $A_{2,3}$ make the world a better place

Q_3 What is the **requirement**?
$A_{3,1}$ the design must make the world a better place $A_{3,2}$ the design must make the world a better place $A_{3,3}$ to purify dirty water so that it becomes potable

Despite the contexts and the questions being nearly synonymous, the answer span extracted does depend on the wording of both. The probabilities of a given word being either the start or end position of the answer to the question Q_2 are calculated based on the BERT-SQuAD model outputs, and are plotted as a distribution over the context length, shown in Figure 2.

Language models are sensitive to even small adjustments in question wording. Q_1 and Q_2 result in similar answers, different from that of Q_3 which uses the less generalized key word "`requirement`." Furthermore, the models are also sensitive to keyword matches between the question and context. For example, without any shared words in these sequences between Q_3 and C_{1-2}, the extracted answer is from the first sentence in the context. However, when the `requirement` is specified in the second sentence C_3, the extracted answer shifts to the second sentence. This is true for each combination where a keyword match occurs, in $A_{1,1}$, $A_{2,2}$, and $A_{3,3}$. Keyword matches can potentially bias the extracted answer span. As a result, the most generalized wording of the question, Q_2 is used in the experiments in this work. A framework for eliminating the possible bias stemming by removing the need to pose a question entirely is suggested in the discussion section of this paper.

4. Experiments

In order to demonstrate the efficacy of applying question-answering to surface functional information, a case study from a complex real-world design problem is considered. In the field of Micro-Electromechanical Systems (MEMS), vibrational energy harvesting is such a complex problem. Natural linear resonance ω_0 scales inversely with mass m, as shown by the fundamental relationship $\omega_0 = \sqrt{\frac{k}{m}}$. Simply because many naturally occurring vibrations are low frequency (>100Hz), it is a challenge to create a micro-scale device to operate in this range.

Two abstracts of papers describing designs of micro-scale vibrational energy harvesters are analyzed. Excepts from academic literature are used because of the expectation that they explicitly state the key functional requirement of the design being presented. Furthermore such texts may be publicly shared whereas actual design specifications from industry are less structured and often protected by stringent intellectual property restrictions. The first abstract (A) [18] describes a buckled-beam oscillator design, and the second (B) [7] describes a non-linear doubly-clamped beam design. For each paper, the key functional information from the abstract are surfaced using extractive question-answering from the BERT-SQuAD model. The pre-trained neural network parameters assess a probability distribution of the most likely positions of the start and end indices of the answer to be returned, similar to the scoring visualized in Figure 2. For these experiments, the context C is the entire abstract passage, and the question Q posed is `What is the goal?`. The answer spans A of which the model has the highest confidence are shown, highlighted, in Figure 3, where the abstracts are indexed according to word-units to illustrate how the context is tokenized by the model. The two answer A spans of surfaced functional information are shown in (A) and (B) below.

Fig. 3. Key functional information is highlighted in red in the design context, showing the span demarcated by BERT as having the maximum likelihood of containing the answer; context is indexed by word-unit

(A) lowering the operating frequency window below 100Hz for the first time at the MEMS scale

(B) overcomes the limitations of conventional linear resonance beam-based piezoelectric energy harvesters in terms of power bandwidth and power density

The purpose of this experiment is to demonstrate the ability to take in a long-form textual context and return a single key functional requirement. This validates a critical part of what can be a powerful framework of design document decomposition if NLP-based question-answering is applied recursively to exhaustively return all functional information contained within text.

5. Discussion

The use of extractive question-answering results in subsequences from the original which provide opportunity for further decomposition of design context. The surfaced spans of functional information correctly identify high-level functional requirements of vibrational energy harvester design. Background information in the abstracts are correctly disregarded, as are portions which detail the solution domain instead of functional requirements. By referencing the outputted surfaced spans from the initial processing, which appear to be high-level functional information, more detailed sub-requirements may potentially be mapped through recursive use of this method.

The main drawback of directly applying a question-answering model to extractively surface functional information is that a question must be posed. In Section 3.3, the potential for bias arising from wording of the question was identified as a consequence of introducing a question input. Question-answering was identified as an appropriate source task from which a pre-trained model could be applied to the target task of surfacing function because of the sizeable data available for this standard NLP task, and the lack thereof in design. However, if such a dataset of context – extracted functional information examples could be crowdsourced or generatively created, pre-trained language models such as BERT could be easily fine-tuned directly on the task targeted by this paper, bypassing the need for a question input. The learnable parameters introduced during fine-tuning of BERT would now be trained directly on producing high scores for functional information, without a question input. The experiments in this paper demonstrate the value of this application, and the challenges with wording sensitivity can be addressed with design-specific data. A limitation of using academic design documentation for testing is that industry design specification are generally less structured and documented with less elaboration. When analyzing such documentation, additional processing may be required before the method outlined in this paper can be deployed.

6. Conclusion

The goal of this work is to enhance a designer's practice by providing a tool to automate the time-consuming, tedious steps in the design process using Machine Learning methods such that a human designer may focus their creative energy on their craft which cannot be performed by machines. The task of surfacing functional information from large corpora of design research documentation (prior to abstraction to formal requirements by a human) was identified as such a step apt for automation. In the absence of an annotated dataset specific to this design task, a Natural Language Processing (NLP) model trained on extractive question-answering was demonstrated to perform the task of functional extraction on engineering design publications in the field of MEMS.

This method is superior to simple keyword-searches widely used in design-related information retrieval tasks because rather than matching sequences of characters to a query, the BERT language model seeks to identify spans of text containing answer information. This ability is learned from a massive pre-training exercise teaching the model the meaning of language, followed by a detailed fine-tuning step where the model learns from examples of question-answering. The drawbacks due to sensitivity to wording stemming from the introduction of a question input necessitated by the question-answering model were discussed, and the process for bypassing the need of such an input through direct fine-tuning of a language model on a design-specific dataset was discussed. In conclusion, this paper presents a modular functionality for automating design information processing through the application of Artificial Intelligence in language processing to design, for future development to the goal of transforming the field of Design with AI methods in a manner similar to which many other fields have benefited over the past decade.

7. Acknowledgement

This work was supported by MIT/SenseTime Alliance on Artificial Intelligence, and the National Science Foundation (NSF) Leading Engineering for America's Prosperity, Health, and Infrastructure (LEAP HI) program, award number 1854833.

References

[1] Akay, H., Kim, S.G., 2020a. Design transcription: Deep learning based design feature representation. CIRP Annals 69, 141–144.
[2] Akay, H., Kim, S.G., 2020b. Measuring functional independence in design with deep-learning language representation models. Procedia CIRP 91, 528–533.
[3] Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C., 2003. A neural probabilistic language model. Journal of machine learning research 3, 1137–1155.
[4] Chen, W., Ahmed, F., 2020. Padgan: Learning to generate high-quality novel designs. Journal of Mechanical Design , 1–17.
[5] Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 .
[6] Griffin, A., Hauser, J.R., 1993. The voice of the customer. Marketing science 12, 1–27.
[7] Hajati, A., Xu, R., Kim, S.G., 2011. Wide bandwidth piezoelectric micro energy harvester based on nonlinear resonance, in: Proc. PowerMEMS, pp. 3–6.
[8] Kim, S.G., Yoon, S.M., Yang, M., Choi, J., Akay, H., Burnell, E., 2019. Ai for design: Virtual design assistant. CIRP Annals 68, 141–144.
[9] Mandelbaum, A., Shalev, A., 2016. Word embeddings and their use in sentence classification tasks. arXiv preprint arXiv:1610.08229 .
[10] Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 .
[11] Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P., 2016. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250 .
[12] Suh, N.P., 1990. The principles of design. Oxford University Press.
[13] Suh, N.P., 1998. Axiomatic design theory for systems. Research in engineering design 10, 189–209.
[14] Torrey, L., Shavlik, J., 2010. Transfer learning, in: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, pp. 242–264.
[15] Ulrich, K.T., 2003. Product design and development. Tata McGraw-Hill Education.
[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Advances in neural information processing systems, pp. 5998–6008.
[17] Wu, J., Zhang, C., Xue, T., Freeman, B., Tenenbaum, J., 2016. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling, in: Advances in neural information processing systems, pp. 82–90.
[18] Xu, R., Akay, H., Kim, S.G., et al., 2019. Buckled mems beams for energy harvesting from low frequency vibrations. Research 2019, 1087946.