FROM TEXT TO IMAGES: LINKING SYSTEM REQUIREMENTS TO IMAGES USING JOINT EMBEDDING

Chen, Cheng;
Carroll, Cody;
Morkos, Beshoy

University of Georgia

ABSTRACT

Smart manufacturing enterprises rely on adapting to rapid engineering changes while minimizing the generated risk. Making informed decisions related to engineering changes and managing risks against unexpected costs requires more information to be extracted from limited data. However, limited information in early-stage design can come in many forms, namely text and images. The development of innovative design tools and processes to link multisource data together is essential to assist designers in building model-based engineering (MBE) systems. However, the formal computational linking of multisource data is yet to be realized in MBE. We propose a framework to implement transfer learning and integrate domain specific knowledge to bridge this information gap. A synthetic dataset is created using web scraping techniques based on keywords extracted from the requirements. Requirement-image pairs are used to fine tune a contrastive language-image pretraining model to acquire domain knowledge. The results demonstrate how the content of images can be used to indicate all affected requirements for tracing engineering changes in a complex system.

Keywords: Requirements, Product Lifecycle Management (PLM), Complexity, Engineering Change, Contrastive Language-Image Pre-training (CLIP)

Contact:
Morkos, Beshoy
University of Georgia
United States of America
bmorkos@uga.edu

Cite this article: Chen, C., Carroll, C., Morkos, B. (2023) 'From Text to Images: Linking System Requirements to Images Using Joint Embedding', in *Proceedings of the International Conference on Engineering Design (ICED23),* Bordeaux, France, 24-28 July 2023. DOI:10.1017/pds.2023.199

1 INTRODUCTION

Advances in smart factories, coupled with the disruptions of supply chains, have created a turning point for manufacturing industries in terms of improving design resilience. Model-based approaches have become increasingly prominent in manufacturing applications as machine learning is used more extensively in design automation to assist in early-stage design. Despite the improvement in manufacturing resilience, we have not fully exploited semi-structured (i.e., Sensor data) or unstructured data (i.e., Text in natural language) for design improvement. Designing a complex system that takes advantage of existing data would require the development of new tools and processes Castet (2017). This means that domain experts should actively develop different design techniques to resolve dynamic design issues. How to process multi-source data to aid knowledge acquisition during the design process has received attention in recent years from other industry environments, such as the process industry, the manufacturing execution system , and the cyber-physical system. Implementing different types of models in design and manufacture continues to present challenges and opportunities.

Current approaches to managing design changes still present unique challenges and are heavily dependent on domain experts. In current systems of tracking design changes, domain experts are relied upon to identify and interpret each change instance, which can be time-consuming and prone to errors. To mitigate this deficiency, various design change propagation models are developed, including complex network approaches Hein et al. (2021) and design structure matrices Stirgwolt et al. (2022). Besides modeling engineering changes within requirement documents, few approaches exist to track design changes across domains. Data analysis across multiple domains presents a number of unique challenges. First, due to confidentiality, few design documents are publicly available or can be used for benchmark data sets. Second, it is not well understood how to extract meaningful information from a variety of unstructured data sets. As natural languages possess greater lexical diversity, it is not well understood how to extract domain knowledge from latent content from language models. In requirement management (RM), the corresponding image data sets are rarely documented. To compensate for the missing information, image scraping techniques can be deployed to gather online images based upon the keywords provided. This study presents a framework for bridging the information gap on how to trace design information across design domains by combining textual and visual representations.

A promising solution for generating correlation between text and image is to implement joint embedding, which is a machine learning technique designed to capture the association between different types of data sets. Early studies in this area employed different approaches to analyze texts and images in relation to each other. In recent years, a contrastive language-image pre-training (CLIP) model was developed to deal with out-of-distribution predictions by using zero-shot learning Radford et al. (2021). For typical image and text classification problems, both training and test data sets are from the same distribution. In contrast, the CLIP model uses a dot product to learn the joint embedding space and perform zero-shot prediction on images with truly out-of-distribution samples. Several pre-trained CLIP models containing general knowledge can be further fine-tuned to learn domain-specific designs. Because of these factors, we selected the CLIP model to learn the correspondence between requirements documents and images.

This study proposes a method to address the current information gap of multi-source data issues within MBE. We present a framework that can learn domain-specific knowledge by building correlations between images and texts. As a result of this method, engineers can visualize the interconnections between subsystems and manage the propagation of changes. It is important to note that in this paper we don't focus on a particular application as the research is still in the beginning phases and we are attempting to provide a general framework. The study has three major contributions:

1. An alternative method is presented for filling in the missing visual information relating to each requirement.
2. A framework is developed for leveraging transfer learning and establishing correlations between design requirements and images of physical components.
3. This study contributes to the understanding of the factors affecting the fine-tuning of a model and its ability to predict mechanical design outcomes.

2 BACKGROUND

We discuss the technical background of three topics relevant to this research: design requirements in product lifecycle management (PLM), image scraping, and joint embedding literature.

2.1 Requirements management in PLM

This study was prompted by earlier research on requirements management. As requirements documents are collaboratively developed based on different domain knowledge, implementing and tracking engineering changes across various fields can be problematic. A requirement risk is a potential mismatch between stakeholder expectations and the outcome of a project. The evolution and management of design changes can be challenging in a team environment Morkos et al. (2019); coupled with distribution management, a process for approving engineering changes for documents within PLM Saaksvuori and Immonen (2008) where engineers spend 15-40 percent of their time searching and checking information within PLM systems while making difficult trade-off decisions to comply with customer requirements, adds to those difficulties. There is evidence that requirements may not always correlate well with other populated design documents within a project Morkos et al. (2010). Such discrepancies may result in information loss during the propagation of engineering changes. Because each engineering change can propagate through functional or non-functional requirements, an automated requirement change propagation prediction (ARCPP) tool was developed to simulate the affected requirements based on keywords Morkos (2012). With the integration of RM tools into PLM Violante et al. (2017), a product-centric approach becomes increasingly critical to trace design information related to the physical product. In contrast to functional requirements, a case study demonstrated that engineering design decisions are often influenced by non-functional requirements in the automotive OEM industry Shankar et al. (2012). To mitigate unexpected changes, the complexity of design change propagation necessitates the development of a volatility measure to estimate the reactions of engineering changes to the basic change instances Hein et al. (2021, 2022). Further, a topic model approach reduces the risks associated with unexpected change propagation by dividing requirements documents into interpretable groups from which propagation can be estimated Chen et al. (2021). Despite these advances, the propagation of requirements must ultimately reflect on the physical component, and such information links are essential for smart manufacturing.

2.1.1 Requirements in smart manufacturing

For smart manufacturing to achieve higher production, higher quality, and cost-effective rates, unstandardized or unstructured data such as requirements must be reevaluated Wang et al. (2021). With the integration of data science and manufacturing, the direction of requirement management in PLM will undergo a paradigm shift. Future cloud manufacturing (CMfg) will be dependent on customers' service requirements Tao et al. (2015), such as decentralized production 3D printing. As blockchain-based PLM advances, individual designers will be able to secure and maintain a record of their requirements documents with other stakeholders, as well as other design information, including text, images (i.e., drawings) and 3D models. Adding a new block to the network will allow all systems from stakeholders will verify and update synchronously on historical records. As design manufacturing requirements continuously evolve, the direction of product requirements will increase in variety, quality, and service while maximizing the satisfaction of customers.

2.1.2 Challenges in multi-source data and PLM

Current manufacturing sectors still face several major data challenges in implementing PLM (i.e., product design, manufacturing, and customer service). The PLM front-end is a centralized network where vendors manage all product information. Although data collection has grown rapidly, the "Big Data" concept and technique still have limited application in the PLM domain Li et al. (2015) resulting in current solutions still relying on software for managing, analyzing, and simulating engineering changes. For instance, image files are often included with technical notes in a folder tree sent to suppliers David and Rowe (2016). The automatic correlation of images with natural language, however, remains a work in progress for manufacturers. Data driven approaches in manufacturing are often hindered by the lack of open datasets. For manufacturers, however, the automatic correlation of images with natural language remains a work in progress due to privacy, biases, and security concerns. Further, the lack of open access

https://doi.org/10.1017/pds.2023.199 Published online by Cambridge University Press

datasets will make it difficult to visualize the paired design changes with images of the corresponding mechanical designs. Exploiting collected data from multiple sources is a challenging task that must be accomplished to satisfy the requirements and ensure the success of industrial projects. Various industries may utilize different formats or standards for production and design. Often, manufacturers are unaware of how to obtain and store design information Li et al. (2015). It is equally challenging to explore unstructured datasets from various types of data, such as natural language, images, and voice in addition to structured datasets. To integrate unstructured data into current MBE systems, more data-driven approaches and strategies should be developed to provide cost-effective solutions.

2.2 Image scraping

Image retrieval techniques are used to correlate the most related images from a large database, which are widely used in social web applications. As image data is not always accessible, image retrieval is used to search and collect images from the Internet. Image retrieval can be divided into three categories: text-based image retrieval (TBIR) such as Google, content-based image retrieval (CBIR), and semantic-based image retrieval (SBIR) Van Gemert (2003). Through a query, text-based retrieval can be simplified into a keyword-based search, and the returned results can be visualized as images with semantic similarity Datta et al. (2008). Combined with TBIR systems, web scraping is a technique which can collect information from Google. Such scraping tasks include reading HTML links, image files, and audio records. The challenge of collecting information online involves complicated website structures and bot access as known as the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). Many libraries are built to aid designers to automatically download images based on queries, such as the Selenium, the Google-image-download, and the Beautiful Soup libraries. These tools allow users to search and modify the raw content through appropriate parsers using Python. Based on targeted image URL links, information is downloaded for downstream joint embedding analysis.

2.3 Joint embedding

As digital threads become more prevalent in industry, computer vision techniques are making their way into other fields, such as manufacturing. Joint embedding involves learning different types of data, such as images, texts, speech, and video, into a common latent vector space. Common research challenges in this area are topics such as bi-directional image and text retrieval, and image captioning. Canonical correlation analysis is often used to determine the linear combination of image and textual data that maximizes the correlation between image-text pairs at a high memory overhead Hardoon et al. (2004). Using this method, correlations can be built between images (i.e., engineering drawings or photos) and text documents (i.e., requirements documents). A variety of loss functions were developed to overcome the memory cost problem, including triplet loss function Schroff et al. (2015) and multi-class N-pair loss Sohn (2016). The CLIP model, trained on 400 million images and texts from publicly available data sets, is a supervised zero-shot learning technique. The zero-shot learning approach is characterized by the fact that no classes are presented during testing that were presented during training Socher et al. (2013). As the CLIP model can be boiled down to image and text embedding during training, this structure allows diverse types of neural networks to be applied to the image and text encoders. Several common building blocks are used in the construction of a text encoder, including the BERT or transformer-base models Sanh et al. (2019). Meanwhile, a cosine similarity is then calculated among images and texts and evaluated with a chosen loss function. As a part of testing, the zero-shot CLIP model provided more reliable results for out-of-distribution image prediction. Often, neural network models are divided into trained and tested groups based on the same distribution of data sets for downstream tasks. Without the assumption that the new testing images come from the same distribution, it is difficult for models to differentiate the new testing images into new classes. Various techniques are developed to overcome such challenge of out-of-distribution images, including, few-shot learning Sung et al. (2018), and N-shot learning Sirinam et al. (2019). In managing requirements documents, utilizing these techniques can be helpful in controlling unexpected design changes that may occur in requirements documents.

3 EXPERIMENT

We describe how a synthetic image dataset can be created using existing requirements documents obtained from industry projects to enhance the efficiency of engineering change management. Figure 1 shows the pipeline of the proposed pipeline using a fine-tuned CLIP model.

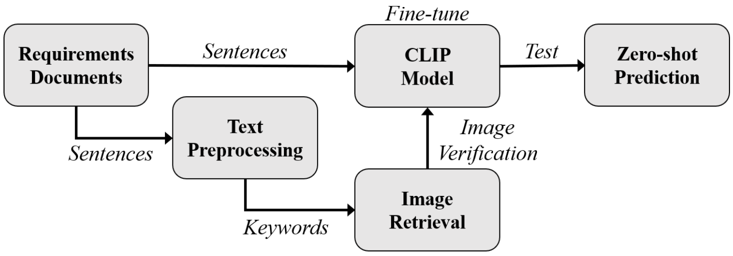


Figure 1. Pipeline of proposed framework

3.1 Text preprocessing

The purpose of a text preprocessing step is to extract the most relevant keywords for image scraping, as certain words contribute more to connecting visual ideas than others. Our first step is to eliminate all non-alphanumeric characters and stopwords (e.g., "shall," "etc,." or "must"). The remainder of the corpus consists of nouns, verbs, and adjectives filtered by part of speech (POS). A previous study determined that nouns and verbs can be used to describe the physical architecture and functional characteristics of projects Hein et al. (2018). Search queries are based on these keywords.

3.2 Synthetic image dataset

An industrial requirement dataset describing the design of a pipe assembly line is implemented in this study Morkos (2012). The study incorporates requirements documents designed for a pipe threading station from an industrial project. In total, 350 requirements are included, containing both functional and non-functional requirements. Project details include topics such as design specifications, project descriptions, equipment supplies, installation procedures, and shipping. After text preprocessing, each sentence is reduced to phrases for retrieving online images.

A version of the search model is implemented to scrape images from online text searches. As the order of keywords does not significantly affect the search results, queries are automatically sent to online servers to retrieve images as a browser user. The original resolution image is downloaded locally using several techniques and packages, including BeautifulSoup, Request, LXML XML toolkit, and regular expressions (RegEx). As images can be extracted from several sources, a verification procedure is implemented to ensure that all images are accessible through the Pillow library. For example, some images cannot be downloaded from an online PDF document or a website protected by anti-bot tools such as CAPTCHA. This requires manual verification to replace irrelevant images. Because images come in a variety of sizes, we use the resampling LANCZOS[1] filter to rescale each image into a 300×300 pixel size. By doing so, we avoid losing information on the edges.

3.3 CLIP model

As the number of image-requirement pairs is relatively small, directly training the model on CLIP might not be effective. Instead, transfer learning allows the model to integrate previous knowledge with domain-specific knowledge. In this experiment, we compared the performance of pre-trained and fine-tuned CLIP models. Conducting an overall evaluation of zero-shot prediction accuracy is beyond the scope of this study.

3.3.1 Prediction on pre-trained CLIP

A pre-trained model is typically trained on a large dataset that is intended for general use. Pretrained models are typically generated using a variety of large open-source datasets. The direct application of a pretrained model to a domain-specific task could provide a broadline assessment of performance. Using a pre-trained CLIP model, we select an unforeseen image that is closely associated with the industrial design to predict the most likely requirements. Similarly, the new image is subjected to the same filters before being passed on to the image encoder. The transformer model is used to encode requirements. By utilizing zero-shot predictions, the most relevant requirements are identified. Comparing to the baseline model, a fine-tuned prediction model should provide improved performance for domain-specific tasks.

---

[1] https://pillow.readthedocs.io/en/stable/releasenotes/2.7.0.html

https://doi.org/10.1017/pds.2023.199 Published online by Cambridge University Press

3.3.2 *Prediction on fine-tuned CLIP*

The requirement-image pair is first randomly shuffled into a training set with a batch size of ten. To make a zero-shot prediction, out-of-distribution images should be selected from variant designs of manufacturing pipe stations. The total number of epochs is twenty. Image and text losses are calculated individually using cross-entropy. The Adam optimizer is implemented with a learning rate of 5e-6 and decoupled weight decay regularization of 0.4 for all layers. These values are adjusted based on analysis and evaluation to fine-tune the hyperparameters. Similarly, the same prediction procedure is implemented to output the top five requirements with their probabilities.

4 RESULT AND DISCUSSION

4.1 Synthetic dataset

As the created synthetic image dataset contains various types of images, Figure 2 presents several search results from the industrial trial project. In response to different search terms, the collected images include photographs, drawings, and document scans. Note that not all the returned images accurately reflect the details of a search query, and we assume that the top image represents the most relevant results. When the first image is not available, the next image is manually downloaded. Further, some images may not capture the meaning of the requirements due to ambiguous words and short search queries. In such cases, we consider some images to be noise. For example, in Figure 2 (e), upon sending the query "threading, line, Bucker, station," the retrieved result depicts a picture of a train departing the Buckner station.

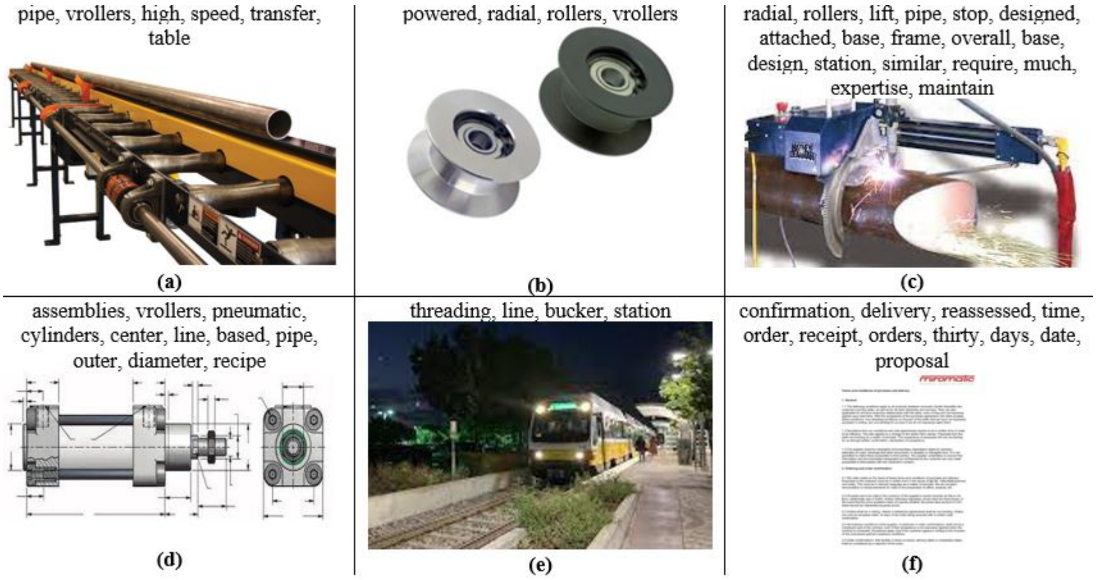


Figure 2. Samples of collected synthetic image datasets with requirement keywords

Although a fine-tuned model may not learn valuable knowledge from irrelevant images, it is still possible to obtain limited useful information. In Figure 2 (f), many search queries related to non-functional requirements contain the words "proposal," "description," "specification," and "criteria," which result in a screenshot of a document. Though CLIP models may not capture detailed content from images, they may still recognize these keywords as representing the concept of documents. In context-rich design projects that include more image documents, designers may fine-tune the model or combine it with additional neural networks to further extract textual information from images.

Similar search queries might return the same image. As an example, after word preprocessing, query numbers 158 ('box', 'end', 'threading', 'station', 'idler', 'radial', 'rollers', 'vrollers') and 167 ('box', 'end', 'threading', 'inspection', 'station', 'idler', 'pipe', 'radial', 'rollers') have the same image result. As both sentences contain many similar words and describe similar objects, the study uses the same pictures to represent both requirements.

https://doi.org/10.1017/pds.2023.199 Published online by Cambridge University Press

4.2 Improvements in design

With the increasing number of epochs, the total loss decreases, as shown in Figure 4. The loss function is averaged based on the cross-entropy loss between the image and the text. As a result of model fine-tuning, training loss is significantly reduced (around 65%) after 10 epochs. As a trade-off decision, fine-tuning a model could result in the loss of transfer knowledge and the acquisition of more domain-specific information while increasing the number of epochs. Thus, we employed an early stop strategy during the fine-tuning process to prevent overfitting. The CLIP model stops learning requirement-image pairs after 20 epochs and provides the most interpretable results. It is important to recognize that fine-tuning increases the risk of losing previous knowledge and gaining excessive domain-specific knowledge.

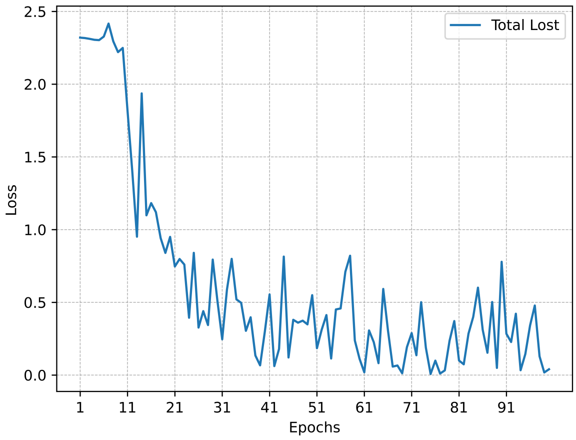
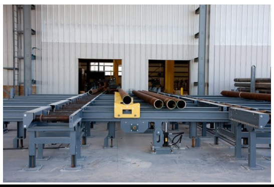


Figure 3. Variation of training error with increasing epochs

For the purpose of maintaining confidentiality, each requirement sentence is represented by certain keywords, as shown in Figure 4. A fine-tuned CLIP model is tested using an out-of-distribution image from a variant design in Figure 4, which has certain similarities to the actual pipe threading stations.



Pre-trained Model		Fine-tuned Model	
Keywords	Percentage	Keywords	Percentage
'lift', 'pipe', 'entry', 'end', 'table', 'paddle', 'threading', 'conveyor', 'transfer', 'box'	10.23%	'vrollers', 'pipe', 'transfer', 'table', 'gravity', 'roll', 'towards', 'exit', 'conveyor'	32.66%
'station', 'bucker', 'structural', 'contructed', 'frame', 'members'	5.09%	'rail','assemblies','spaced','half','feet', 'spanning','length','transfer','table'	13.12%
'thirteen', 'table', 'line', 'threading', 'transfer'	3.62%	'pipe','rest','adjustable','pipe','stop', 'exit','conveyor'	11.98%
'project', 'description'	3.59%	'pipe','secured', 'vrollers','clamp','high', 'speed','transfer','table'	7.31%
'constructed', 'inch', 'structural', 'table', 'walls', 'tubing', 'transfer', 'quarter'	2.89%	'pipe','secured','vrollers','clamp','high', 'speed','transfer','table'	7.31%

Figure 4 & Table 1 An image of conveyor system[2] with model predictions

The pipe threading equipment outlined in the requirements document, as well as the storage equipment shown in Figure 4, contain several types of conveyor systems that can be potentially adapted from one to another. Rather than viewing this problem as a pure classification process, each requirement might

[2] https://steelstorage.com/resources/photo-galleries/

https://doi.org/10.1017/pds.2023.199 Published online by Cambridge University Press

correspond to multiple images or vice versa. Therefore, the zero-shot prediction method is employed to compute the probability for each requirement-image pair. Based on the results, the best result (10.23%) is considered the most relevant requirement for the pre-trained model. In contrast, the top prediction result from the fine-tuned model achieves higher accuracy by providing more relevant information. Upon interpretation, the improved results have a closer relationship to functional requirements pertaining to “pipe stations” or “transfer tables.” As the fine-tuned model can recognize the concept from images and find the most relevant requirements, engineers should determine the appropriate number of relevant requirements and make corresponding engineering adjustments.

A particularly interesting and noteworthy observation is the use of images that contain both image and text data. The image in Figure 5 is chosen as a challenge for the fine-tuned model recognizing shapes and text information simultaneously. The image depicts a conveyor ball transfer table, on which hardened carbon steel balls are used to replace rollers. In such images, the fine-tuned CLIP model did not result in significant performance improvements for the top prediction (2.5% improvement), as shown in Figure 5. In images that contained only photographic images and no text, the fine-tuned CLIP model was considerably improved (22.4% improvement).

In the pre-trained model of Figure 5, two distinct requirements resulted in the same keyword phrases after the pre-processing step. Depending on the design needs, adjusting the strategies in the preprocessing step can greatly affect the interpretation of the final result. Aside from the differences in the resulted requirements, both baseline and fine-tuned models recognize the new image as a transfer table. Upon interpretation, the fine-tuned model predicts requirements that are more closely aligned with the transfer table, both quantitatively and qualitatively.

Pre-trained Model		Fine-tuned Model	
Keywords	Percentage	Keywords	Percentage
‘threading’, ‘line’, ‘thirteen’, ‘transfer’, ‘table’	15.76%	‘pipe’, ‘pin’, ‘threading’, ‘station’, ‘transfer’, ‘table’, ‘towards’, ‘end’, ‘threading’, ‘inspection’,	18.03%
‘pipe’, ‘next’, ‘transfer’, ‘table’	7.65%	‘design’, ‘vrollers’, ‘many’, ‘similar’, ‘features’, ‘vrollers’, ‘tube’, ‘uses’, ‘exception’, ‘high’, ‘temperature’, ‘designs’	14.06%
‘pipe’, ‘next’, ‘transfer’, ‘table’	7.65%	‘threading’, ‘station’, ‘base’, ‘design’, ‘similar’, ‘stations’	9.63%
‘pipe’, ‘gravity’, ‘roll’, ‘transfer’, ‘table’, ‘towards’, ‘box’, ‘drift’, ‘threading’, ‘protector’, ‘station’	4.87%	‘pipe’, ‘gravity’, ‘roll’, ‘transfer’, ‘table’, ‘towards’, ‘bucker’, ‘station’	5.83%
‘transfer’, ‘table’, ‘designed’, ‘located’, ‘previous’, ‘next’, ‘operation’	3.41%	‘pin’, ‘end’, ‘blast’, ‘station’, ‘design’, ‘identical’, ‘box’, ‘blast’,	5.45%

Figure 5 & Table 2 An image of conveyor ball transfer table[3] with requirement predictions

The out-of-distribution images are selected from a variant design as indicated earlier. As similarities can be defined from different perspectives, the out-of-distribution images may take different forms. For instance, taking images of the same object from various angles with a variety of backgrounds may also be considered as testing images. As not all the mechanical components are symmetrical, different angles of the same part might have an impact on the predictions.

The study suggests that the proposed framework could potentially be used to visualize requirement traceability by taking images of various physical components. The most relevant requirements should be determined for each image and evaluated regarding engineering changes. Although the synthetic dataset contains some irrelevant images as noise, the fine-tuned CLIP model is still capable of learning useful information and improving out-of-distribution prediction.

[3] https://omtec.com/catalog/f1-conveyor-table/

https://doi.org/10.1017/pds.2023.199 Published online by Cambridge University Press

Through a synthetic dataset, the fine-tuned model can identify standard mechanical components from collected images. For specialized mechanical parts, the image obtained from the internet may not accurately reflect their physical components. A minor change in design may, however, be treated by an out-of-distribution prediction and not necessitate a new simulation. As requirements are often added or deleted during the reengineering process, designers need to repeat the analysis to achieve higher accuracy. The proposed process would allow engineers to realize the interconnection of heterogeneous data quickly and reduce human error in the design process. Future work should explore different rotation-invariant techniques to build a more robust model and integrate this framework into digital threads. Rather than using 2D images, 3D point clouds could be another future direction. Further, the fine-tuned model can be combined with augmented reality for industrial applications.

4.3 Study limitations

While there are benefits this approach provides to designers, it is important to note that it does not replace the domain engineering design knowledge and expertise. Ultimately, the designer will need to make a determination if the images are relevant and offer utility. The use of the images will also depend on the designer's familiarity with the working principle. This study does not consider the designer element of this study - which is necessary to ensure such a tool could be implemented in design practice.

5 CONCLUSION

We propose a framework for bridging gaps in and synthesizing multi-source data to facilitate knowledge acquisition and improve design efficiency. As image data may not always be available, we collected online images by using keywords filtered from requirements documents using POS tagging which aids in tracking engineering change propagation. To collect images from Google search results, a web scraping technique is used. Images are manually verified and modified according to the closest interpretation of requirements. The collected image dataset is verified and resampled to the same size. We demonstrate an improvement in model prediction by showing the top five most relevant requirements after fine-tuning the CLIP model. Testing images are selected from a variant design to assess the robustness of the model. The major contributions of this work are threefold. First, we provide a method for constructing a synthetic image dataset containing the physical components of requirements. Secondly, using transfer learning, we combine prior knowledge with domain-specific information to understand the connection between requirements and images forming out-of-distribution image datasets that can be identified for identifying and interpreting requirements. Third, the predicted results illustrate the performance and limitations of the models by indicating the most relevant requirements for invariant designs which engineers can determine which components are affected to minimize risks for a complex system.

Future work can be extended in several directions. Several CLIP model architectures and other industrial design documents should be considered to provide useful insights into different types of product designs through comparison of various model architectures with publicly available design documentation.

REFERENCES

Castet, C.e.a. (2017), "A point of view from mbse practitioners", *NASA JPL*.

Chen, C., Mullis, J. and Morkos, B. (2021), "A topic modeling approach to study design requirements", in: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 85383, American Society of Mechanical Engineers, p. V03AT03A021.

Datta, R., Joshi, D., Li, J. and Wang, J.Z. (2008), "Image retrieval: Ideas, influences, and trends of the new age", *ACM Computing Surveys (Csur)*, Vol. 40 No. 2, pp. 1–60.

David, M. and Rowe, F. (2016), "What does plms (product lifecycle management systems) manage: Data or documents? complementarity and contingency for smes", *Computers in Industry*, Vol. 75, pp. 140–150, http://doi.org/10.1016/j.compind.2015.05.005.

Hardoon, D.R., Szedmak, S. and Shawe-Taylor, J. (2004), "Canonical correlation analysis: An overview with application to learning methods", *Neural computation*, Vol. 16 No. 12, pp. 2639–2664.

Hein, P.H., Kames, E., Chen, C. and Morkos, B. (2021), "Employing machine learning techniques to assess requirement change volatility", *Research in Engineering Design*, Vol. 32 No. 2, pp. 245–269.

Hein, P.H., Kames, E., Chen, C. and Morkos, B. (2022), "Reasoning support for predicting requirement change volatility using complex network metrics", *Journal of Engineering Design*, Vol. 33 No. 11, pp. 811–837, http://doi.org/10.1080/09544828.2022.2154051.

Hein, P.H., Voris, N. and Morkos, B. (2018), "Predicting requirement change propagation through investigation of physical and functional domains", *Research in Engineering Design*, Vol. 29 No. 2, pp. 309–328.

Li, J., Tao, F., Cheng, Y. and Zhao, L. (2015), "Big data in product lifecycle management", *The International Journal of Advanced Manufacturing Technology*, Vol. 81 No. 1, pp. 667–684.

Morkos, B., Joshi, S. and Summers, J.D. (2019), "Investigating the impact of requirements elicitation and evolution on course performance in a pre-capstone design course", *Journal of Engineering Design*, Vol. 30 No. 4-5, pp. 155–179, http://doi.org/10.1080/09544828.2019.1605584.

Morkos, B., Joshi, S., Summers, J.D. and Mocko, G.G. (2010), "Requirements and data content evaluation of industry in-house data management system", in: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 44113, pp. 493–503.

Morkos, B.W. (2012), *Computational representation and reasoning support for requirements change management in complex system design*, Ph.D. thesis, Clemson University.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. et al. (2021), "Learning transferable visual models from natural language supervision", in: *International Conference on Machine Learning*, PMLR, pp. 8748–8763.

Saaksvuori, A. and Immonen, A. (2008), *Product lifecycle management systems*, Springer.

Sanh, V., Debut, L., Chaumond, J. and Wolf, T. (2019), "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter", *arXiv preprint arXiv:1910.01108*.

Schroff, F., Kalenichenko, D. and Philbin, J. (2015), "Facenet: A unified embedding for face recognition and clustering", in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823.

Shankar, P., Morkos, B. and Summers, J.D. (2012), "Reasons for change propagation: a case study in an automotive oem", *Research in Engineering Design*, Vol. 23 No. 4, pp. 291–303, http://doi.org/10.1007/s00163-012-0132-2.

Sirinam, P., Mathews, N., Rahman, M.S. and Wright, M. (2019), "Triplet fingerprinting: More practical and portable website fingerprinting with n-shot learning", in: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1131–1148, http://doi.org/10.1145/3319535.3354217.

Socher, R., Ganjoo, M., Manning, C.D. and Ng, A. (2013), "Zero-shot learning through cross-modal transfer", *Advances in neural information processing systems*, Vol. 26.

Sohn, K. (2016), "Improved deep metric learning with multi-class n-pair loss objective", *Advances in neural information processing systems*, Vol. 29.

Stirgwolt, B.W., Mazzuchi, T.A. and Sarkani, S. (2022), "A model-based systems engineering approach for developing modular system architectures", *Journal of Engineering Design*, Vol. 33 No. 2, pp. 95–119, http://doi.org/10.1080/09544828.2021.1980203.

Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H. and Hospedales, T.M. (2018), "Learning to compare: Relation network for few-shot learning", in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1199–1208.

Tao, F., Zhang, L., Liu, Y., Cheng, Y., Wang, L. and Xu, X. (2015), "Manufacturing service management in cloud manufacturing: overview and future research directions", *Journal of Manufacturing Science and Engineering*, Vol. 137 No. 4, http://doi.org/10.1115/1.4030510.

Van Gemert, J. (2003), *Retrieving images as text*, Ph.D. thesis.

Violante, M.G., Vezzetti, E. and Alemanni, M. (2017), "An integrated approach to support the requirement management (rm) tool customization for a collaborative scenario", *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Vol. 11 No. 2, pp. 191–204, http://doi.org/10.1007/s12008-015-0266-3.

Wang, L., Liu, Z., Liu, A. and Tao, F. (2021), "Artificial intelligence in product lifecycle management", *The International Journal of Advanced Manufacturing Technology*, Vol. 114 No. 3, pp. 771–796.