Human Factors in Design, Engineering, and Computing, Vol. 159, 2024, 193–199
https://doi.org/10.54941/ahfe1005582

Ergonomic Problem and Solution Identification by Applying Image Captioning With Embedded Ergonomic Knowledge

Gunwoo Yong[1], Quan Miao[2], Meiyin Liu[2], and SangHyun Lee[1]

[1]Dept. of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[2]Dept. of Civil and Environmental Engineering, Rutgers University, Piscataway, NJ 08854, USA

ABSTRACT

Work-related musculoskeletal disorders (WMSDs) are a primary cause of non-fatal injuries in diverse industries. Traditional manual ergonomic problem and solution identification for reducing WMSDs is time-consuming and limited by expert availability. Image captioning—interpreting images of workers and their workplaces and capturing interactions therein—is one potential alternative. Yet, due to the absence of ergonomic knowledge, conventional image captioning models are limited in generating accurate captions of ergonomic problems and solutions. Therefore, we aim to automatically identify ergonomic problems and solutions from images by applying image captioning embedded with an ergonomic knowledge graph. Specifically, we developed an ergonomic knowledge graph encoder and incorporated it with the state-of-the-art image captioning model. Comparative testing on eight ergonomic problem-solution pairs showed that our model outperformed the state-of-the-art model. This result highlights the critical role of integrating ergonomic knowledge into image captioning models, paving the way for broader workplace applications to reduce WMSDs.

Keywords: Ergonomic problem and solution identification, Image captioning, Knowledge graph, Computer vision

INTRODUCTION AND RELATED WORKS

WMSDs are the leading cause of non-fatal injuries in the U.S. across diverse industries, accounting for 29.6% of injury cases that required days away from work, job restriction, or transfer in 2021–2022 (BLS, 2023). To protect workers from WMSDs, the Occupational Safety and Health Administration (OSHA) recommends that ergonomic experts visit workplaces to identify ergonomic problems, assess ergonomic risks, and provide corresponding solutions (OSHA, 2015). To mitigate the challenge of limited experts by reducing manual effort, significant advancements have been made in machine learning (ML)-based ergonomic risk assessments using workplace images or wearable sensor signals. However, problem and solution identification still

relies on manual efforts and is infrequently performed. As an alternative, image captioning—a computer vision technique that generates captions from images—offers a promising solution, as demonstrated in our earlier work (Yong et al., 2024). However, existing image captioning models often generate incorrect captions of ergonomic problems and solutions, mainly due to a lack of ergonomic knowledge.

In other domains, some studies have explored methods for injecting external knowledge into image captioning models through knowledge graphs (Wajid et al., 2024), for example, in medicine (Zhang et al., 2020) and transportation (Zhang et al., 2023). However, since existing models are tailored to certain knowledge graphs depending on the nature of the corresponding disciplines, there is a lack of such models for image captioning with an embedded ergonomic knowledge graph.

In overcoming this challenge, we aim to automatically identify knowledge-backed ergonomic problems and solutions from workplace images by applying image captioning embedded with an ergonomic knowledge graph.

DEVELOPMENT OF ERGONOMIC KNOWLEDGE EMBEDDED IMAGE CAPTIONING MODEL

For the development of image captioning embedded with ergonomic knowledge, we first crafted an ergonomic knowledge graph based on essential elements for ergonomic problem and solution identification according to ergonomic guidelines from the OSHA and the National Institute for Occupational Safety and Health (NIOSH) (Albers and Estill, 2007; OSHA, 2015; Torma-Krajewski et al., 2009). Given that these guidelines commonly introduce that task information, ergonomic risk factors, and potential symptoms are essential elements in identifying ergonomic problems and solutions, we designed tasks, risk factors, and symptoms as nodes in our ergonomic knowledge graph. For the edges in our knowledge graph, we utilized the semantic relationships between nodes described in the guidelines. Figure 1 depicts an example of our ergonomic knowledge graph based on the ergonomic guideline for construction tasks (Albers and Estill, 2007).

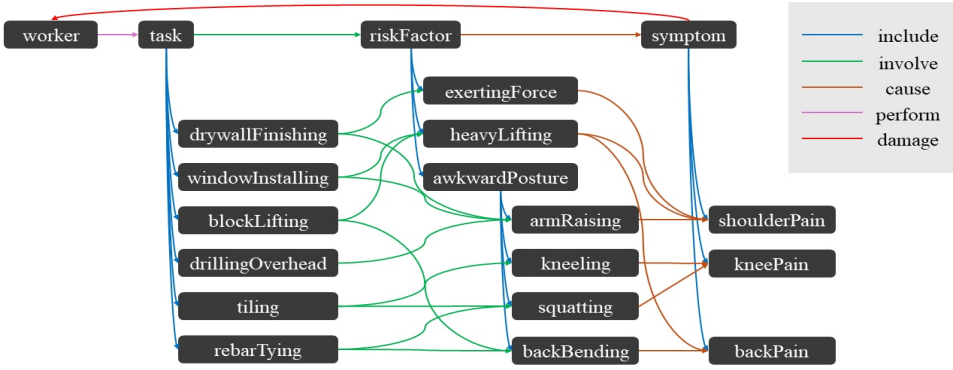

Figure 1: Example of an ergonomic knowledge graph.

Subsequently, unlike general models, we developed an additional encoder to make an image captioning model extract features of ergonomic knowledge from the knowledge graph and inserted it between an image encoder and a text decoder within an image captioning model. Figure 2 displays the detailed architecture of our ergonomic knowledge graph-embedded image captioning model.

Our model is based on Bootstrapping Language-Image Pre-training (BLIP) (Li et al., 2022), the state-of-the-art vision-language model that can perform multiple tasks, including but not limited to image captioning. When inputting an image, our model initially identifies the most suitable task node within the ergonomic knowledge graph through image-text similarity computation. Subsequently, our model retrieves related information from the knowledge graph in triplet form, the most commonly used format to represent knowledge graphs (Kejriwal, 2019). This related information is compressed into triplet embeddings via a language model and 1D convolutional neural networks (CNNs). Finally, the text decoder generates captions from the image embedding and triplet embedding.

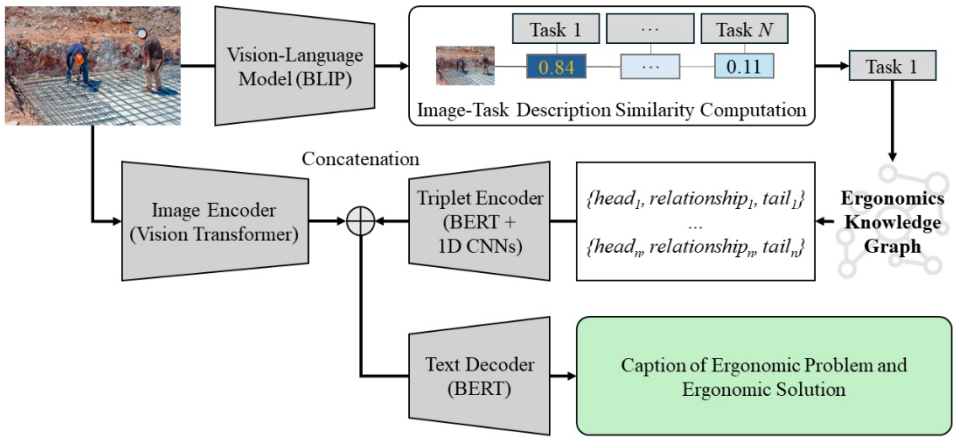

Figure 2: Detailed architecture of our ergonomic knowledge graph embedded image captioning model (peuceta/stock.adobe.com).

TESTING

Dataset Preparation

To demonstrate that leveraging ergonomic knowledge enhances an image captioning model's performance in identifying ergonomic problems and solutions, we tested eight ergonomic problem-solution pairs from the NIOSH ergonomic guideline (Albers and Estill, 2007). This NIOSH guideline introduces ergonomic solutions for specific ergonomic problems occurring in physically demanding construction tasks. Table 1 lists these eight ergonomic problem-solution pairs. The first sentence describes the ergonomic problem, while the second sentence describes the corresponding ergonomic solution.

Table 1. Selected eight ergonomic problem-solution pairs.

Ergonomic Problems and Their Corresponding Solutions.
1. Bending forwards to tie rebars at ground level can cause lower back pain. Use a rebar- tying tool with extension handle.
2. Squatting for rebar tying at ground level can be painful to the knees. Use a rebar-tying tool with extension handle.
3. Tiling by kneeling can harm a worker's knees. Use a portable kneeling creeper with chest support.
4. Tiling requires a worker to squat, causing knee pain. Use a portable kneeling creeper with chest support.
5. Raising arms to drill overhead can lead to shoulder injuries. Use a bit extension shaft for the drill.
6. Raising arms while exerting high forces for drywall finishing can cause shoulder injuries. Use a pneumatic drywall finishing system.
7. Lifting a heavy concrete block requires a worker to bend his back, potentially causing back injuries. Use lightweight concrete block.
8. Installing large windows requires a worker to raise arms, leading to shoulder injuries. Use powered vacuum lifters

We then built an image-caption dataset related to these eight ergonomic problems-solution pairs. We collected 3,600 images from YouTube and 322 images from various other websites, each depicting one of the selected eight ergonomic problems. To demonstrate that our model is not limited to specific workplaces but is applicable across various workplaces, we used the YouTube images for training and the images from other websites for testing. In a manner similar to the caption preparation process used in our prior study (Yong et al., 2024), we prepared a total of 155 synonymic captions for each pair. The statistics of our training and testing datasets are described in Table 2. Note that, for each pair, all 155 captions were used as ground-truth captions since they are synonymic.

Table 2. Statistics of our training and testing datasets.

Ergonomic Problem-Solution Pairs	# of Training Image-Caption Pairs	# of Testing Images	# of Ground-Truth Captions
Pair 1	447	32	155
Pair 2	556	48	155
Pair 3	499	44	155
Pair 4	474	47	155
Pair 5	368	34	155
Pair 6	533	48	155
Pair 7	369	36	155
Pair 8	354	33	155
Sum	3,600	322	1,240

Evaluation Metric

To measure the similarity between ground-truth captions and generated captions, we used the BiLingual Evaluation Understudy (BLEU) metric (Papineni et al., 2002). The BLEU-*N* score, ranging from BLEU-1 to BLEU-4, calculates how many consecutive *N*-word sequences from the generated captions appear in any of the ground-truth captions, on a scale from 0 to 1. As our ground-truth captions are based on the ergonomic guideline, a high BLEU score indicates that our model accurately identifies ergonomic problems and solutions. We compared the BLEU score of our model to that of our backbone model, BLIP, which lacks ergonomic knowledge. This comparison shows whether leveraging an ergonomic knowledge graph is effective in identifying knowledge-backed ergonomic problems and solutions for image captioning models.

RESULTS AND DISCUSSION

Table 3 summarizes the BLEU scores of our model and those of the baseline (i.e., BLIP). For the BLEU-4 score, which compares sequences from a single word up to four consecutive words, ours achieved 0.5810, while the baseline recorded 0.4061. Our ergonomic knowledge embedded model outperformed the baseline across all BLEU scores, from the BLEU-1 score to the BLEU-4 score.

Table 3. BLEU score comparison between our model and the baseline.

Models	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Baseline (BLIP)	0.6892	0.5890	0.5054	0.4061
Ours	0.7833	0.7011	0.6407	0.5810

Examples of captions generated by our model and the baseline is shown in Figure 3. Note that any errors in the generated captions are marked with a stroke. Our model leveraged ergonomic knowledge to more accurately identify tasks, ergonomic risk factors, and potential symptoms, thus generating more accurate captions. These results demonstrate that leveraging ergonomic knowledge is crucial in yielding better captions of ergonomic problems and solutions.

Nonetheless, given that our model requires task identification for accessing the knowledge graph, inaccuracies in task identification from image-text similarity computations may misdirect ergonomic knowledge (Figure 3 bottom). Future research needs to focus on enhancing task identification accuracy to solve this issue.

CONCLUSION

To automatically identify knowledge-backed ergonomic problems and solutions, we applied image captioning embedded with ergonomic knowledge. We constructed an ergonomic knowledge graph and integrated it with the state-of-the-art image captioning model. Upon testing our

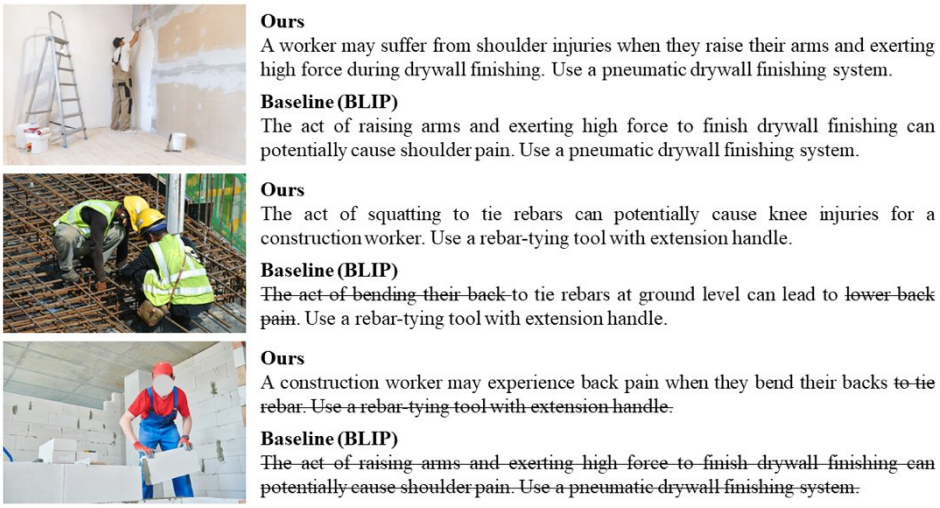

Figure 3: Examples of correctly and incorrectly generated captions by our model and the baseline (perfectlab/Aisyaqilumar/Kadmy/stock.adobe.com).

model with 322 actual workplace images, we identified that our model (BLEU-4 of 0.5810), which leverages ergonomic knowledge, outperforms the state-of-the-art model without such knowledge (BLEU-4 of 0.4061). These results demonstrate the effectiveness of incorporating ergonomic knowledge in image captioning models for ergonomic problem and solution identification. Our accessible automated model is designed to assist in reducing potential WMSDs by intervening in hazardous workplaces where ergonomic knowledge or ergonomic experts are limited.

ACKNOWLEDGMENT

This research was supported by VelocityEHS and the National Safety Council (NSC). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of VelocityEHS and NSC.

REFERENCES

Albers, J. T., Estill, C. F., 2007. Simple Solutions: Ergonomics for Construction Workers (No. 2007–122). NIOSH.

BLS, 2023. Employer-Reported Workplace Injuries and Illnesses - 2021–2022 (No. USDL-23-2359). U. S. Bureau of Labor Statistics.

Kejriwal, M., 2019. Domain-Specific Knowledge Graph Construction, SpringerBriefs in Computer Science. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-12375-8

Li, J., Li, D., Xiong, C., Hoi, S., 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation.

OSHA, 2015. Ergonomics - Overview [WWW Document]. Occupational Safety and Health Administration. URL https://www.osha.gov/ergonomics

Papineni, K., Roukos, S., Ward, T., Zhu, W.-J., 2002. Bleu: a Method for Automatic Evaluation of Machine Translation, in: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL). ACL, Philadelphia, Pennsylvania, USA, pp. 311–318. https://doi.org/10.3115/1073083.1073135

Torma-Krajewski, J., Steiner, L. J., Burgess-Limerick, R., 2009. Ergonomics processes: implementation guide and tools for the mining industry. (No. 2009–107). U. S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH. https://doi.org/10.26616/NIOSHPUB2009107

Wajid, M. S., Terashima-Marin, H., Najafirad, P., Wajid, M. A., 2024. Deep learning and knowledge graph for image/video captioning: A review of datasets, evaluation metrics, and methods. Engineering Reports 6, e12785. https://doi.org/10.1002/eng2.12785

Yong, G., Liu, M., Lee, S., 2024. Explainable Image Captioning to Identify Ergonomic Problems and Solutions for Construction Workers. J. Comput. Civ. Eng. 38, 04024022. https://doi.org/10.1061/JCCEE5. CPENG-5744

Zhang, D., Ma, Y., Liu, Q., Wang, H., Ren, A., Liang, J., 2023. Traffic Scene Captioning with Multi-Stage Feature Enhancement. Computers, Materials & Continua 76, 2901–2920. https://doi.org/10.32604/cmc.2023.038264

Zhang, Y., Wang, X., Xu, Z., Yu, Q., Yuille, A., Xu, D., 2020. When Radiology Report Generation Meets Knowledge Graph. AAAI 34, 12910–12917. https://doi.org/10.1609/aaai.v34i07.6989