

World's most accurate visual question-answering AI

September 15 2021

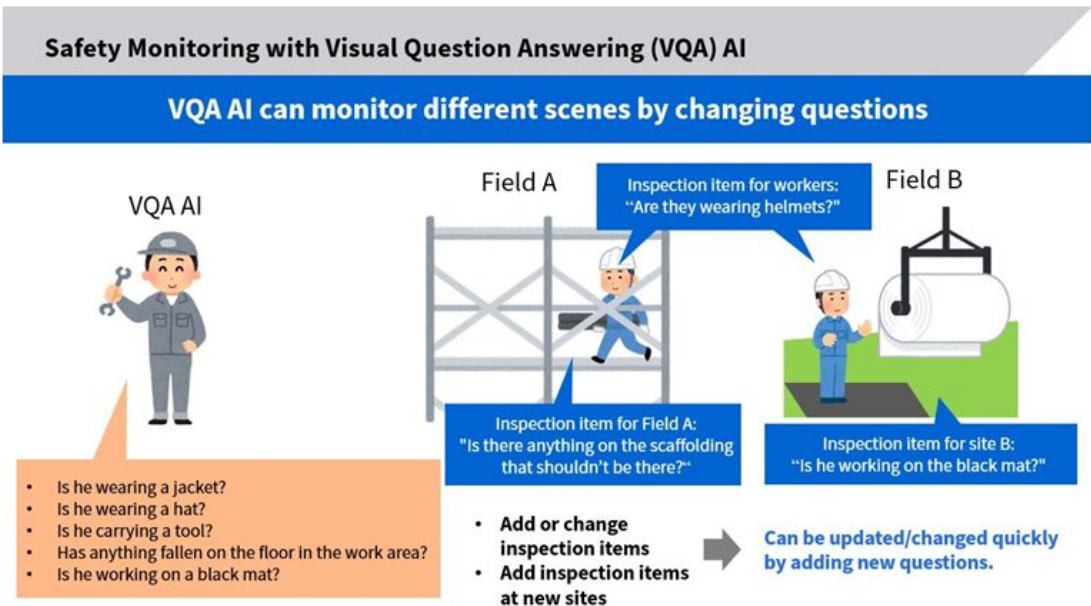


Figure 1: Safety Monitoring with Question-Answering AI. Credit: Toshiba Corporation

Toshiba Corporation has developed the world's most accurate highly versatile Visual Question Answering (VQA) AI, able to recognize not only people and objects, but also colors, shapes, appearances and background details in images. The AI overcomes the long-standing

difficulty of answering questions on the positioning and appearance of people and objects, and has the ability to learn information required to handle a wide range of questions and answers. It can be applied to a wide range of purposes without any need for customization.

In experiments using a public dataset comprising a large volume of images and data text, the VQA AI correctly answered 66.25% of questions without any pre-learning and 74.57% with pre-learning. For example, the AI can find a worker standing in a designated place by asking questions like, "is the person on a black mat?" which requires recognition of the individual, position, shape and color. Applying it to safety monitoring systems at production sites is expected to help improve safety and to reduce workloads on onsite supervisors. It can also be used to identify specific scenes in broadcast content and surveillance video footage.

Toshiba presented the technology at ICANN2021, the international conference for neural networks, on September 14.

Coming years are expected to see growing manpower shortages at production sites in Japan, a trend also become apparent in other advanced nations. This situation is being made all the worse by the emergence of COVID-19, which is making it more essential than ever to ensure worker safety and reduce workloads on site management. One solution is AI, which is being increasingly introduced to production sites. The global AI market, including software, hardware, and services, is forecast to grow 16.4% year over year in 2021 to \$327.5 billion and is expected to reach \$554.3 billion by 2024.

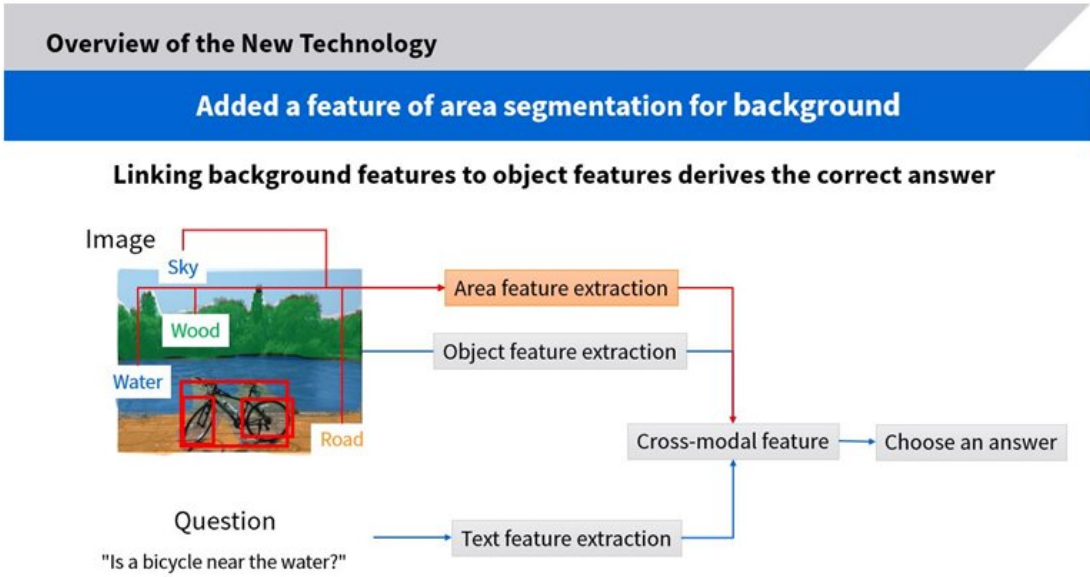


Figure 2: Features of the developed AI. Credit: Toshiba Corporation

Current image recognition AI supports safety inspections at the level where it can detect individual objects learned beforehand, such as people, headwear, and work clothing. This allows it to analyze camera images to determine whether or not someone is wearing a hardhat, or to detect dropped or fallen objects, helping to ensure [workplace safety](#) and reduce the site management workload.

However, getting to this point requires the creation of a determination function that provides a basis for how the AI should recognize an inspection item. For example, when checking for headgear, it must learn how to detect and determine if an individual is wearing a hat—and this has to be done for every individual item that is detected. In a workplace, it is essential to have flexibility that allows immediate changes in

inspection items, but this is difficult with current AI due to time needed to set up and adjust the determination function.

Toshiba's new AI meets the need for flexibility with the world's highest accuracy in answering questions, and it is also able to change or add questions quickly. Its ability to recognize not only people and objects but also image backgrounds, plus the extensive database at its disposal, ensure that it can process quickly the features of images and pre-learned questions to derive the correct answer. After learning a large set of images, questions and answers that cover the presence of people and objects, and information such as their location and status, the AI is able to provide an appropriate answer to a question from approximately 3,000 answer patterns. The AI is highly flexible and can be updated by adding inspection items, or changed to handle a different situation, by a simple "Image and Question" process of adding new question sentences (Fig. 1).

The new AI can answer to questions that requires linking backgrounds to objects – a crucial factor in monitoring safety

Are people standing on the orange floor?*



Correct answer: Yes
Current AI: No ×
New technology: Yes ✓

Is there a person on the green floor??



Correct answer: Yes
Current AI: No ×
New technology: Yes ✓

Figure 3: Example of Question-Answering with AI. Credit: Toshiba Corporation

AI for VQA is a cutting edge-technology now being researched worldwide. The conventional approach mainly relies on the features of people and objects in an image, but Toshiba's new method also extracts background features and spatial areas, including the floors and passageways where these people and objects are to be found (Fig. 2). This feature enables the new AI to derive accurate answers.

For example, the AI can answer questions such as whether there is an [object](#) on a path or if a person is standing in a designated area, as well as whether there is an object (Fig 3 and 4). By applying this AI to safety monitoring at production sites, it is expected to improve workplace safety, to reduce workloads on supervisors, and to contribute to work style improvement.



Figure 4: Example of Question-Answering with AI. Credit: Toshiba Corporation

In a performance evaluation with a global standard public dataset, Toshiba achieved accuracy levels of 66.25% without pre-learning and 74.57% with pre-learning, the highest levels ever recorded, while the results with the current methods were respectively 65.88% and 74.00% (Fig. 5).

Model	Pre-learning	Accuracy (%)
Current model A	None	65.88
Current model A + New technology	None	66.25
Current model B	Yes	74.00
Current model B + New technology	Yes	74.57

Figure 5: Accuracy Comparison with Conventional Methods. Credit: Toshiba Corporation

The versatility of the new AI suits it for application in searches for specific scenes from broadcast content, specific circumstances or people in a disk drive recorders and security footage, and past near-misses in similar situations.

Toshiba will continue system development and accuracy improvement, toward introducing the AI technology into [safety](#) monitoring systems in fiscal 2023.

Provided by Toshiba Corporation

Citation: World's most accurate visual question–answering AI (2021, September 15) retrieved 8 May 2024 from <https://techxplore.com/news/2021-09-world-accurate-visual-questionanswering-ai.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.