

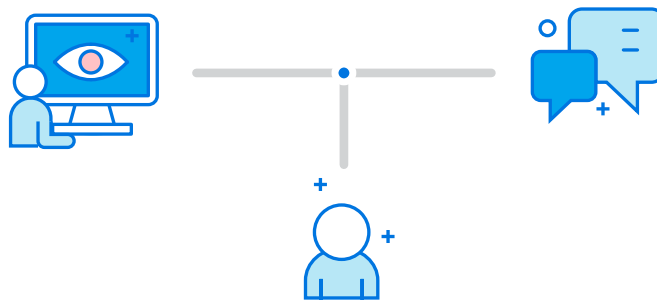
KVQA Dataset

Korean Localization of Visual Question Answering (VQA) for visually impaired People.

From June to December 2019, Testworks and T-Brain, SKT's artificial intelligence research organization, constructed the Korean Visual Question Answering (KVQA) dataset.

What is VQA (Visual Question Answering)?

VQA is an artificial intelligence technology that provides appropriate answers to visual images and image-related questions.



VQA Technology (Source: SK T-Brain Github, KVQA)

Limitations of existing VQA datasets

VQA technology is actively being studied worldwide and requires a learning dataset consisting of images, questions, and answers that fit the linguistic characteristics of each country for development. However, there was no Korean VQA data available in South Korea. Also, most of images in existing dataset were taken by people without visual impairment. There was a limit to developing technologies/services for people with visual impairment because images, taken by people with visual impairment who have difficulty in recognizing the surrounding environment and objects, were different from those taken by people without visual impairment.

The World's First KVQA Dataset

Testworks and T-Brain constructed the world's first Korean-based VQA dataset, KVQA dataset, to improve these VQA technology problems and develop artificial intelligence services for the people with visual impairment.

People with visual impairment participated in the collection of images and questions data during the project period, improving the limitations of the existing VQA dataset.



(a) **Q:** Can I cross the crosswalk now? **A:** No



(b) **Q:** How many lights in this room? **A:** 2



(c) **Q:** What is the person doing in this room? **A:** Piano



(d) **Q:** What kind of flower is this? **A:** Unanswerable

Fig.1 : Examples of KVQA dataset. The most frequent answers are shown for each question. The above examples are image-question pairs of Yes/No, Number, Other, and Unanswerable type in order.

[Download KVQA dataset](#)

KVQA Data (Source: SK T-Brain Github, KVQA)

Image and question data collected directly from people with visual impairment.

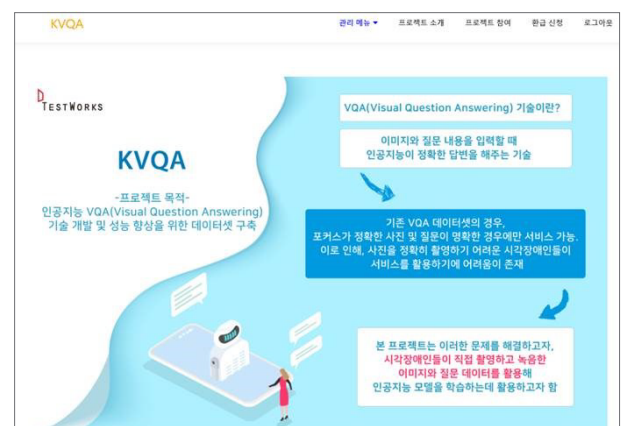
Testworks collaborated with the Korea Blind Association, Siloam Center for the Blind, K.A.B.C.S(Korea Association of Blind College Students), and the Korea Blind Union to collect image and question data.

With the help of the Korea Association for the Blind, we promoted active public relations activities, such as the production of braille newspapers to encourage people with visual impairment to participate in data collection.

Collect large-scale answer data from the crowd-sourcing platform aiworks

Testworks uses aiworks, an online platform for large-scale data collection using crowd-sourcing method, to collect images and question data collected by the visually impaired.

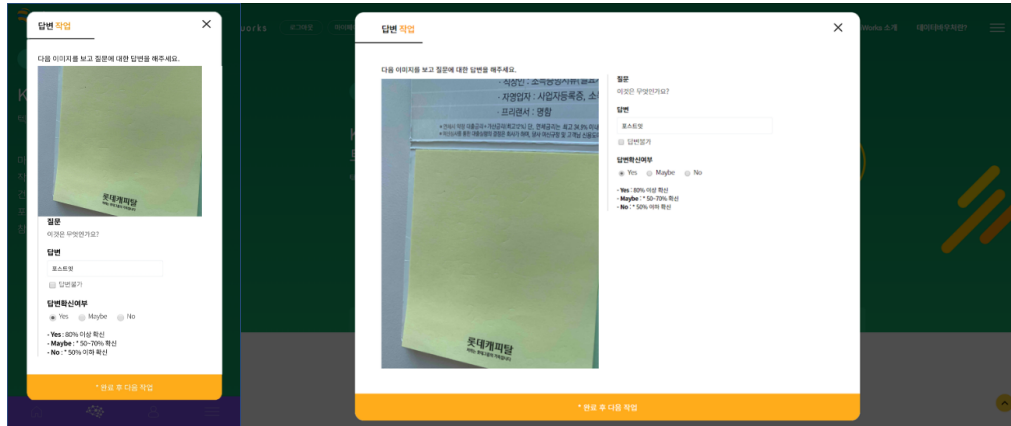
First, we opened a page for the KVQA data collection project and recruited crowd-worker to participate in collecting answer data.



Landing page for KVQA data construction (Source: Testworks Inc.)

In addition, we opened a project that provides cashable points or providing volunteer hours when viewing images and questions on aiworks platform, and writing answers.

Crowd-workers and volunteers who participated in the project connected to aiworks through PCs and smartphones and participated in collecting answer data



Collection of answer data through aiworks, left smartphone screen, right PC screen (Source: Testworks Inc.)

Performance of constructing KVQA datasets

By constructing KVQA data, about 250 people with visual impairment from across the country participated in the image and question data collection for about six months, and around 4,500 people and volunteers provided answer data. As a result, we were able to build data with 100,000 images, 100,000 questions, and 1 million answers.

V1.0 (Jan. 2020)

	Overall (%)	Yes/No (%)	Number (%)	Other (%)	Unanswerable (%)
# images	100,445 (100)	6,124 (6.10)	9,332 (9.29)	69,069 (68.76)	15,920 (15.85)
# questions	100,445 (100)	6,124 (6.10)	9,332 (9.29)	69,069 (68.76)	15,920 (15.85)
# answers	1,004,450 (100)	61,240 (6.10)	93,320 (9.29)	690,690 (68.76)	159,200 (15.85)

KVQA dataset construction statistics (Source: SK T-Brain Github, KVQA)

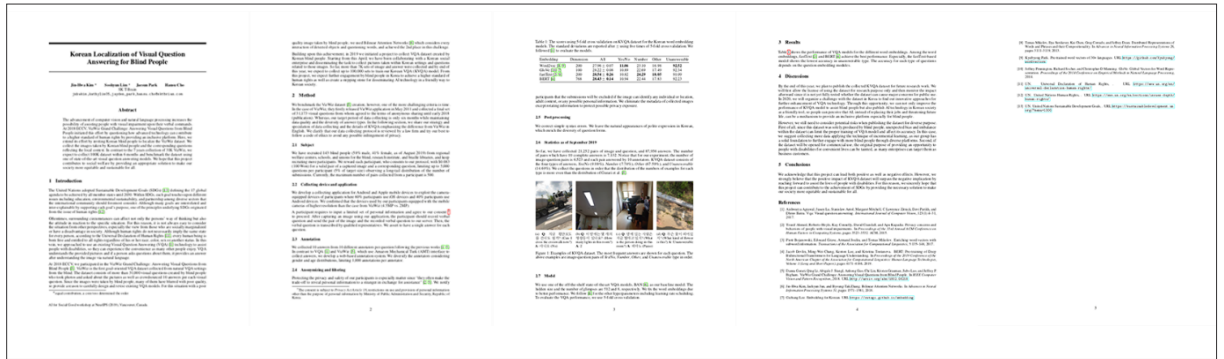
T-Brain, taking second place after Facebook in the VQA Challenge before the KVQA project, has written paper on the constructed dataset and introduced the achievements of KVQA at the workshop AI for Social Good in NeurIPS, the world's best artificial intelligence conference 2020.

During the KVQA project, Testworks successfully established a large-scale Korean-based VQA dataset and created shared value for solving social problems by linking social services programs to aiworks, data collection processing platform.

Testworks continues to participate in various shared value projects to solve social problems with customers based on technology-based growth.

Paper

AI for Social Good workshop at NeurIPS (Kim & Lim et al., 2019)



KVQA dataset paper (Source: SK T-Brain Github, KVQA)

PDF Download

Description of the Data Field

Name	Type	Description
VQA	[dict]	list of dict holding VQA data
+ image	str	filename of image
+ source	str	data source ["kvqa" "vizwiz"]
+ answers	[dict]	list of dict holding 10 answers
+ answer	str	answer in string
+ answer_confidence	str	["yes" "maybe" "no"]
+ question	str	question about the image
+ answerable	int	answerable? [0 1]
+ answer_type	str	answer type ["number" "yes/no" "other" "unanswerable"]

KVQA data field (Source: SK T-Brain Github, KVQA)

Example of data

```
[{
  "image": "KVQA_190712_00143.jpg",
  "source": "kvqa",
  "answers": [{
    "answer": "피아노",
    "answer_confidence": "yes"
  }, {
    "answer": "피아노",
    "answer_confidence": "yes"
  }, {
    "answer": "피아노 치고있다",
    "answer_confidence": "maybe"
  }, {
    "answer": "unanswerable",
    "answer_confidence": "maybe"
  }, {
    "answer": "게임",
    "answer_confidence": "maybe"
  }, {
    "answer": "피아노 앞에서 무언가를 보고 있음",
    "answer_confidence": "maybe"
  }, {
    "answer": "피아노치고있어",
    "answer_confidence": "maybe"
  }, {
    "answer": "피아노치고있어요",
    "answer_confidence": "maybe"
  }, {
    "answer": "피아노 연주",
    "answer_confidence": "maybe"
  }, {
    "answer": "피아노 치기",
    "answer_confidence": "yes"
  }
  ],
  "question": "방에 있는 사람은 지금 뭐하고 있지?",
  "answerable": 1,
  "answer_type": "other"
},
{
  "image": "VizWiz_train_000000008148.jpg",
  "source": "vizwiz",
  "answers": [{
    "answer": "리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "티비 리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "리모컨",
    "answer_confidence": "maybe"
  }, {
    "answer": "리모컨",
    "answer_confidence": "yes"
  }, {
    "answer": "리모컨",
    "answer_confidence": "yes"
  }
  ],
  "question": "이것은 무엇인가요?",
  "answerable": 1,
  "answer_type": "other"
}
]
```

KVQA data example (Source: SK T-Brain Github, KVQA)