Photo-based Question Answering

Tom Yeh\textsuperscript{2}, John J. Lee\textsuperscript{1}, Trevor Darrell\textsuperscript{2}

EECS & CSAIL, MIT\textsuperscript{1}
EECS & ICIS, UC Berkeley\textsuperscript{2}

Is there any problem?\textsuperscript{1}
How many stories?\textsuperscript{2}
Who is the architect?\textsuperscript{2}

Many people found text-based QA useful.

Text-based QA has become increasingly popular.

Many of our questions have been asked already.
Text-based QA sometimes can be difficult.

Photo-based QA sometimes can be more desirable.

The community is receptive of photo-based QA.

Some photo-based QA can be automated.

Prototype 1: Adding photos to a text-based QA system

Prototype 2: Adding QA to a photo-album system

I'm looking for a poster of Phil Daniels on Vempa with "The Who" logo on the back of his coat and the fingers of 'The Who' faces in the bike's mirrors.
Prototype 3: Applying photo-based QA to mobile devices.

Here are some photo-based questions!

- How many floors?
- Is there any problem?
- Who is the architect?
- How many stories?
- What labs are here?
- How tall?
- What is the rating of this book?
- Is this book good?
- How tall?
- Who wrote this book?
- When was it written?
- Is there a sequel?
- Who lives here?

Three-layer architecture for Photo-based QA

Easiest questions are handled by Template-based QA.

- How many floors?
- Is there any problem?
- Who is the architect?

Layer 1

Template-based QA

- Books
- Buildings

Layer 2

IR-based QA

Layer 3

Human-based QA

Resolved Questions

Community

Easiest questions are handled by Template-based QA.

Layer 1

Template-based QA

- Books
- Buildings

Layer 2

IR-based QA

Layer 3

Human-based QA

WWW
Amazon
Wiki...
Easiest questions are handled by Template-based QA.

Frequent questions are handled by IR-based QA.

Hard questions are handled by Human-based QA.
Hard questions are handled by **Human-based QA.**

Images can be indexed based on visual properties.

It is often advantageous to partition the index.

But, choosing the right partition is tricky.

Question can help select the partition.

Question can help filter within the partition.
Evaluation is based on a dataset of 30,000+ images.

Sample match results

Image matching may perform poorly without any filtering.
Category-based filtering can improve performance.

Questions?