

# AN ARTIFICIAL INTELLIGENT CHAT-BOT TO HELP THE BLIND

Daniela Massiceti, Siddharth N., Puneet K. Dokania, Philip H.S. Torr

Department of Engineering Science, University of Oxford

{daniela, sid, puneet, phst}@robots.ox.ac.uk



## Blindness is a challenge

- Visual impairment affects 253 million people globally and 4 million in the UK
- Blindness costs the NHS £4.9 billion per year
- For the blind, independent living is a big challenge & assistive services are few & costly

## Visual chat-bots are a solution

- Visual chat-bots are artificially intelligent (AI) agents that can have conversations about visual scenes – like *Siri* with “eyes”
- The blind user asks the visual chat-bot a question
- Using **machine learning**, the visual chat-bot can helpfully respond

Is there an obstacle in front of me?

Yes, to your left

## VISUAL CHAT-BOT

### 1 Vision Neural Network

An image is a matrix of pixels. Each pixel is represented by three numbers:



The Vision NN transforms the matrix into a number vector

The vector represents the image:  
\* objects  
\* layout  
\* colours

question

Is she smiling?

image



dialogue history

How old does the girl appear to be?	Four
What color hair does the girl have?	Blonde
Is this inside a coffee shop?	No
Are there any other people?	No

Language NN

Vision NN

Fusion NN

Language NN

Looking into our **neural network** visual chat-bot model

No, she's eating

answer

### 2

### Language Neural Network

We define a vocabulary of words (1. obstacle, 2. person, 3. avoid...)

Using the vocabulary indices, each question/answer can be represented as a number vector

The Language NN transforms each vector into another vector which captures:

- \* word sequence
- \* grammar
- \* sentence meaning

### 3 Fusion Neural Network

The Fusion NN fuses the question, image and history into a single vector

This vector corresponds to a sequence of vocabulary indices – the answer words!

We reward the chat-bot if the answer is grammatical & sensible

We penalise the chat-bot otherwise, thus ‘teaching’ it

### Future

Collect a dataset that captures real-world scenarios faced by the blind

Scale-up the visual chat-bot to handle real-time video

STEM for Britain  
March 2018

[www.robots.ox.ac.uk/~daniela](http://www.robots.ox.ac.uk/~daniela)