Unreasonable Effectiveness of OCR in Visual Advertisement Understanding

Mayu Otani, Yuki Iwazaki, Kota Yamaguchi
CyberAgent, Inc.
Challenges in visual ads

Illustrations

Symbolism

Complecated Compositions
Key observations

Texts on ads seem the most powerful clues
Our approach

1. Extract image2text and ocr2text relevance
2. Rank the statement given both scores

Statement text: I should shop for jeans at this store because they use their profits to build homes for people.

Visual and OCR clues from ad: old jeans new homes

Image2text relevance

Ocr2text relevance
Our model

"I should shop for jeans ..."

image2text network

Faster RCNN

old jeans new homes

ocr2text network

Text CNN

1-layer net

weighted average

L2 distance (negative relevance)

1-layer net

normalize

weighted average

normalize
Per-word weights

Vector similarity between OCR detected words and words in a statement text
Long-term dependency harms. We trim texts.

I should shop for jeans at this store because they use their profits to build homes for people.

I should <action> because <reason>

shop for jeans at this store

they use their profits to build homes for people

Action to image / OCR relevance estimation

Reason to image / OCR relevance estimation
Training

Contrastive Loss

\[ L_n = \frac{1}{2} \left( y_n d_n^2 + (1 - y_n) \max(\mu - d_n, 0)^2 \right) \]

\( y_n \): Label. 1 = correct text, 0 = incorrect text

\( d_n \): Distance between text and image/OCR embeddings

“Old, jeans, new, homes, …”

OCR words embedding

“New, jeans, old, homes, …”

Image embedding
Examples of QA results

A. I should shop for jeans at this store Because they use their profits to build homes for people

Per-word weights for action

old jeans new homes we partnered exclusively with brad make it right foundation non profit organization that builds sustainable homes for people in need

Per-word weights for reason

old jeans new homes we partnered exclusively with brad make it right foundation non profit organization that builds sustainable homes for people in need
Examples of QA results

Per-word weights for reason

toyota leadership award the spirit of the leader part courage part determination single minded dedication to personal excellence and the ability to inspire excel in others what makes leader as leader in automotive design and technology toyota recognizes and this spirit by presenting the toyota leadership award to an each team competing in college football games …

A. I should buy Toyota Because they support college leaders
Examples of QA results

No words detected

A. I should shop at the gap Because it will make me sexy
How much does OCR help?

I should shop for jeans at this store Because they use their profits to build homes for people

old jeans
new homes

P@1
0.42
0.82
How much does appearance help?

I should shop for jeans at this store because they use their profits to build homes for people.

old jeans → new homes

new homes → old jeans

P@1 (validation set)

0.83

0.85
Is OCR quality crucial?

A. Yes

<table>
<thead>
<tr>
<th></th>
<th>Score on validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Cloud Vision OCR</td>
<td>0.85</td>
</tr>
<tr>
<td>Tesseract OCR v3</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Tesseract v3

Water, Hydration and Health
A Toolkit for Registered Dietitians
Nestle Waters
The Healthy Hydration Company

Google Cloud Vision

Water, Hydration and Health
A Toolkit for Registered Dietitians
Nestle Waters
The Healthy Hydration Company
Things we didn’t try
but look important

OCR results correction

BLAC K B E R R Y R E M E M B E R ...

Multilingual support

Words detected but discarded

'誕生。美容オイル生答転のルージュ。
それは、美容オイルがスティックになった、愛らしく色艶。
とろけて密着。色っぽくうるんだ官能の唇へ。
マキアージュドラマティックルージュ全10色新発売
MA K E u P DATA : p クアチイックルー問" RD425 / トカルーアイ
Pipp-V1233: 殿定カラート
2 Mn 合わ □ 0 1 2 0 . 3 0 - 4 7 1 0 0 9 0 0 -
2100席楽ギ始收定AMHEMO\nwww.shiseido.cojpmq\nレディにし

NEWMAQUIIAGE\nとの唇、女っぽくて、ごめんなさい。

'
Limitations

- Multilingual support
- Decorative font
- Paintings, drawings
Misc findings

- Higher resolution helps
  - Perhaps because CNNs can literally read texts
  - Also object detection?

- Annotators really read texts
  - And perhaps logos
Symbolism might be used for rhetorics, but not necessarily for surface messages.
Remaining questions

• How well do humans understand ad messages without language clues?

• Effects of designs (colors, layout, font style, etc.) in message telling

• Better task / dataset design?
  • Hiding / blurring texts enough?