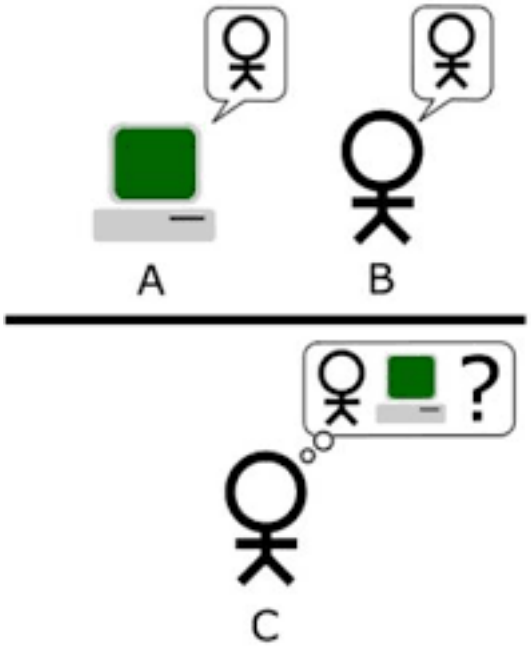
VQA
视觉问答机器人

七月在线 加号

微博：@翻滚吧_加号

目录

VQA的起源



source: https://zh.wikipedia.org/wiki/%E5%9B%BE%E7%81%B5%E6%B5%8B%E8%AF%95

VQA的起源

Computer AI passes Turing test in 'world first'

9 June 2014 | Technology

source: http://www.bbc.co.uk/news/technology-27762088

33%

VQA的起源

新图灵测试的探讨

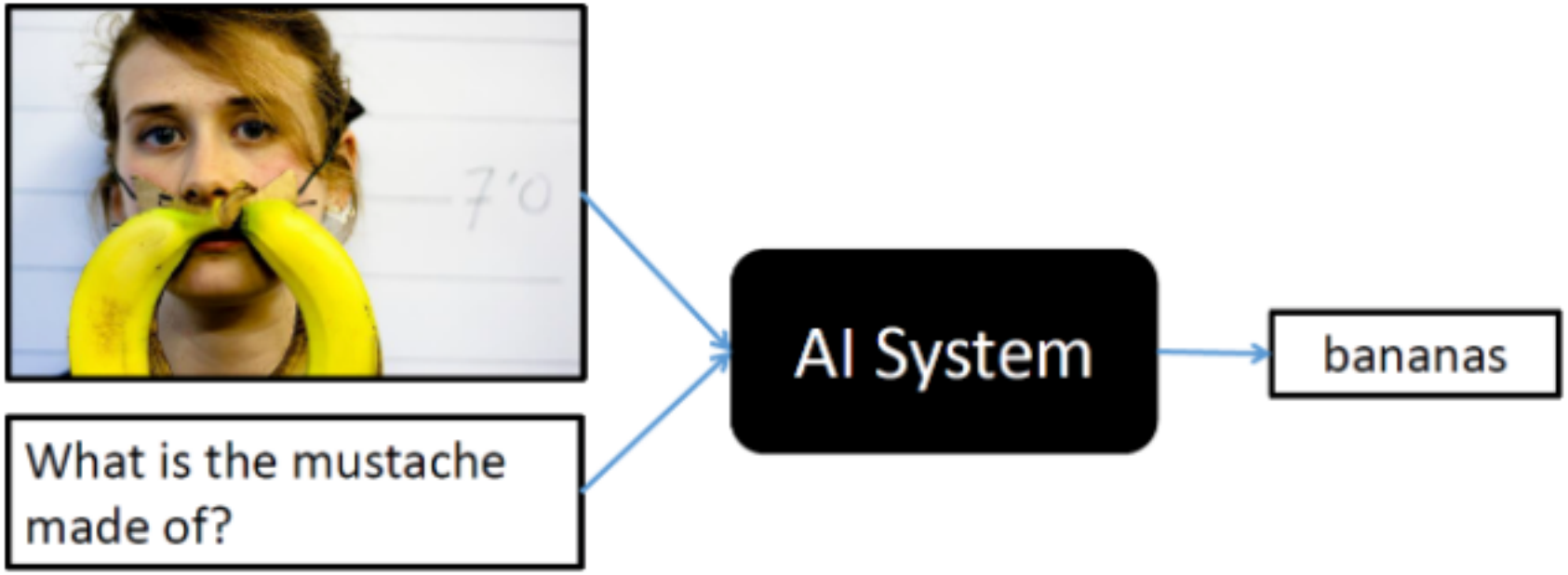
Malinowski, M. and Fritz, M., 2014. Towards a visual turing challenge. *arXiv preprint arXiv:1410.8027*. https://arxiv.org/abs/1410.8027

Geman, D., Geman, S., Hallonquist, N. and Younes, L., 2015. Visual turing test for computer vision systems. *Proceedings of the National Academy of Sciences*, *112*(12), pp.3618-3623. http://www.pnas.org/content/112/12/3618.abstract

VQA的起源

VQA Visual Question Answering

http://www.visualqa.org/

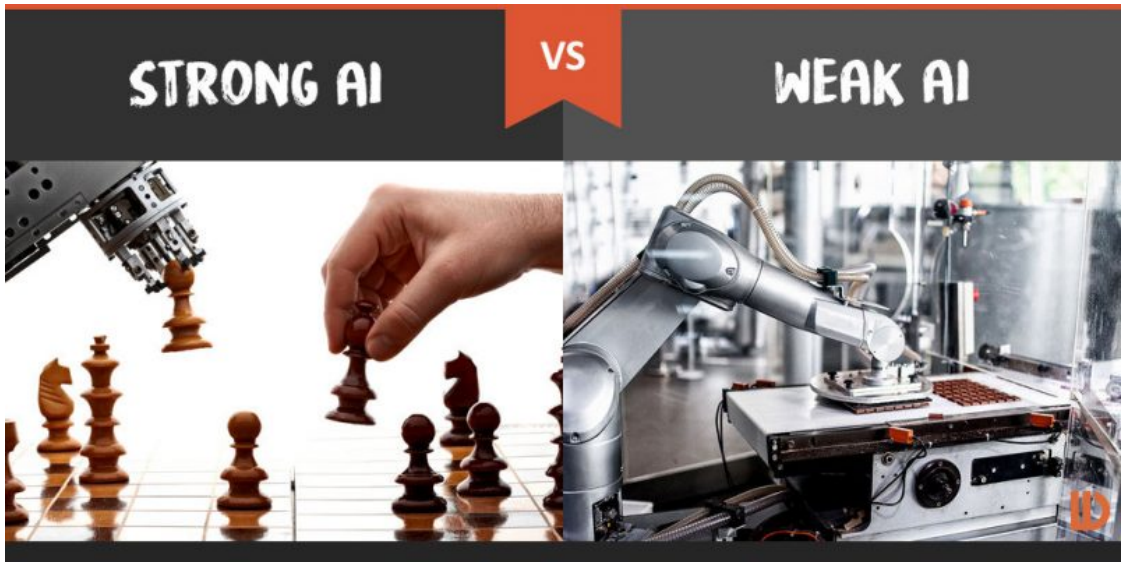


VQA的起源

主要涉及：
NLP
Computer Vision
“common-sense” reasoning

VQA模型

第一步 生成答案



VQA模型

第一步 生成答案

Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C. and Parikh, D., 2015. Vqa: Visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2425-2433). https://arxiv.org/pdf/1505.00468v6.pdf

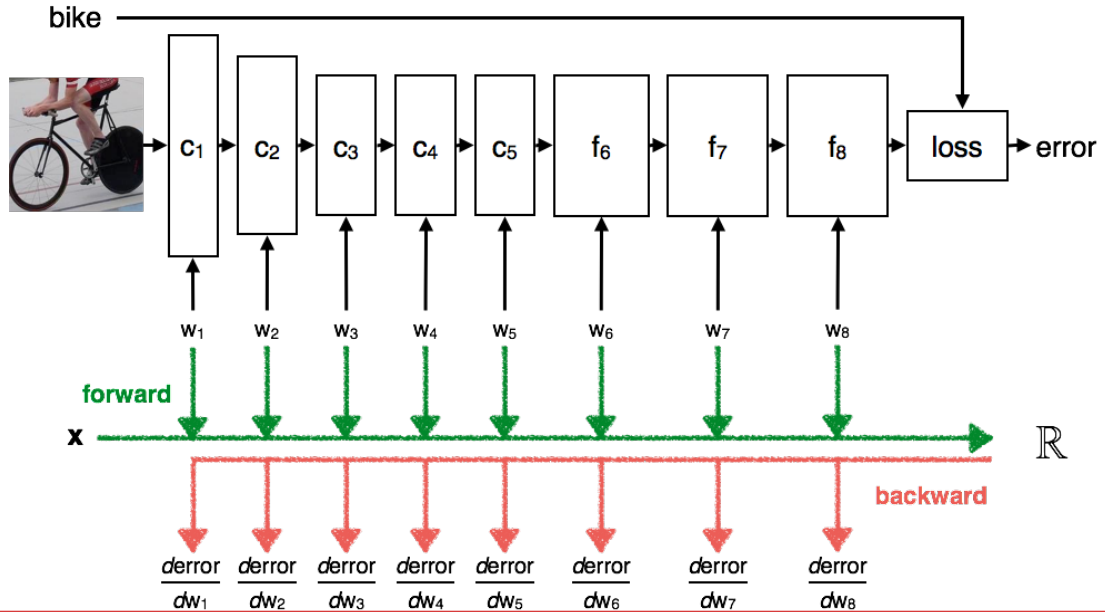
VQA模型

第二步 处理输入源数据：图片

CNN：

VGG-16(Visual Geometry Group, Uni of Oxford)

http://www.robots.ox.ac.uk/~vgg/



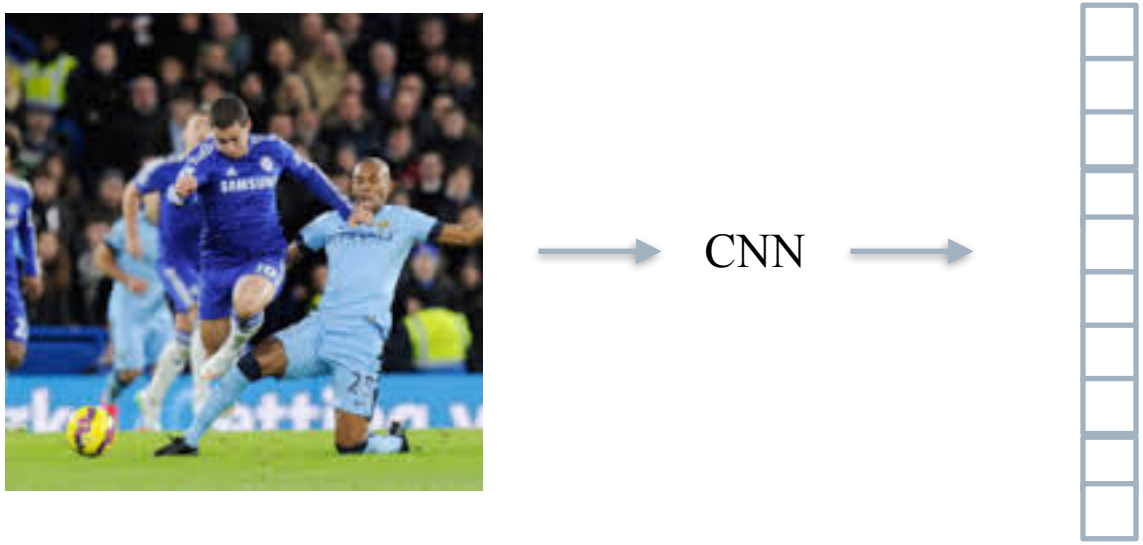
VQA模型

VGG-16的标准构造 (keras)

```
def VGG_16(weights_path=None):
    model = Sequential()
    model.add(ZeroPadding2D((1,1),input_shape=(3,224,224)))
    model.add(Convolution2D(64, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(64, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))

    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(128, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(128, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))

    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(256, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(256, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(256, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))

    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))

    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))

    model.add(Flatten())
    model.add(Dense(4096, activation='relu'))
    model.add(Dropout(0.5))
    model.add(Dense(4096, activation='relu'))
    model.add(Dropout(0.5))
    model.add(Dense(1000, activation='softmax'))

    if weights_path:
        model.load_weights(weights_path)

    return model
```

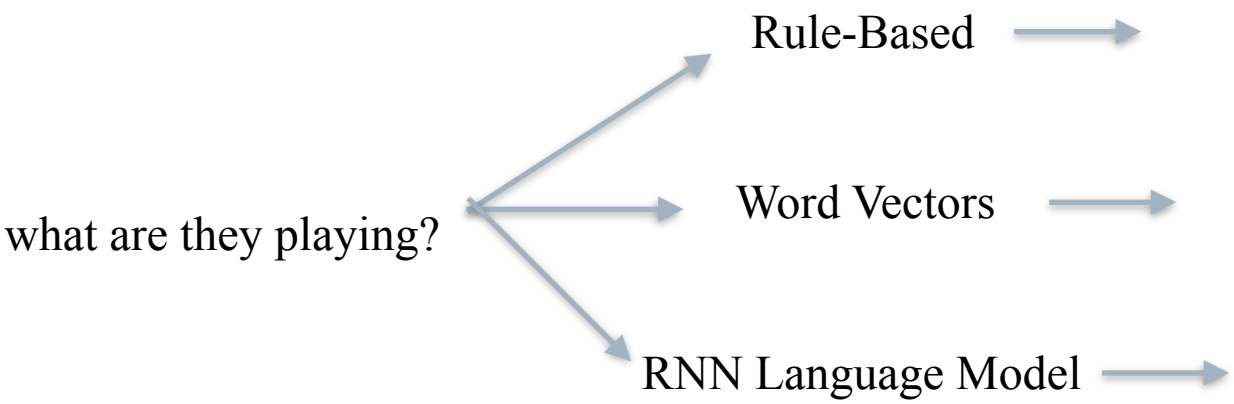
VQA模型

第二步 处理输入源数据：图片



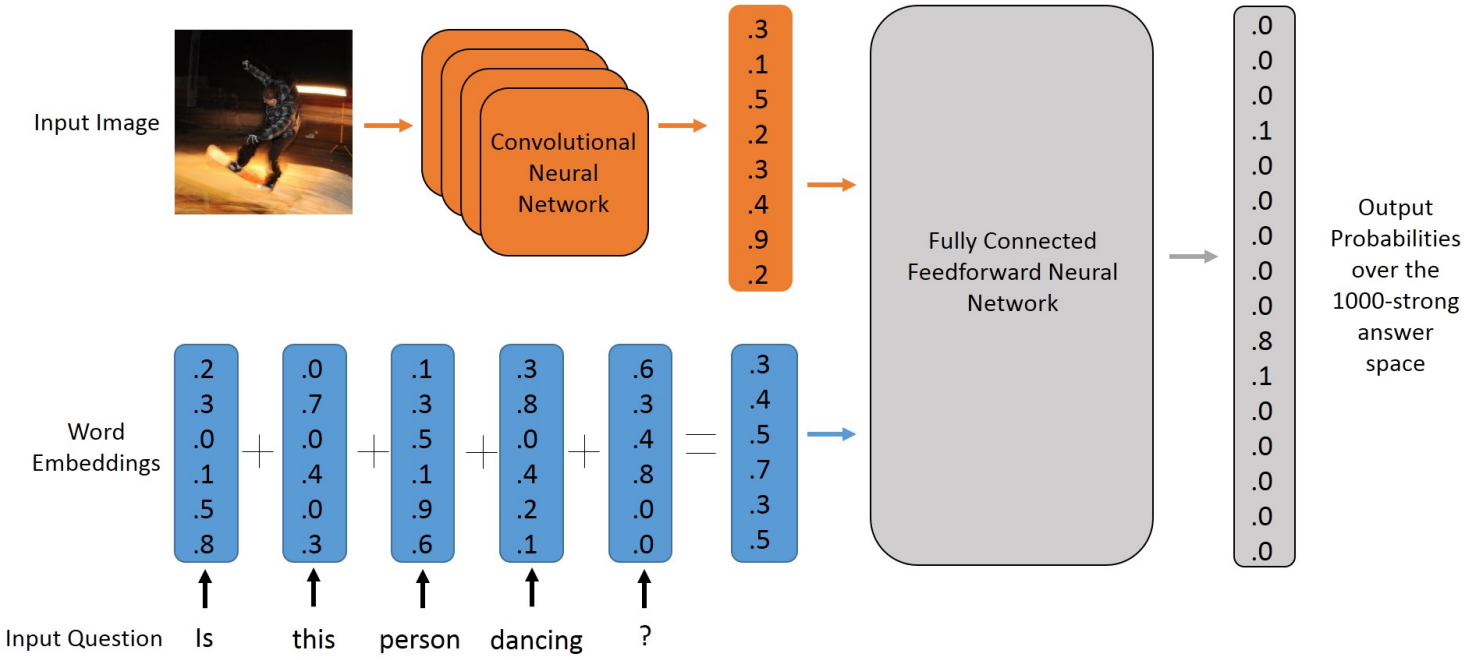
VQA模型

第二步 处理输入源数据：文字



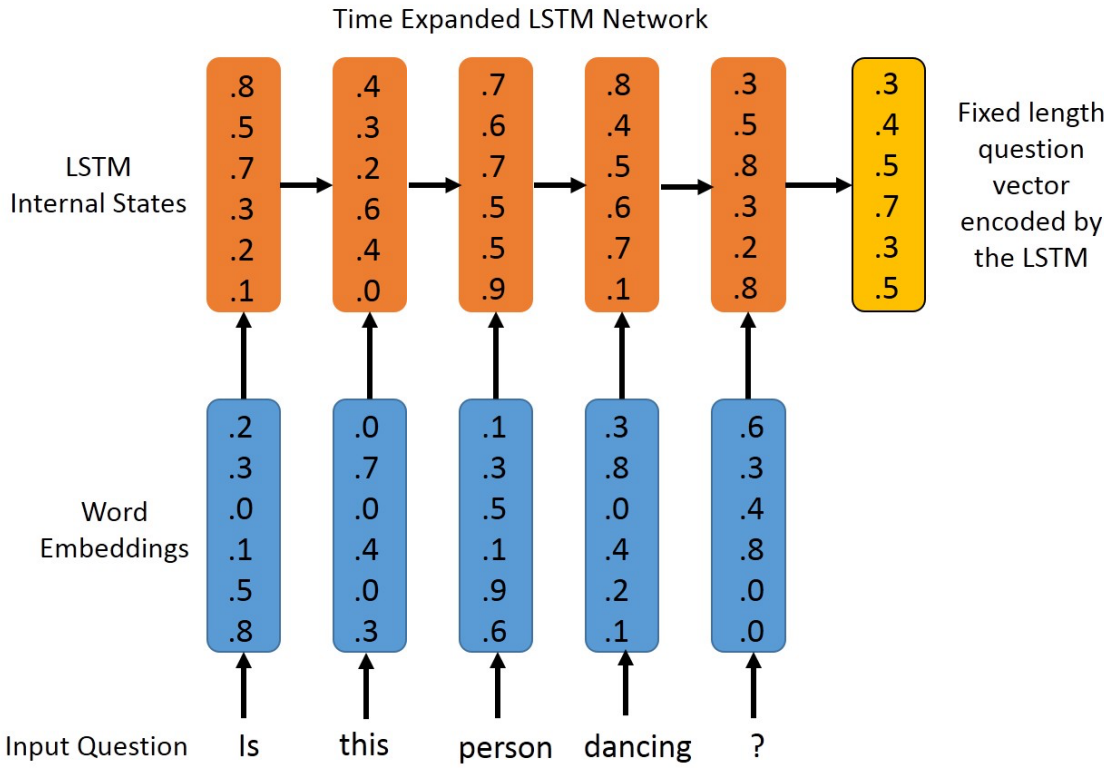
VQA模型

第三部 选取VQA模型-MLP



VQA模型

第三部 选取VQA模型-LSTM



VQA模型

VQA实战

【代码见相应iPythonNotebook】

VQA一些开源代码参考

https://github.com/VT-vision-lab

https://github.com/abhshkdz/neural-vqa

https://github.com/iamaaditya/VQA_Demo

VQA进阶发展

带上注意力的模型

Lu, J., Yang, J., Batra, D. and Parikh, D., 2016. Hierarchical question-image co-attention for visual question answering. In *Advances In Neural Information Processing Systems* (pp. 289-297). https://arxiv.org/abs/1606.00061

『真正』的VQA（参考）

Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104-3112). https://arxiv.org/abs/1409.3215

感谢大家！

恳请大家批评指正！