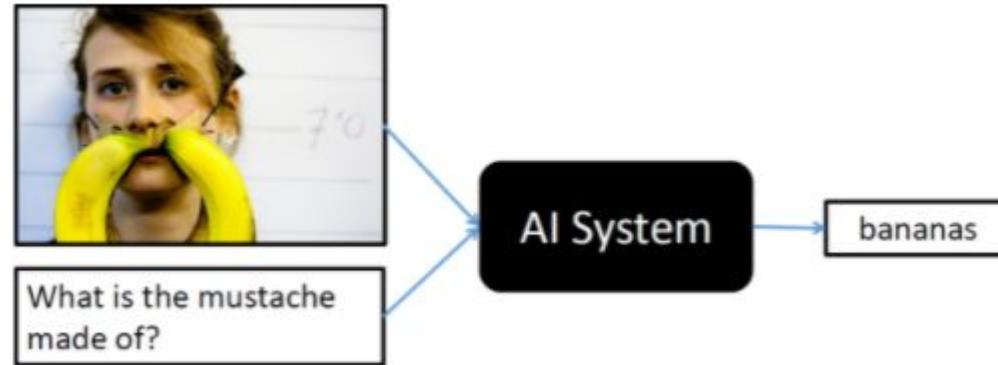


# Medical Domain VQA

Project 5  
Ruiling Zhang  
Zihao Shen  
Yuko Ishikawa

# What is VQA?

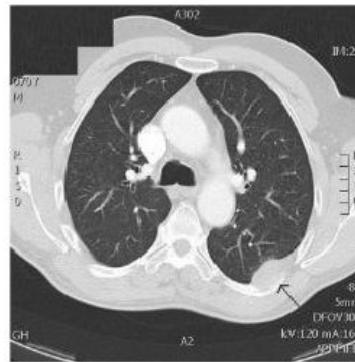
Visual Question Answering (VQA) is a recent problem in computer vision and natural language processing. An algorithm takes as input an image and a natural language about the image and generate a natural language answer as output.



# What is VQA in the medical domain?

It takes as input a medical image and a clinical relevant question and output the answer based on the visual content.

Image



Question

what does ct image show?

Answer

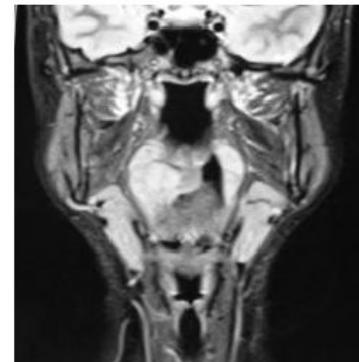
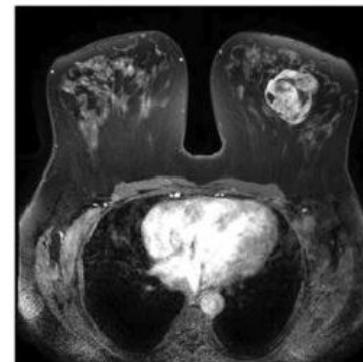
chest wall lesion encroaching on  
intercostal nerve.

when was imaging discordant on  
subsequent mri?

after growth

how does the mass look?

irregular oval-shaped



# Literature Review

# Medical Visual Question Answering (VQA-Med) task at ImageCLEF

ImageCLEF is an evaluation campaign offering several research tasks related to information retrieval, machine learning, natural language processing etc.,

It has included medical tasks since 2004 and teams from over the world have been competing to create a good model every year.

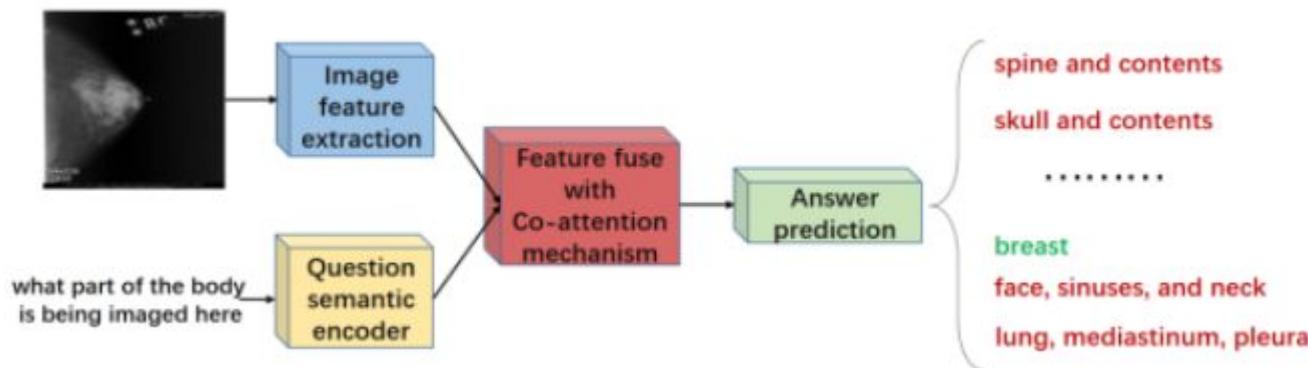
We focused on papers on ImageCLEF in the past few years.



# Common Approach

The model basically consists of the following modules:

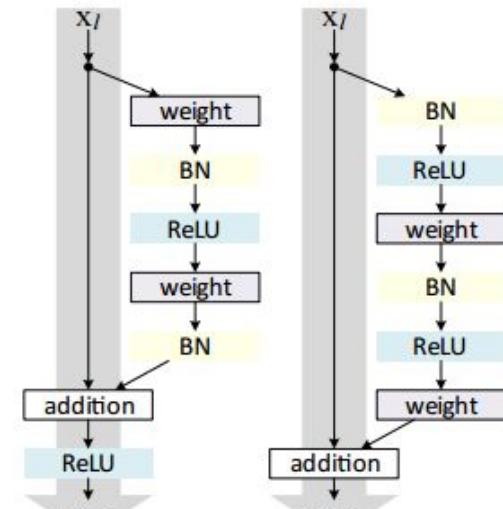
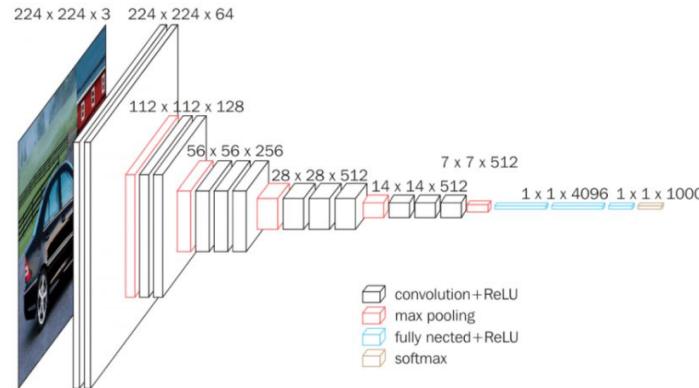
- Image feature extraction
- Question feature extraction
- Merge image features and question features



# Image feature extraction

Pretrained Convolutional Network (CNN) models

- VGG16
- VGG19
- ResNet-50
- ResNet-152
- (AlexNet)

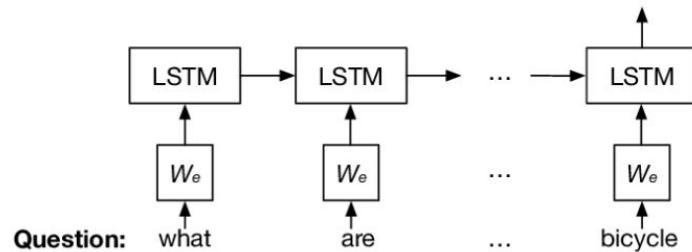


(a) original

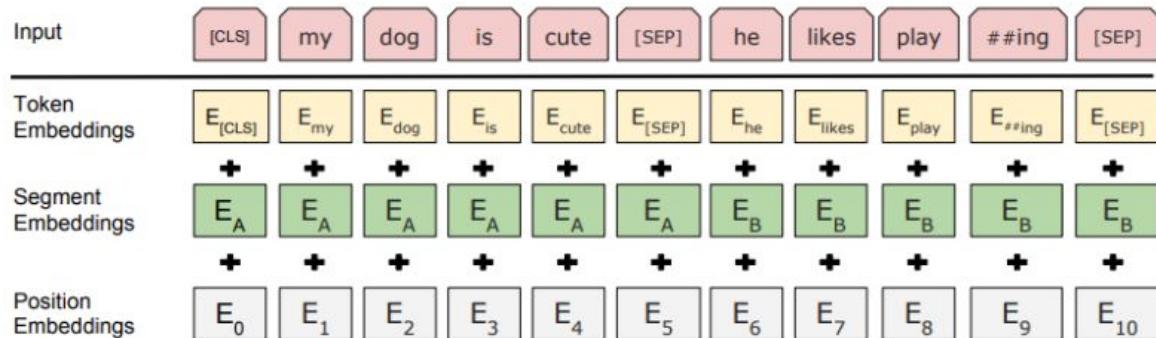
(b) proposed

# Question feature extraction

- LSTM (Long Short-Term Memory networks)



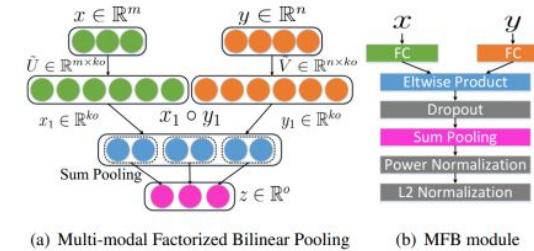
- BERT (Bidirectional Encoder Representations from Transformers)



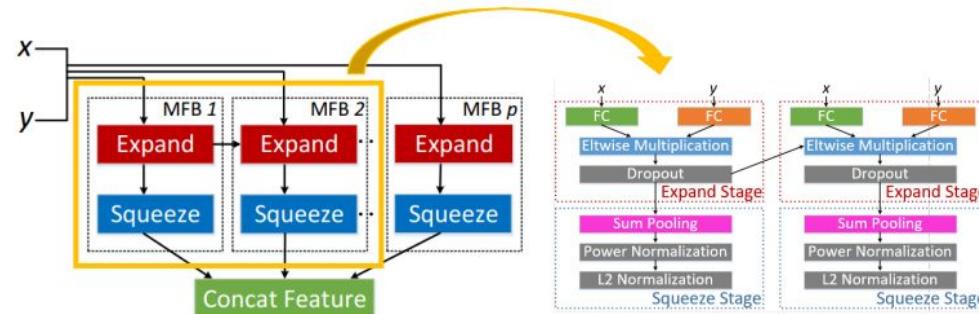
# Merge image features and question features

## Pooling approach

- Multimodal Factorized Bilinear (MFB) pooling



- Multimodal Factorized High-order (MFH) pooling



# Evaluation Methodology

- Accuracy:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

- BLEU (Bilingual Evaluation Understudy):

Calculate the similarity between the predicted answer and the actual answer

$$\text{BLEU} = \underbrace{\min\left(1, \exp\left(1 - \frac{\text{reference-length}}{\text{output-length}}\right)\right)}_{\text{brevity penalty}} \underbrace{\left(\prod_{i=1}^4 \text{precision}_i\right)^{1/4}}_{\text{n-gram overlap}}$$

# Our Project

# Product Mission

Our product is for doctors and patients to provide an answer on a specific question about a given clinical image in order to support clinical decision making and improve patient engagement opportunities without physically visiting a doctor.

# User Stories

- Doctors
  - Want to ask a question to interpret patients' clinical images
  - Want to get an answer to make a clinical decision
- Patients
  - Want to ask a question to know their disease-status based on their clinical image
  - Want to get an answer without visiting a doctor and searching engines by themselves

# MVP

- Input a clinical image and a question
- Output an answer based on the image

# Technologies

We have not decided yet, but will work on figuring out

- What fine-tuning CNN to pick? VGG16, VGG19 or ResNet?
- What optimizer to pick?
- What value to set for training rate?
- What techniques to use? Cross-validation, regularization, hyperparameter tuning etc..?

# Setup of Development Environment

- Tensorflow
- Keras
- ImageCLEF VQA-Med Dataset
- GPU

# References

- [1] Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C. L., Parikh, D., & Batra, D. (2016). VQA: Visual Question Answering. *International Journal of Computer Vision*, 123(1), 4–31. <https://doi.org/10.1007/s11263-016-0966-6>
- [2] Al-Sadi, A., Al-Ayyoub, M., Jararweh, Y., & Costen, F. (2021). Visual question answering in the medical domain based on deep learning approaches: A comprehensive study. *Pattern Recognition Letters*, 150, 57–75. <https://doi.org/10.1016/j.patrec.2021.07.002>
- [3] Talafha, B., & Al-Ayyoub, M. (2018). JUST at VQA-Med: A VGG-Seq2Seq Model. CLEF.

Thank you