Continual VQA for Disaster Response Systems

Aditya Kane

V Manushree

Sahil Khose

Introduction

- Scenario: During floods aerial data is crucial to assess the affected areas.

Introduction

- Scenario: During floods aerial data is crucial to assess the affected areas.
- Existing work: VQA on FloodNet Kane and Khose (2022). However their system lacks real world deployability.

Introduction

- Scenario: During floods aerial data is crucial to assess the affected areas.
- Existing work: VQA on FloodNet Kane and Khose (2022). However their system lacks real world deployability.
- Problem: Delay for label generation.

Introduction

- Scenario: During floods aerial data is crucial to assess the affected areas.
- Existing work: VQA on FloodNet Kane and Khose (2022). However their system lacks real world deployability.
- Problem: Delay for label generation.
- Solution 1: Don't use labels! - *Zero-shot VQA*

Introduction

- Scenario: During floods aerial data is crucial to assess the affected areas.
- Existing work: VQA on FloodNet Kane and Khose (2022). However their system lacks real world deployability.
- Problem: Delay for label generation.
- Solution 1: Don't use labels! - *Zero-shot VQA*
- Solution 2: Train your model continually as labels are generated - *Continual VQA*

FloodNet dataset

Reference: https://arxiv.org/abs/2012.02951

FloodNet dataset

Types of tasks

1. Classification
2. Segmentation
3. VQA

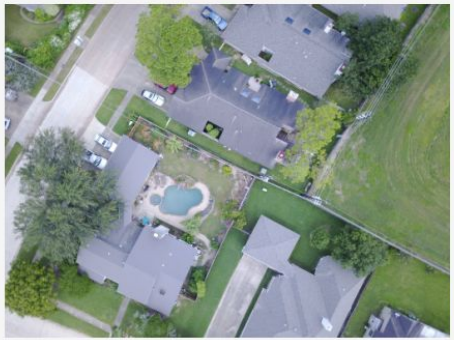


Image Class: Non-Flooded

Image Class: Flooded

Reference: https://arxiv.org/abs/2012.02951

FloodNet dataset

Types of tasks

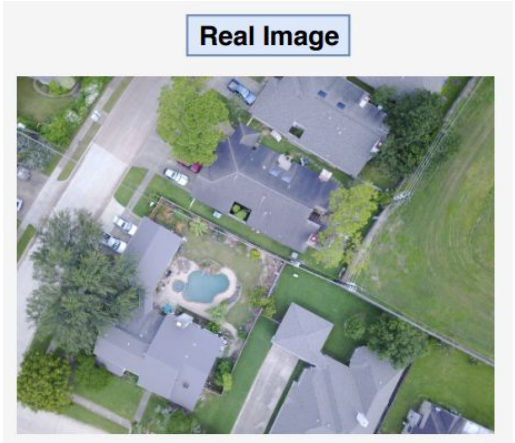
1. Classification
2. Segmentation
3. VQA

Reference: https://arxiv.org/abs/2012.02951

FloodNet dataset

Types of tasks

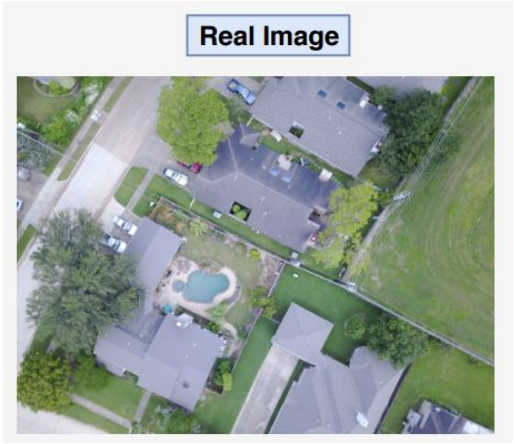
1. Classification
2. Segmentation
3. VQA

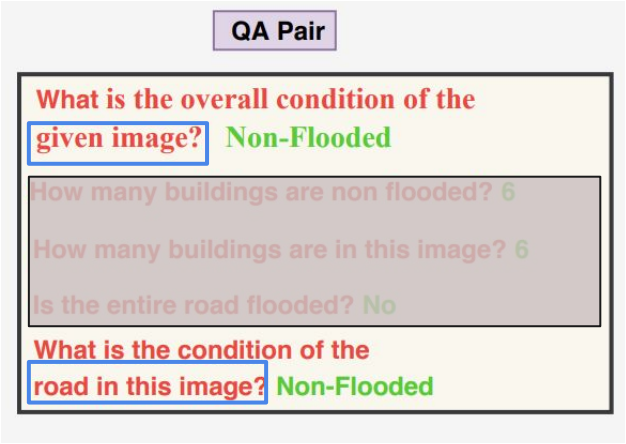


Reference: https://arxiv.org/abs/2012.02951

Types of questions for VQA

1. Condition Recognition
   1.1. Image Condition
   1.2. Road Condition
2. Yes/No
3. Counting Problem
   3.1. Simple Count
   3.2. Complex Count



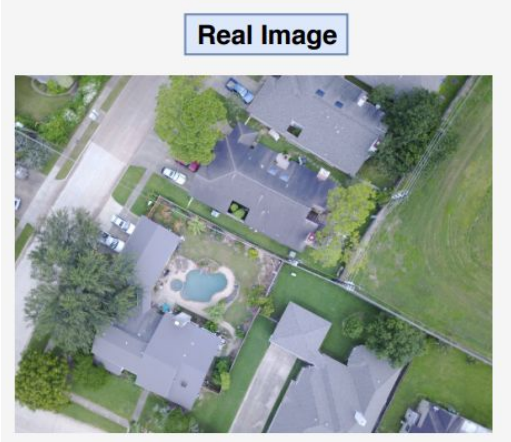


Types of questions for VQA

1. Condition Recognition
   1.1. Image Condition
   1.2. Road Condition
2. Yes/No
3. Counting Problem
   3.1. Simple Count
   3.2. Complex Count

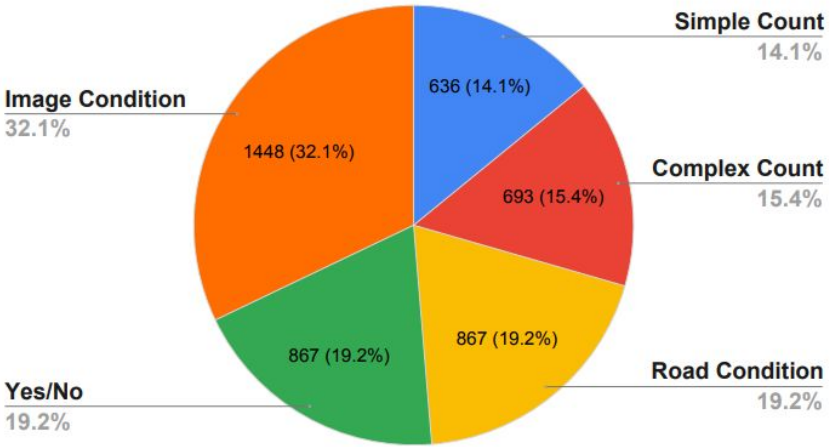
Types of questions for VQA

1. Condition Recognition
   1.1. Image Condition
   1.2. Road Condition
2. Yes/No
3. Counting Problem
   3.1. Simple Count
   3.2. Complex Count

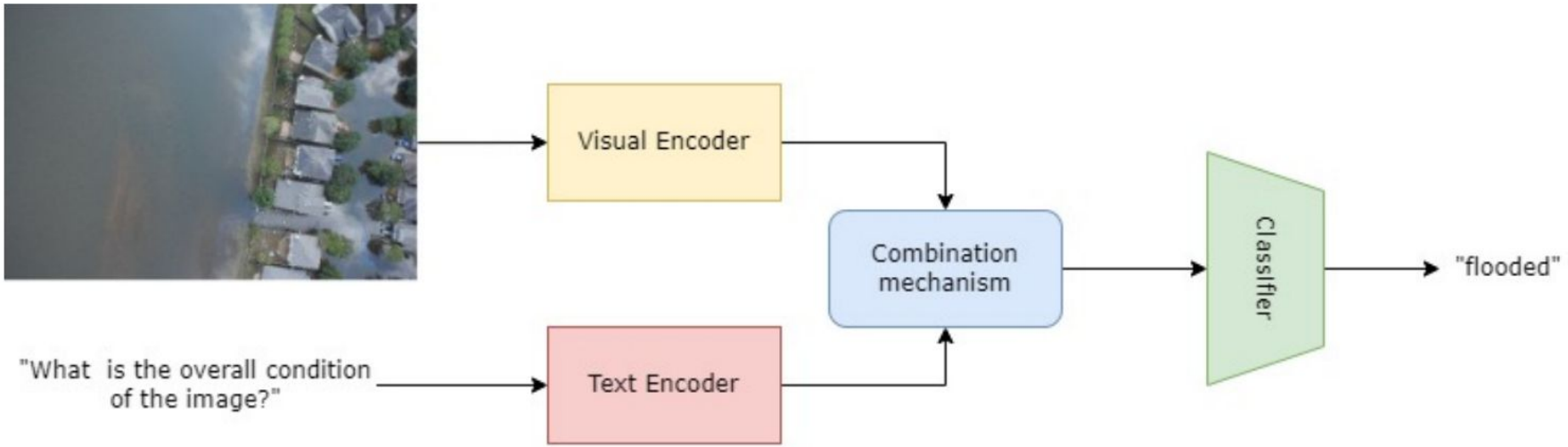


Types of questions for VQA

1. Condition Recognition
   1.1. Image Condition
   1.2. Road Condition
2. Yes/No
3. Counting Problem
   3.1. Simple Count
   3.2. Complex Count

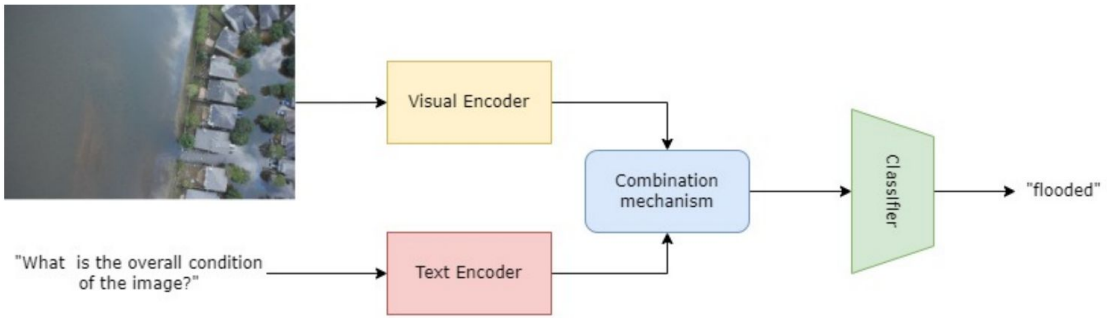


VQA Pipeline



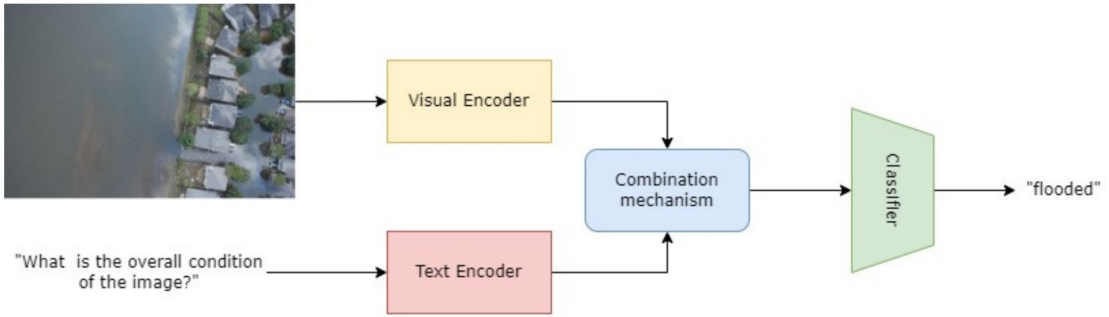
Our Approach for FloodNet VQA

- **Zero-shot VQA system**
  - Multi-modal pre-trained models
- **Continual VQA system**
  - Trained on continuous stream of data



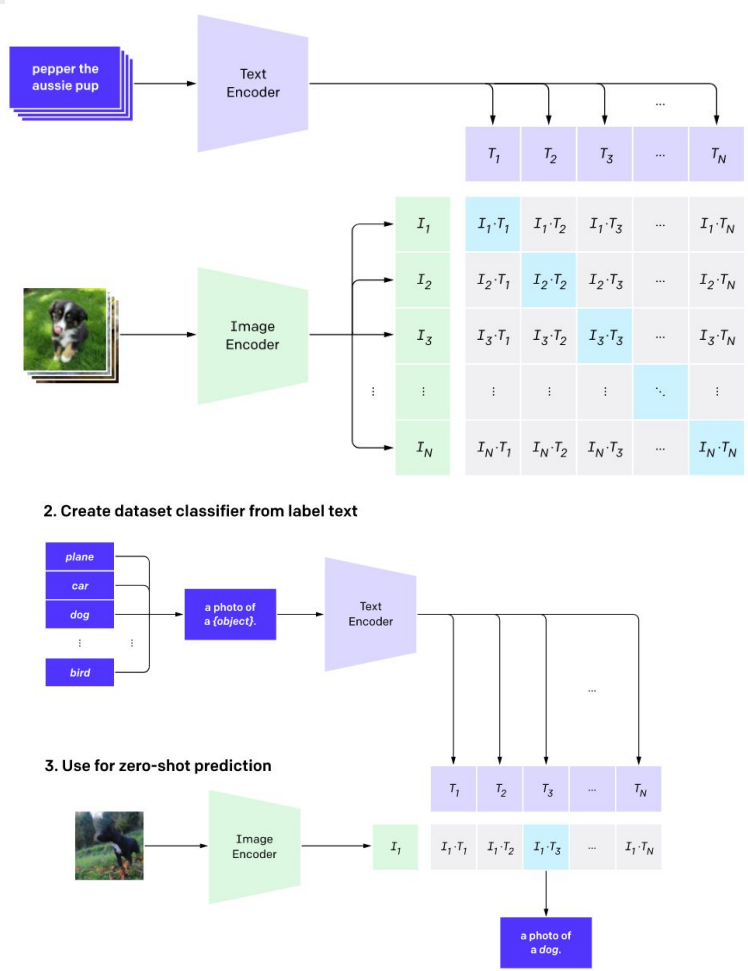
Our Approach for FloodNet VQA

- **Zero-shot VQA system**
  - Multi-modal pre-trained models
- **Continual VQA system**
  - Trained on continuous stream of data



Our Approach for FloodNet VQA

- **Zero-shot VQA system**
  - 1. CLIP out-of-the-box
  - 2. CLIP features for supervised training
- **Continual VQA system**
  - Trained on continuous stream of data



Results

- **Zero-shot VQA system**
  - 1. CLIP out-of-the-box
  - 2. CLIP features for supervised training

Kane and Khose (2022)

Method	Taskwise Accuracy			
	Overall	Yes/No	Image Condition	Road Condition
CNX-mul	98.03	98.31	**98.62**	97.18

Results

- **Zero-shot VQA system**
  - 1. CLIP out-of-the-box
  - 2. CLIP features for supervised training

Kane and Khose (2022)

CLIP out-of-the-box (ZS)

Method	Taskwise Accuracy			
	Overall	Yes/No	Image Condition	Road Condition
CNX-mul	98.03	98.31	**98.62**	97.18
CLIP-ZS	35.56	15.12	41.72	83.14

Results

- **Zero-shot VQA system**
  - 1. CLIP out-of-the-box
  - 2. CLIP features for supervised training

Kane and Khose (2022)

CLIP out-of-the-box (ZS)

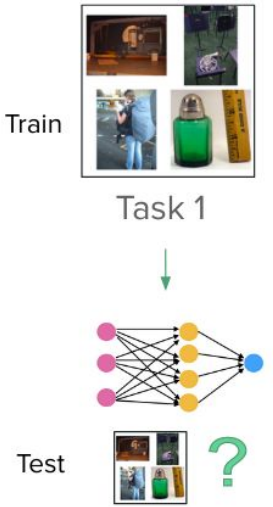
CLIP supervised (Best)

Method	Taskwise Accuracy			
	Overall	Yes/No	Image Condition	Road Condition
CNX-mul	98.03	98.31	**98.62**	97.18
CLIP-ZS	35.56	15.12	41.72	83.14
CLIP-add	93.99	88.37	95.17	97.67
CLIP-cat	92.97	81.97	95.86	97.06
CLIP-mul	96.4	97.71	95.17	97.14
CLIP-mul-taskwise	**98.33**	**98.85**	98.43	**97.71**

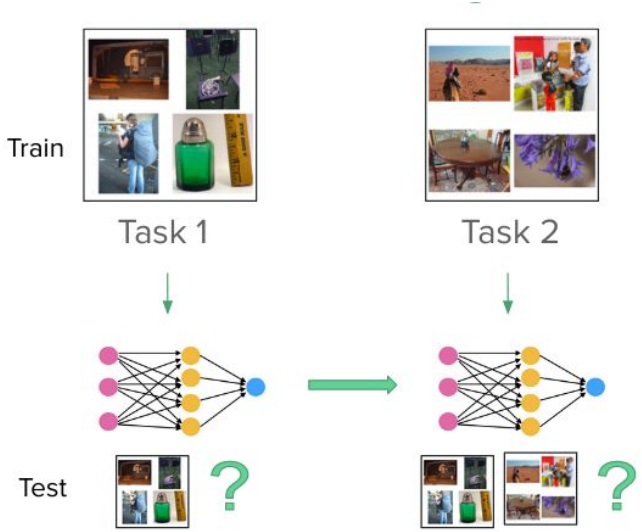
Our Approach for FloodNet VQA

- Zero-shot VQA system
  - 1. CLIP out-of-the-box
  - 2. CLIP features for supervised training
- Continual VQA system
  - Trained on continuous stream of data

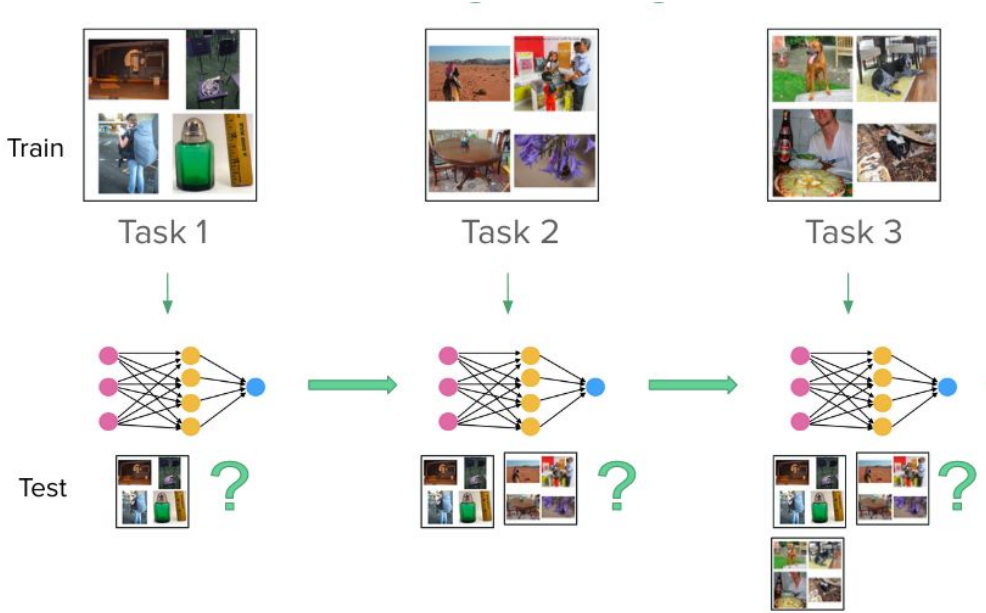
Continual Learning



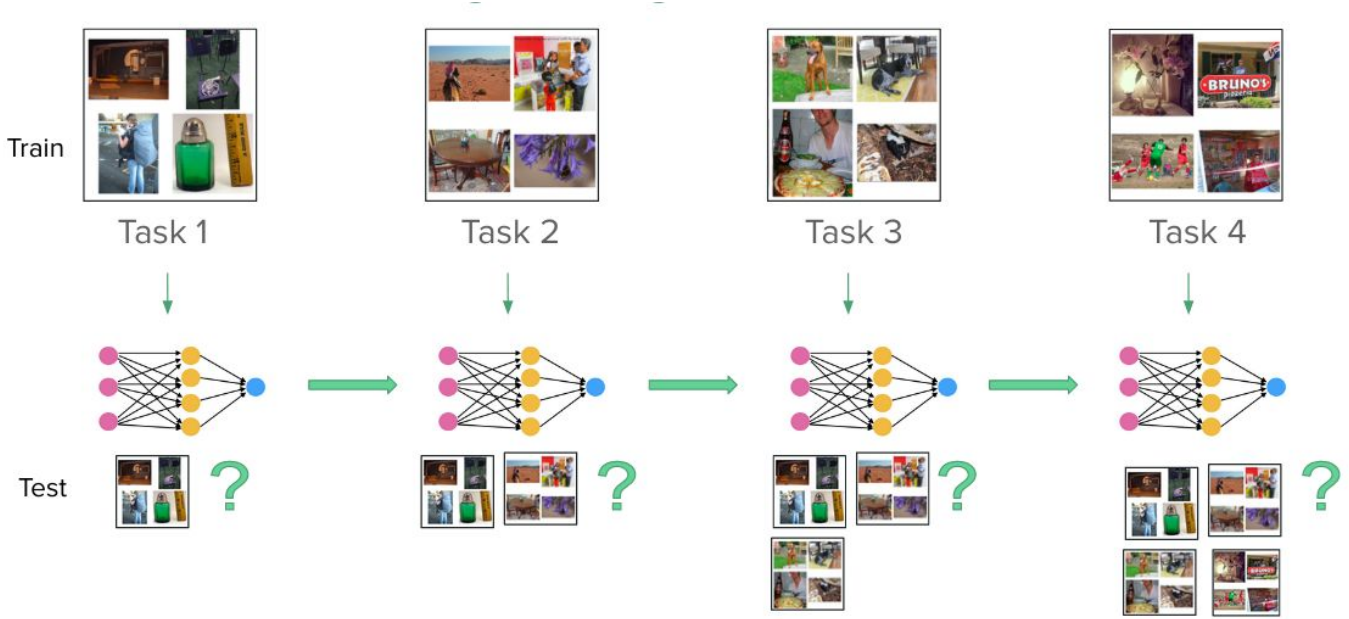
Continual Learning



Continual Learning

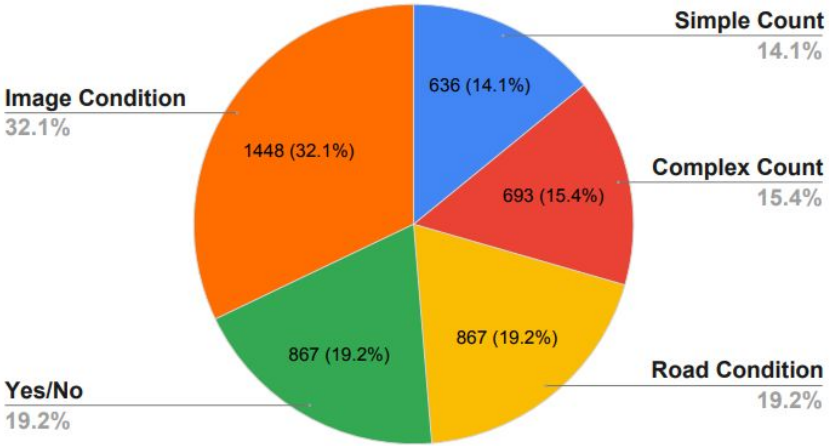


Continual Learning



Our Tasks for Continual VQA

1. Task 1: Image Condition
2. Task 2: Road Condition
3. Task 3: Yes/No



Experience replay methods used

1. Reservoir Sampling Update
2. Ring Buffer
3. Mean of features

(a) Reservoir Sampling Update

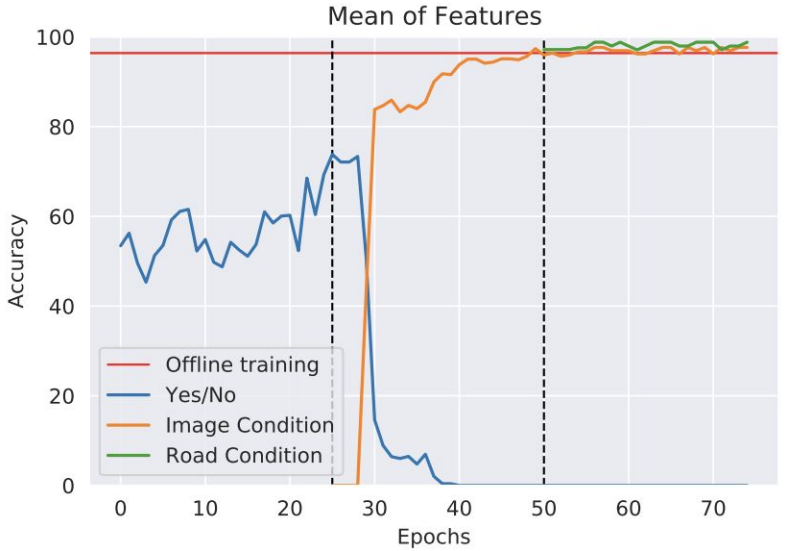
Experience replay methods used

1. Reservoir Sampling Update
2. Ring Buffer
3. Mean of features

(b) Ring Buffer

Experience replay methods used

1. Reservoir Sampling Update
2. Ring Buffer
3. Mean of features

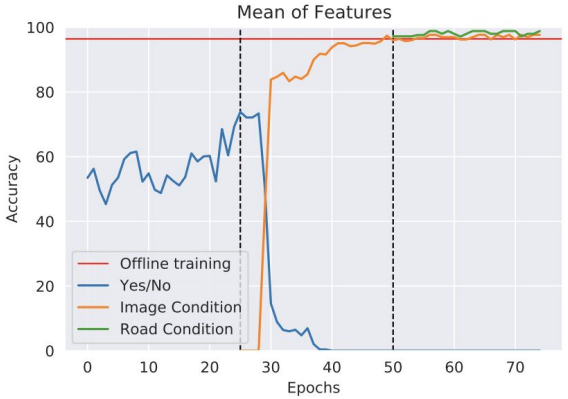


(c) Mean of Features

CL Results

(a) Reservoir Sampling Update

(b) Ring Buffer



(c) Mean of Features

Observations

1. CLIP out-of-the-box performs poorly
2. CLIP features for supervised training outperforms state-of-the-art on FloodNet VQA
3. For Continual Learning: Reservoir Sampling performs the best.
4. "Image Condition" and "Road Condition" tasks yield consistent results.
5. "Yes/No" task hampers the performance.

Observations

1. CLIP out-of-the-box performs poorly
2. CLIP features for supervised training outperforms state-of-the-art on FloodNet VQA
3. For Continual Learning: Reservoir Sampling performs the best.
4. "Image Condition" and "Road Condition" tasks yield consistent results.
5. "Yes/No" task hampers the performance.

Observations

1. CLIP out-of-the-box performs poorly
2. CLIP features for supervised training outperforms state-of-the-art on FloodNet VQA
3. For Continual Learning: Reservoir Sampling performs the best.
4. "Image Condition" and "Road Condition" tasks yield consistent results.
5. "Yes/No" task hampers the performance.

Observations

1. CLIP out-of-the-box performs poorly
2. CLIP features for supervised training outperforms state-of-the-art on FloodNet VQA
3. For Continual Learning: Reservoir Sampling performs the best.
4. "Image Condition" and "Road Condition" tasks yield consistent results.
5. "Yes/No" task hampers the performance.

Observations

1. CLIP out-of-the-box performs poorly
2. CLIP features for supervised training outperforms state-of-the-art on FloodNet VQA
3. For Continual Learning: Reservoir Sampling performs the best.
4. "Image Condition" and "Road Condition" tasks yield consistent results.
5. "Yes/No" task hampers the performance.

Impact

- We build efficient systems to mitigate delay in assessing flood affected areas.
- Our model learns on-the-fly on new tasks.
- We tackle the delay in label generation.

Impact

- We build efficient systems to mitigate delay in assessing flood affected areas.
- Our model learns on-the-fly on new tasks.
- We tackle the delay in label generation.

Impact

- We build efficient systems to mitigate delay in assessing flood affected areas.
- Our model learns on-the-fly on new tasks.
- We tackle the delay in label generation.

Conclusion

- We explore the Continual VQA setup, which has been unexplored before, especially in disaster response tasks like FloodNet.
- We propose a light-weight model with online training capability and also suggest the order of task labels to be provided for optimal performance of the deployed model.

Conclusion

- We explore the Continual VQA setup, which has been unexplored before, especially in disaster response tasks like FloodNet.
- We propose a light-weight model with online training capability and also suggest the order of task labels to be provided for optimal performance of the deployed model.

Thank you!

Feel free to reach us out in case of any questions

Aditya Kane
adityakane1@gmail.com

V Manushree
manushree635@gmail.com

Sahil Khose
sahil.khose@gatech.edu

Check out our paper and code!

- Paper: arxiv.org/abs/2209.10320
- Code: github.com/AdityaKane2001/continual_vqa
  (give us a ⭐ if you like our work!)