

MULTI-MODAL MISINFORMATION DETECTION USING VQA





MOHIT SHARMA

PROBLEM STATEMENT & DATASET:

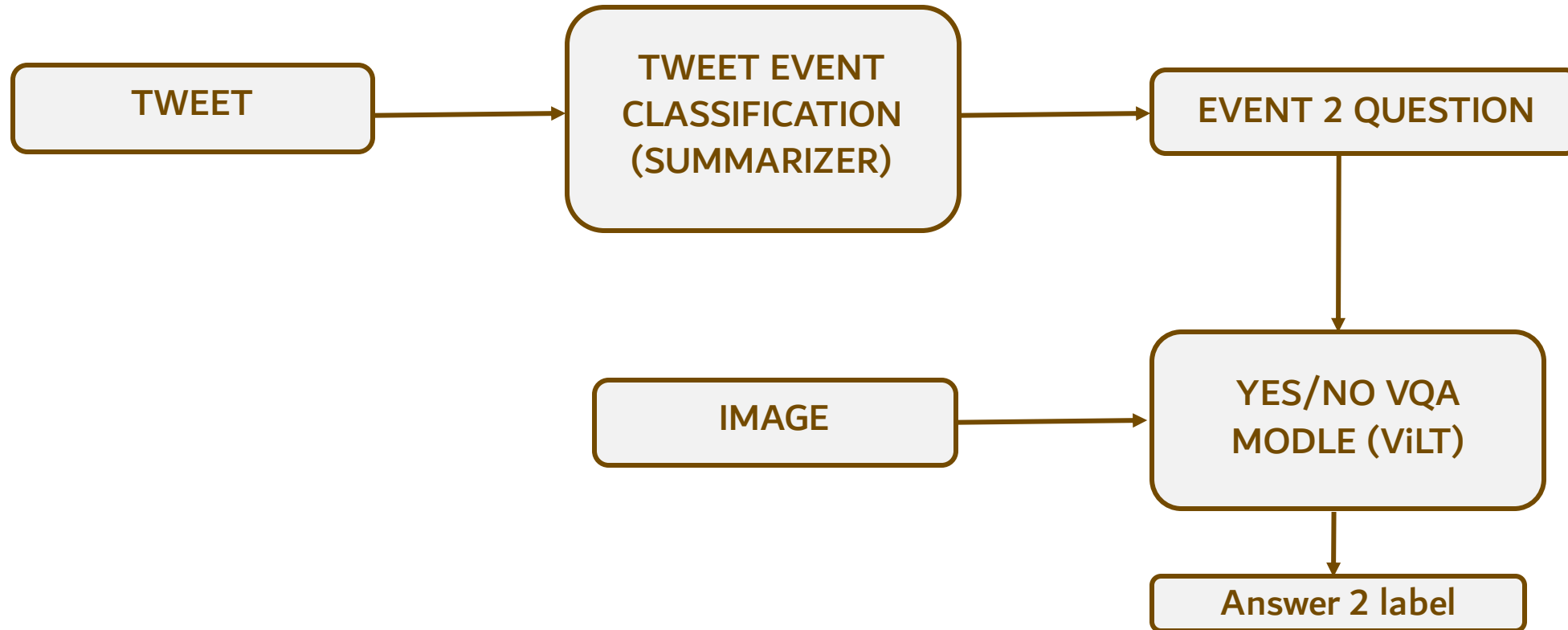


- Classify a multi-modal information sample (text-image pair) as fake or real
- VQA should be utility of choice to query the image
- Dataset: MediaEval 2015 task, tweet and image pairs
- A total of 16326 samples generated from around 500 unique images

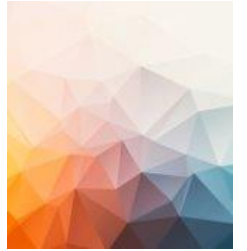
DATASET:

			
<p>Just known this bcs of #jason #chen updated the pic! Everyone be safe! #newyork #sandy #hurricane #nature</p>	<p>Wow RT@billmckibben: Boardwalk floating in sections through the flooded streets of Atlantic City</p>	<p>Wow, so Sunil Tripathi is a Brown U Student that has been missing since March...</p>	<p>RIP Tamerlan and free Dzhokhar Tsarnaev. This corrupt society will never stop blaming people.</p>
<p>fake</p>	<p>real</p>	<p>fake</p>	<p>real</p>

MODEL:



DESIGN CHOICES:

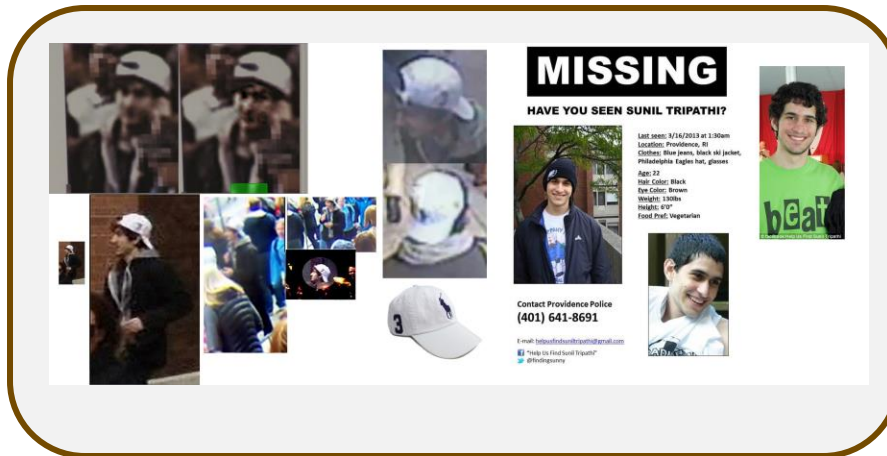


- Extraction of data from tweet: Tweet Event Classification (15 unique events), Summarisation as alternative.
 - Event 2 Question (yes/no questions, Is this & Does this as prefixes)
 - Binary VQA module (pre-trained Vision-and-Language Transformer (ViLT), fine-tuned on VQAv2)
-

**PIC: Comparison
of #Boston
suspect Sunil
Tripathi's FBI-
released
images/video and
his MISSING
poster**

**TWEET EVENT
CLASSIFIED AS
"boston marathon
bombing"**

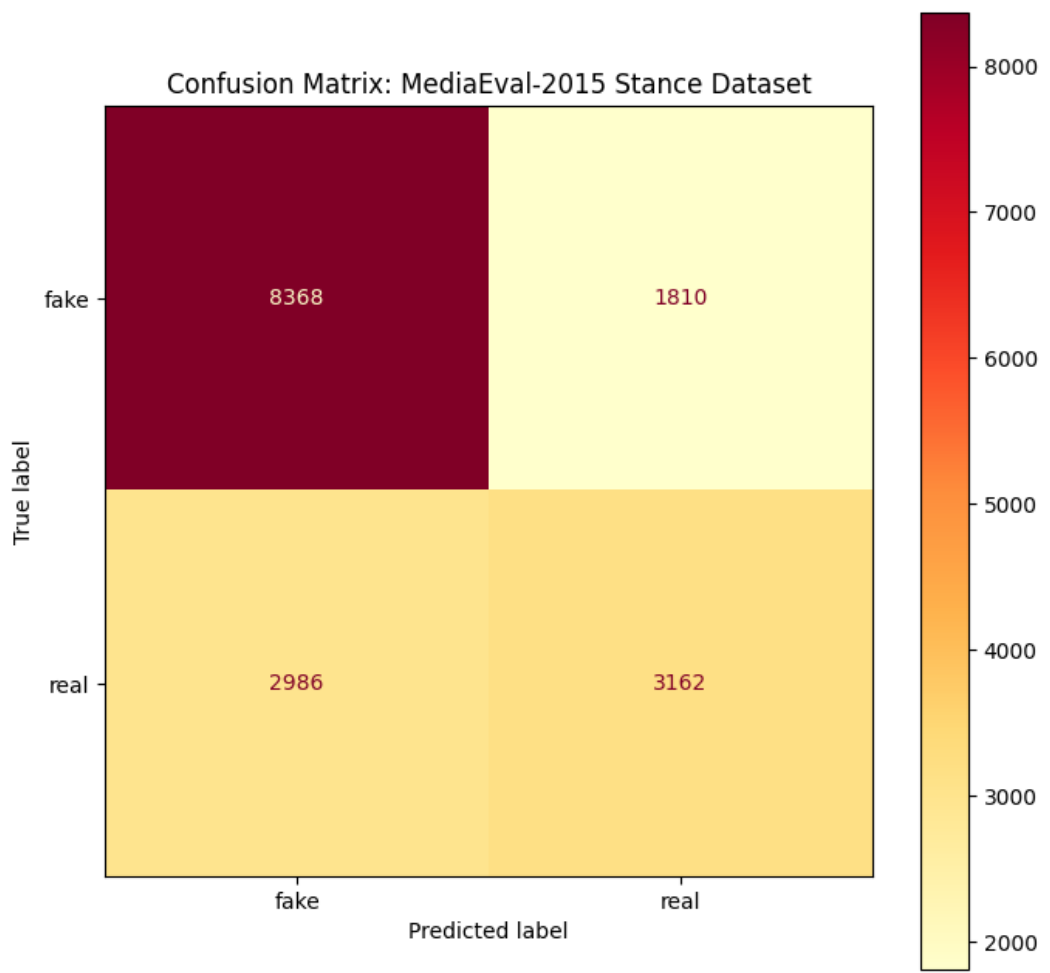
**Question: Does this image
correspond to boston marathon
bombing?**



**YES/NO VQA
MODLE (ViLT)**

no (fake)

RESULTS:



	precision	recall	f1-score	support
fake	0.74	0.82	0.78	10178
real	0.64	0.51	0.57	6148
accuracy			0.71	16326
macro avg	0.69	0.67	0.67	16326
weighted avg	0.70	0.71	0.70	16326