

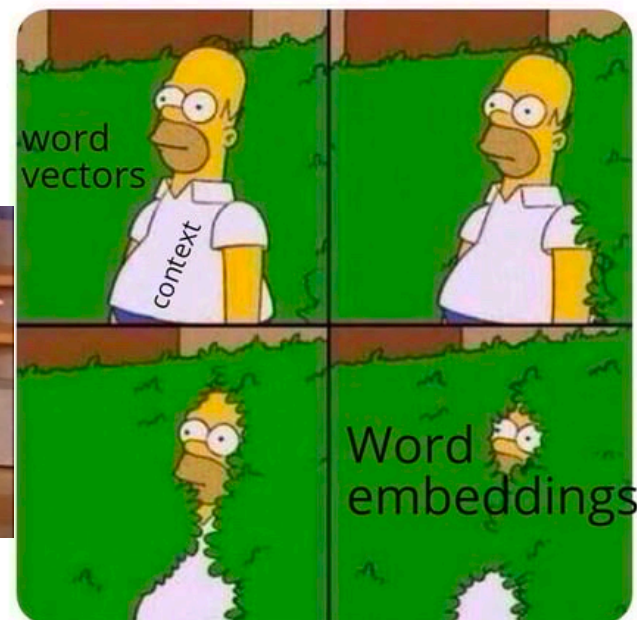
May 10, 2023.

DSML: NLP module.

Word embeddings
in a nutshell

Attention

Class starts
@ 9:05



When you penalize your Natural Language Generation model for large sentence lengths

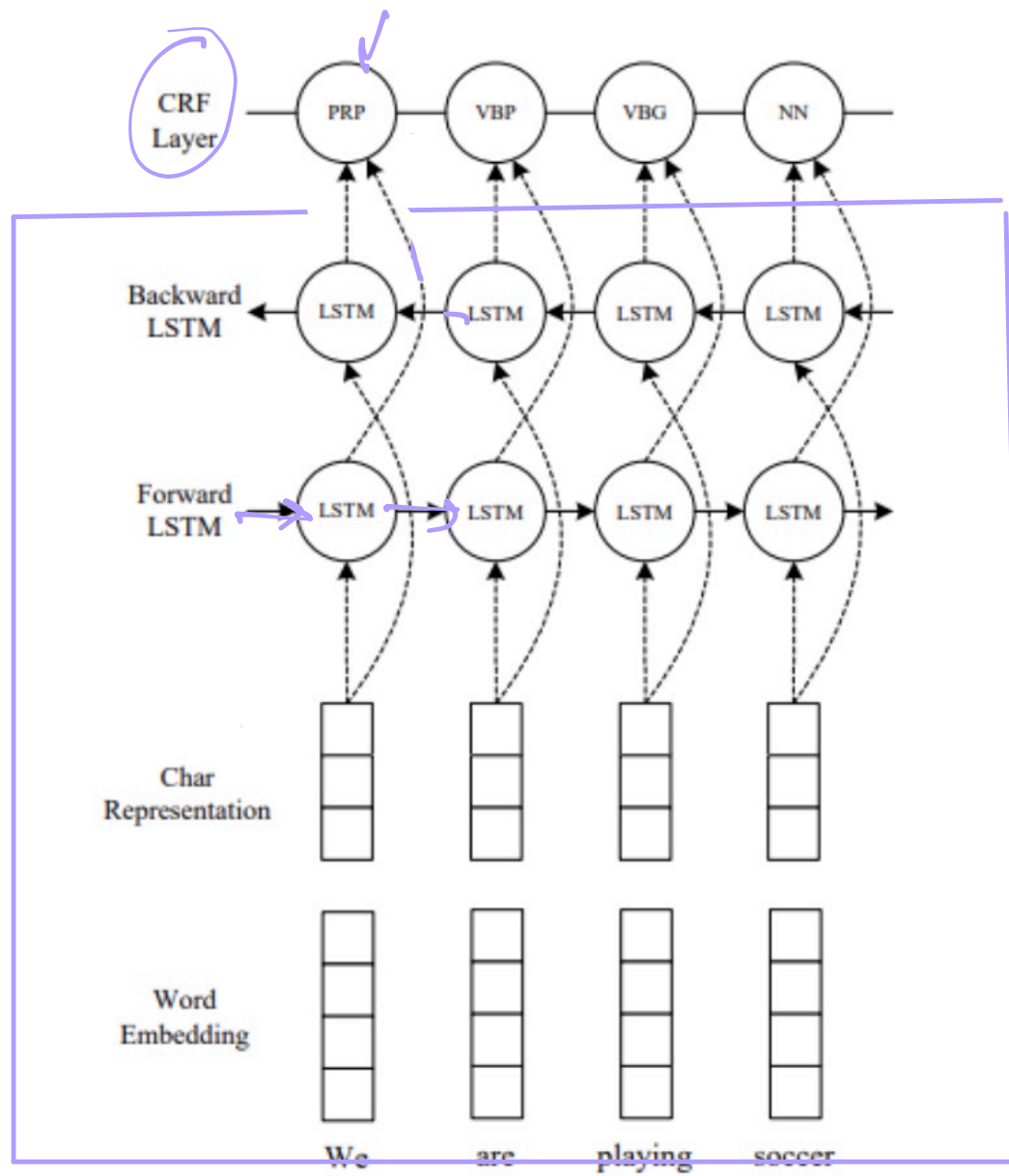


Recap:

- * Named Entity Recognition.
- * Bidirectional LSTMs.
- * Conditional Random Field (CRF).

Agenda:

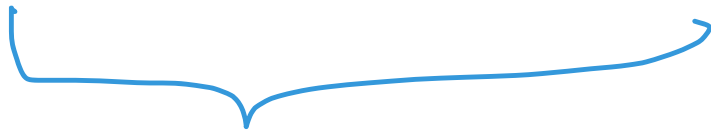
- * Bidirectional LSTMs + CRF. (code)
- * Stanza → Uses this technique.
- * Introduction to attention mechanisms.



An orange fat cat called Garfield...

<DT>

<N>



I - Animal.

Attention:

RNN

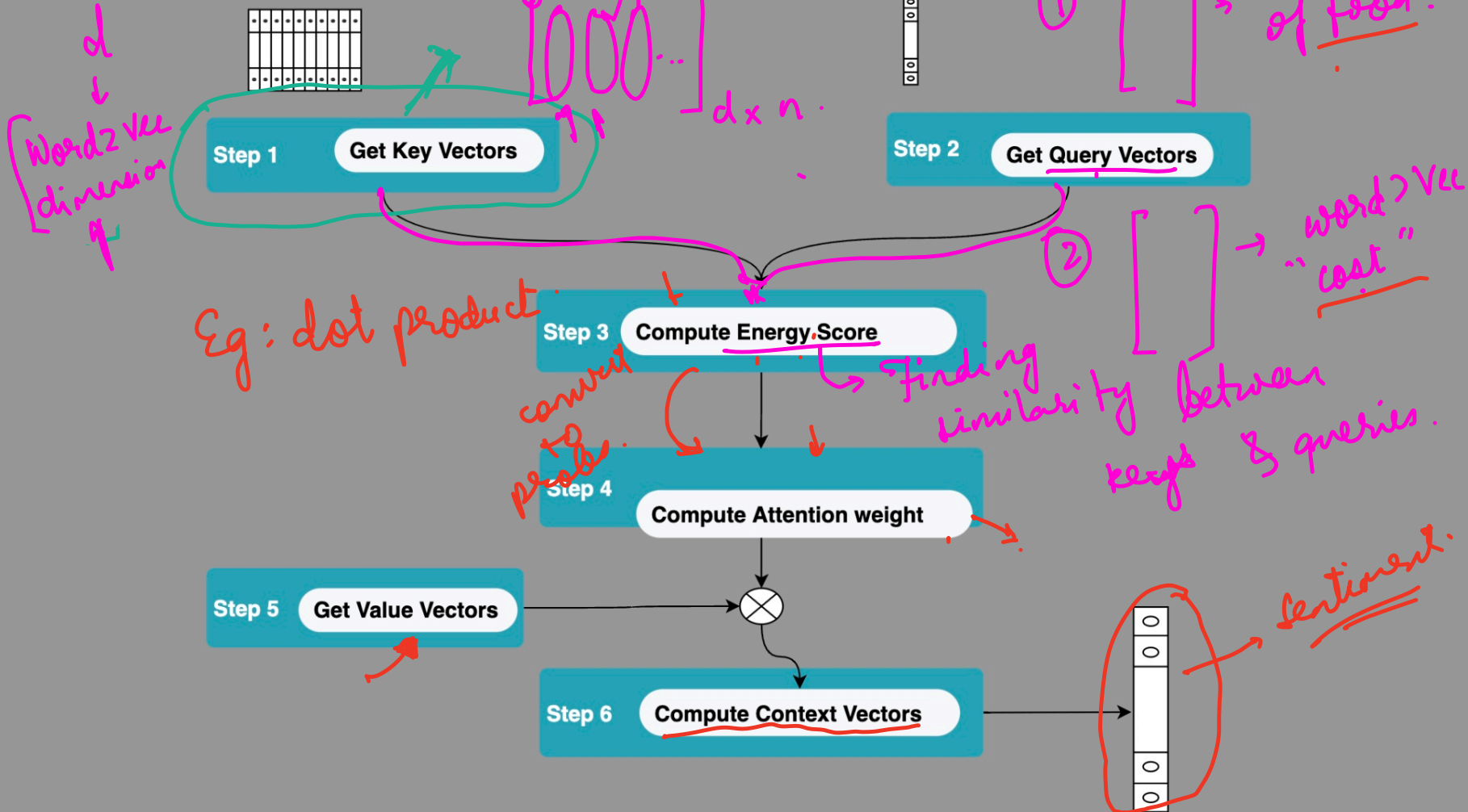
LSTM

GRUs etc.



"food is delicious but price is high" n

Attention Mechanisms Computation Flow



In the example,

the example,

① KEY $\rightarrow n \times d$

\uparrow length of sentence

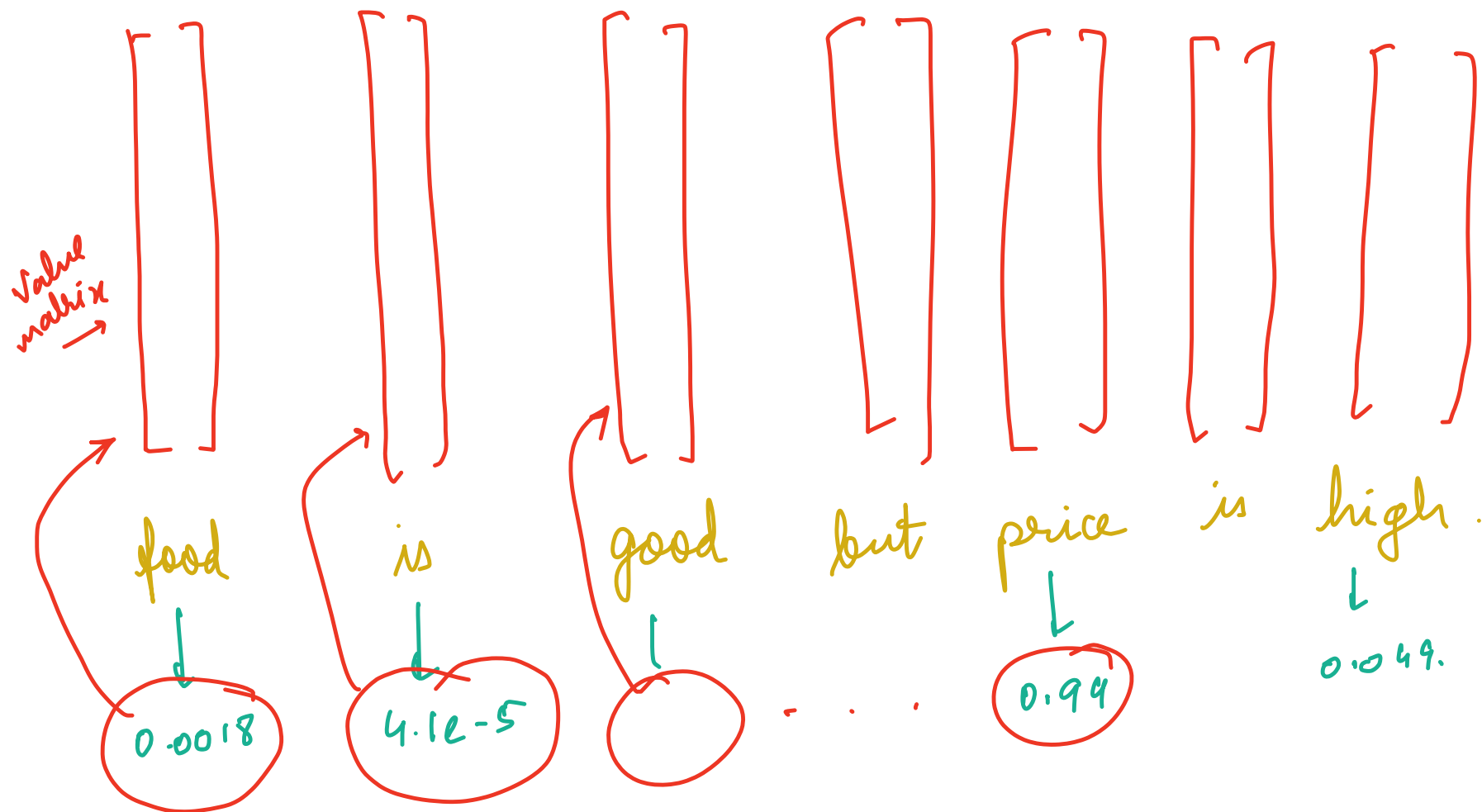
length of embedding.

② QUERY $\rightarrow d \times 1$ vector.

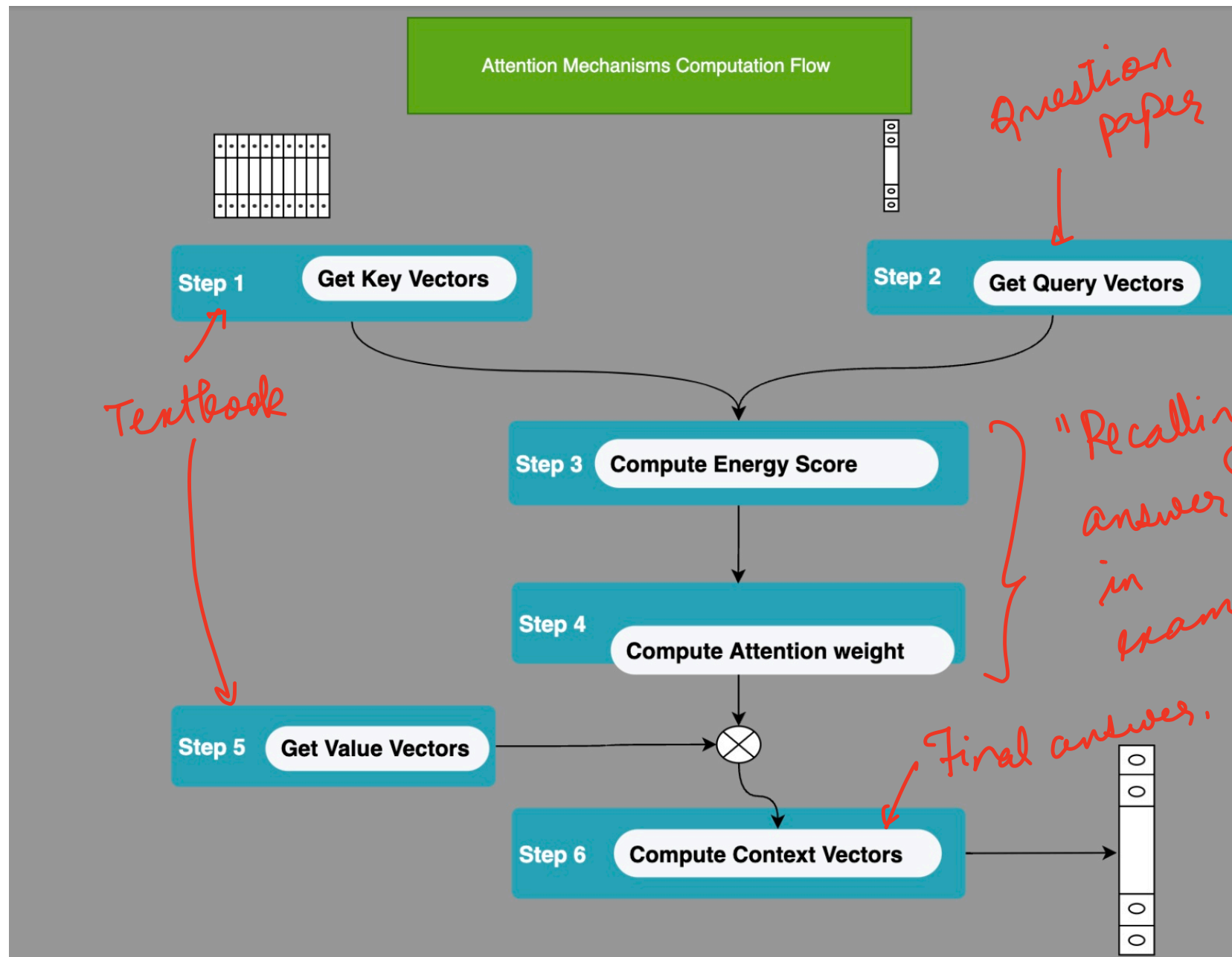
③ Energy score $\rightarrow n \times 1$

④ Attention weights $\rightarrow n \times 1$





Analogy:



$$-(1 - p_t)^\gamma \log(p_t)$$

Issue: Class imbalance.

$t \rightarrow$ high representation, $p_t \rightarrow$ high.

Q] Paper + Code for vision based attention.