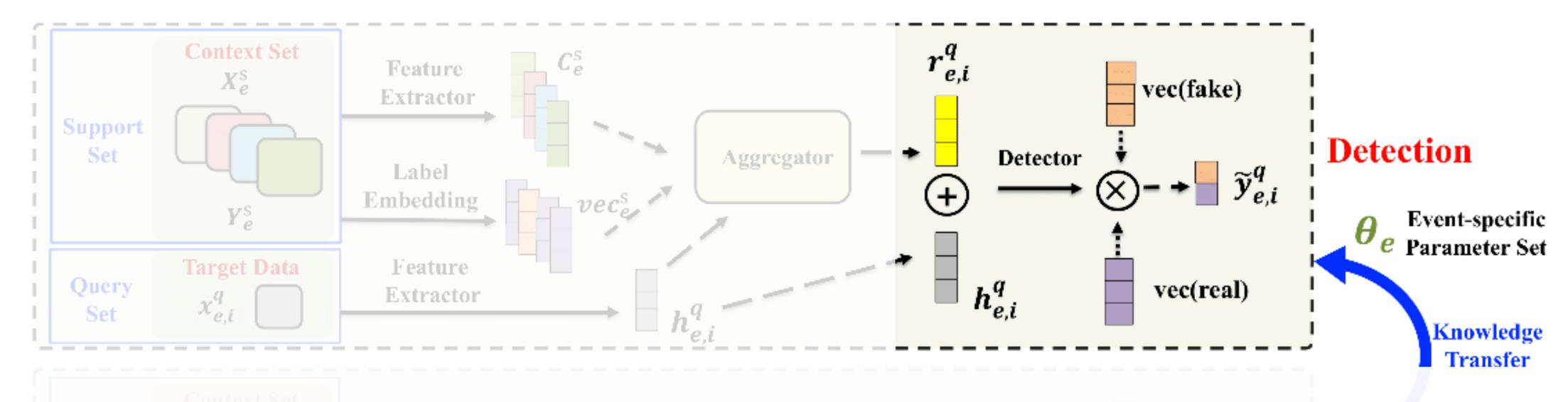# Methodology
Neural Network Architecture

- Thereafter both context embedding and target feature vector are fed into detector to output a vector.

- The final prediction is based on the similarities between this output vector and label embeddings.

Methodology
Feature Extractor

- **Textual feature extractor**

  - Text-CNN

  - Use 300-dimensional FastText pre-trained word-embedding

- **Visual feature extractor**

  - Pre-trained VGG-19

- On the top of extractors, both add a fully connected layer to adjust dimension to d_f.

- The output of two extractors are concatenated together to form a feature vector.