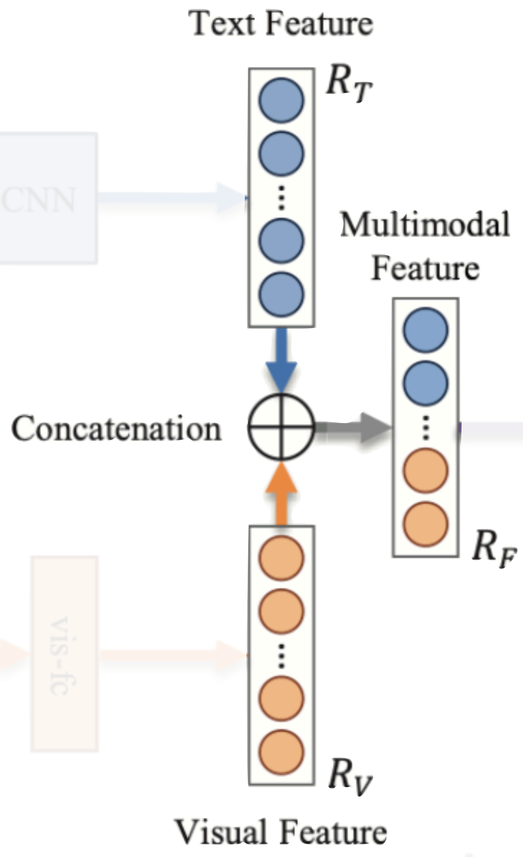
# Methodology....

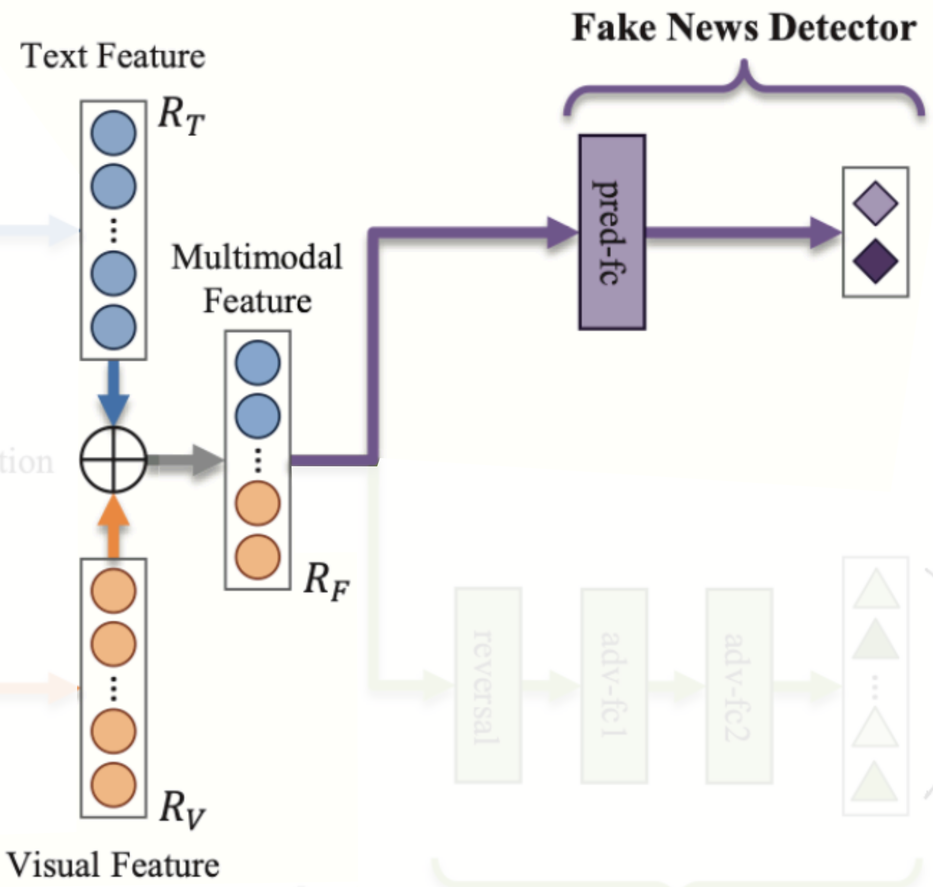
## Multi-model Feature Extractor




- Concatenated to form the multi-model feature representation denoted as
  - $R_F = R_T \oplus R_R \in \mathbb{R}^{2p}$
- Denote the multi-model feature extractor
  - $G_f(M; \theta_f)$
  - $M$: a set of textual and visual posts
  - $\theta_f$: parameter to be learned

# Methodology.....

## Fake News Detector




- Denote as $G_d(\cdot;\theta_d)$ , $\theta_d$: detector parameters
- Deploy a fully connected layer with softmax to predict the post are fake or real.
- **Probability** of post $m_i$ being a fake one:
  - $P_\theta(m_i) = G_d(G_f(m_i;\theta_f);\theta_d)$
- Employ <u>**cross entropy**</u> to calculate the **detection loss**:
  - $L_d(\theta_f,\theta_d) = -\mathbb{E}_{(m,y)\sim(M,Y_d)}[y\log(P_\theta(m)) + (1-y)(\log(1-P_\theta(m))]$
- <u>Minimize loss</u> function by seeking the optimal parameters $\theta_f, \theta_d$
  - $(\hat{\theta}_f,\hat{\theta}_d) = arg\ \min_{\theta_f,\theta_d} L_d(\theta_f,\theta_d)$