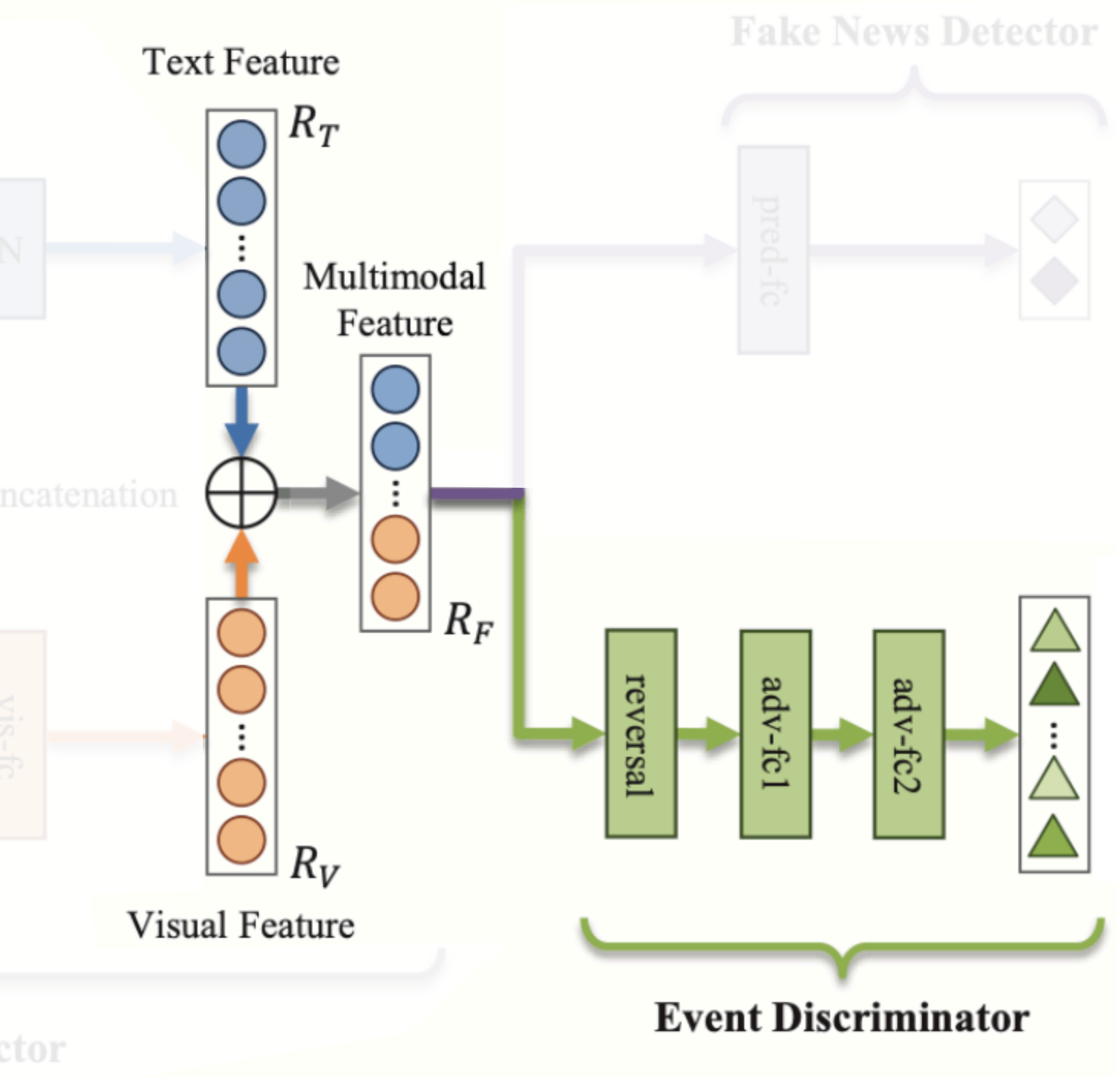
Methodology........

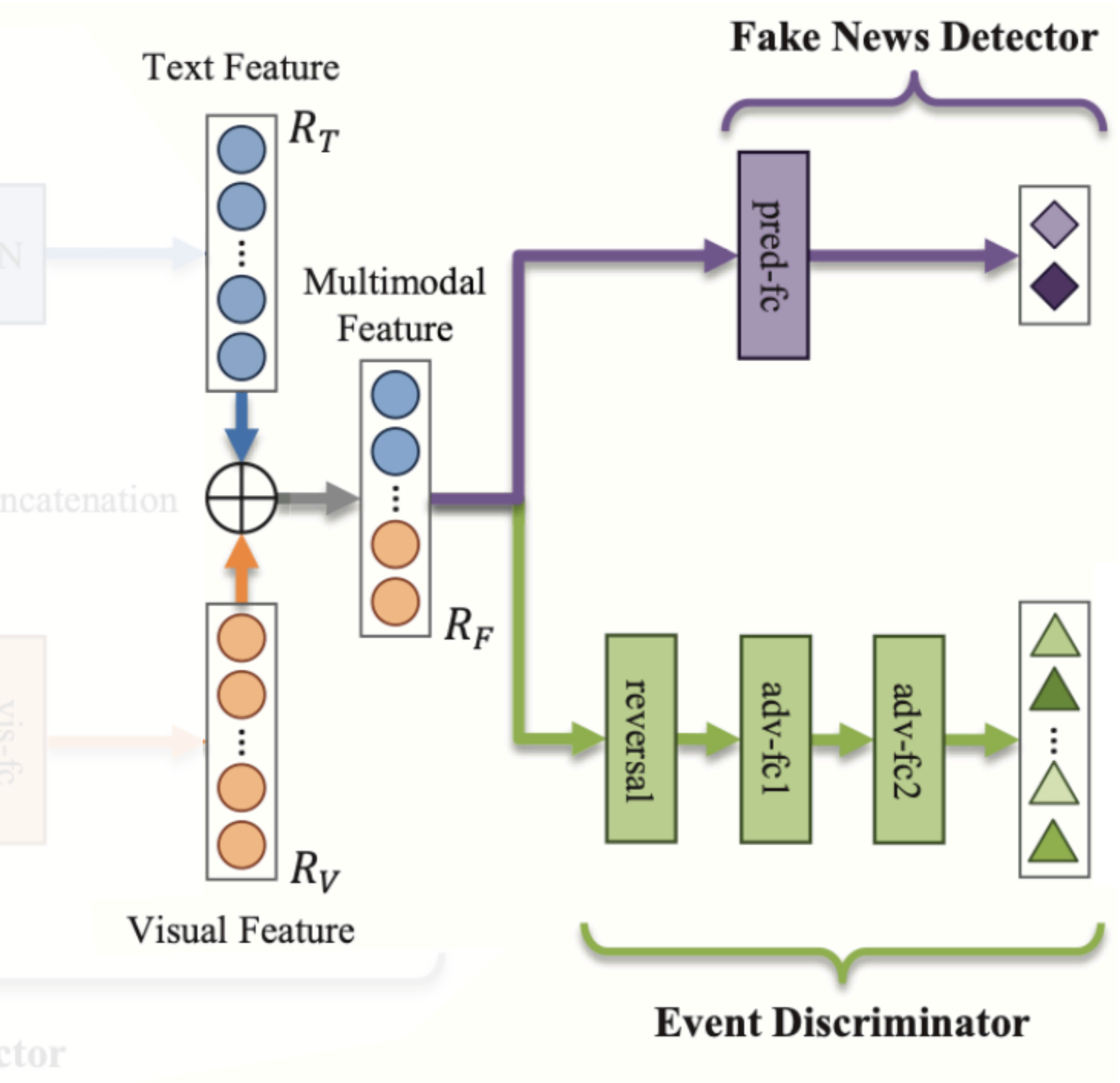
Event Discriminator



- Loss function by <u>cross entropy</u>:
  - $L_e(\theta_f, \theta_e) = -\mathbb{E}_{(m,y)\sim(M,Y_e)}[\sum_{k=1}^{K} 1_{[k=y]} \log(G_e(G_f(m;\theta_f);\theta_e))]$
- Parameters minimizing the loss $L_e(\cdot, \cdot)$:
  - $\hat{\theta}_e = arg \min_{\theta_e} L_e(\theta_f, \theta_e)$
- Large loss means the events' representations are <u>similar and the learned feature are event-invariant</u>.
  - Need to <u>**maximize**</u> the $L_e(\theta_f, \hat{\theta}_e)$ by seeking the optimal parameters θ_f

Methodology………

Model Integration



- $G_f(\cdot\,;\theta_f)$ need to cooperate with $G_d(\cdot\,;\theta_d)$ to **minimize the $L_d(\theta_f, \theta_d)$** to improve performance

- $G_f(\cdot\,;\theta_f)$ tries to fool $G_e(\cdot\,;\hat{\theta}_e)$ to achieve event-invariant representations by maximizing $L_e(\theta_f, \theta_e)$

- Define loss of this three-player game as

  - $L_{final}(\theta_f, \theta_d, \theta_e) = L_d(\theta_f, \theta_d) - \lambda L_e(\theta_f, \theta_e)$

  - In this paper, simply set $\lambda = 1$ to without tuning the trade-off parameter.