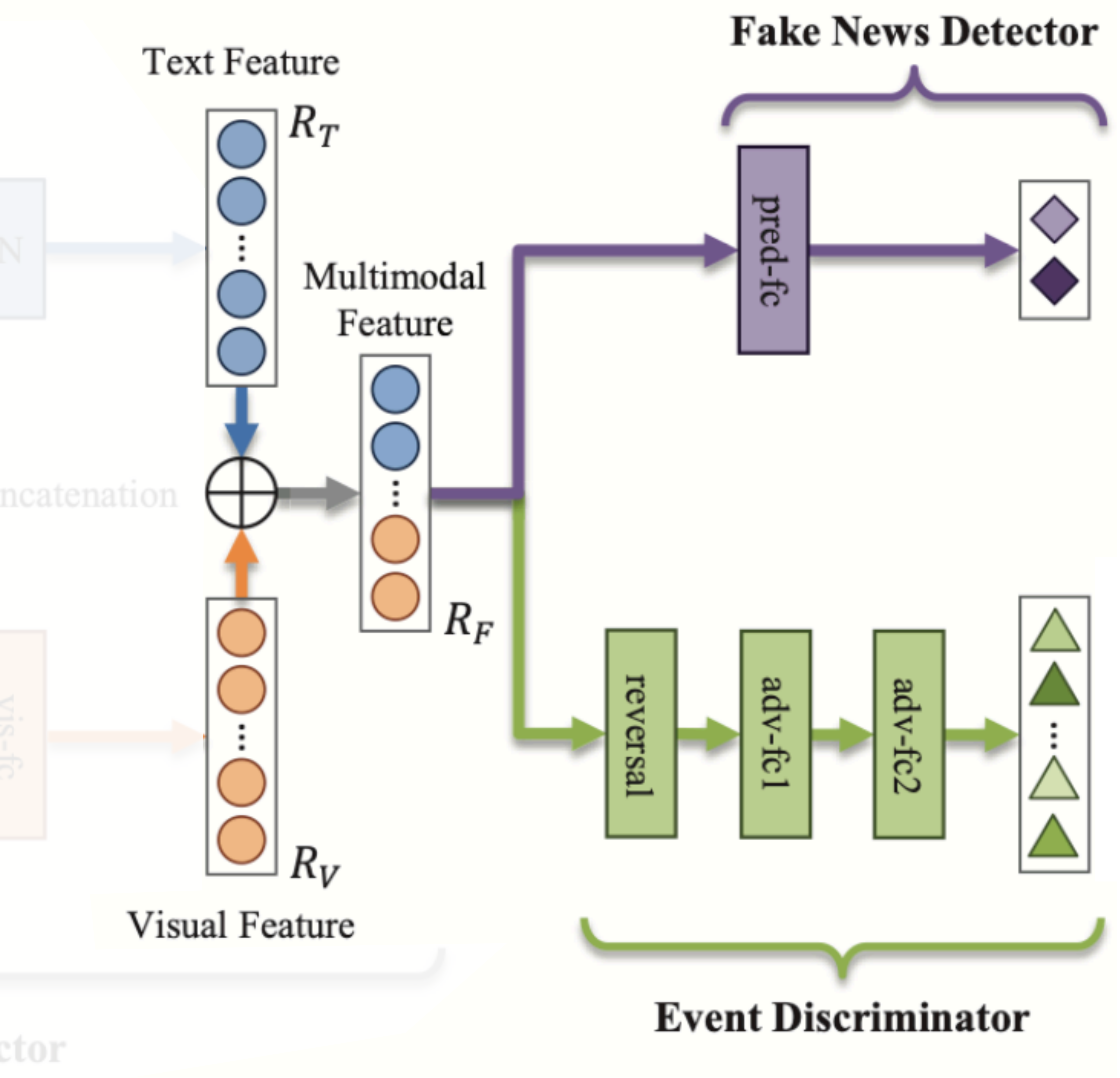
Methodology……….

Model Integration



- parameter set we seek is the saddle point of the final objective function, use SGD to solve problem:
  - $(\hat{\theta}_f, \hat{\theta}_d) = \underset{\theta_f, \theta_d}{arg\ min}\ L_{final}(\theta_f, \theta_d, \hat{\theta}_e)$
  - $\hat{\theta}_e = \underset{\theta_e}{arg\ max}\ L_{final}(\hat{\theta}_f, \hat{\theta}_d, \theta_e)$
- Here adopt the gradient reversal layer (GRL)
  - Acts as an identity function during forward stage, and it multiplies gradient with $-\lambda$ and passes the results to the preceding layer during back-prop stage.
  - GRL easily added between $G_f(\cdot\,; \theta_f)$ and $G_e(\cdot\,; \hat{\theta}_e)$

Methodology…………

Gradient Reversal Layer

- $L_{final}(\theta_f, \theta_d, \theta_e) = L_d(\theta_f, \theta_d) - \lambda L_e(\theta_f, \theta_e)$
- $(\hat{\theta}_f, \hat{\theta}_d) = \underset{\theta_f, \theta_d}{arg\ min}\ L_{final}(\theta_f, \theta_d, \hat{\theta}_e)$
- $\hat{\theta}_e = \underset{\theta_e}{arg\ max}\ L_{final}(\hat{\theta}_f, \hat{\theta}_d, \theta_e)$

Input: The multi-modal input $\{m_i\}_{i=1}^N$, the auxiliary event label $\{e_i\}_{i=1}^N$, the detection label $\{y_i\}_{i=1}^N$ and the learning rate η

1. **for** number of training iterations **do**
2. Decay learning rate: $\eta' = \dfrac{\eta}{(1 + \alpha \cdot p)^\beta}$
3. Update $\theta_f \leftarrow \theta_f - \eta(\dfrac{\partial L_d}{\partial \theta_f} - \lambda \dfrac{\partial L_e}{\partial \theta_f})$
4. Update $\theta_e \leftarrow \theta_e - \eta \dfrac{\partial L_e}{\partial \theta_e}$
5. Update $\theta_d \leftarrow \theta_d - \eta \dfrac{\partial L_d}{\partial \theta_d}$
6. **end for**