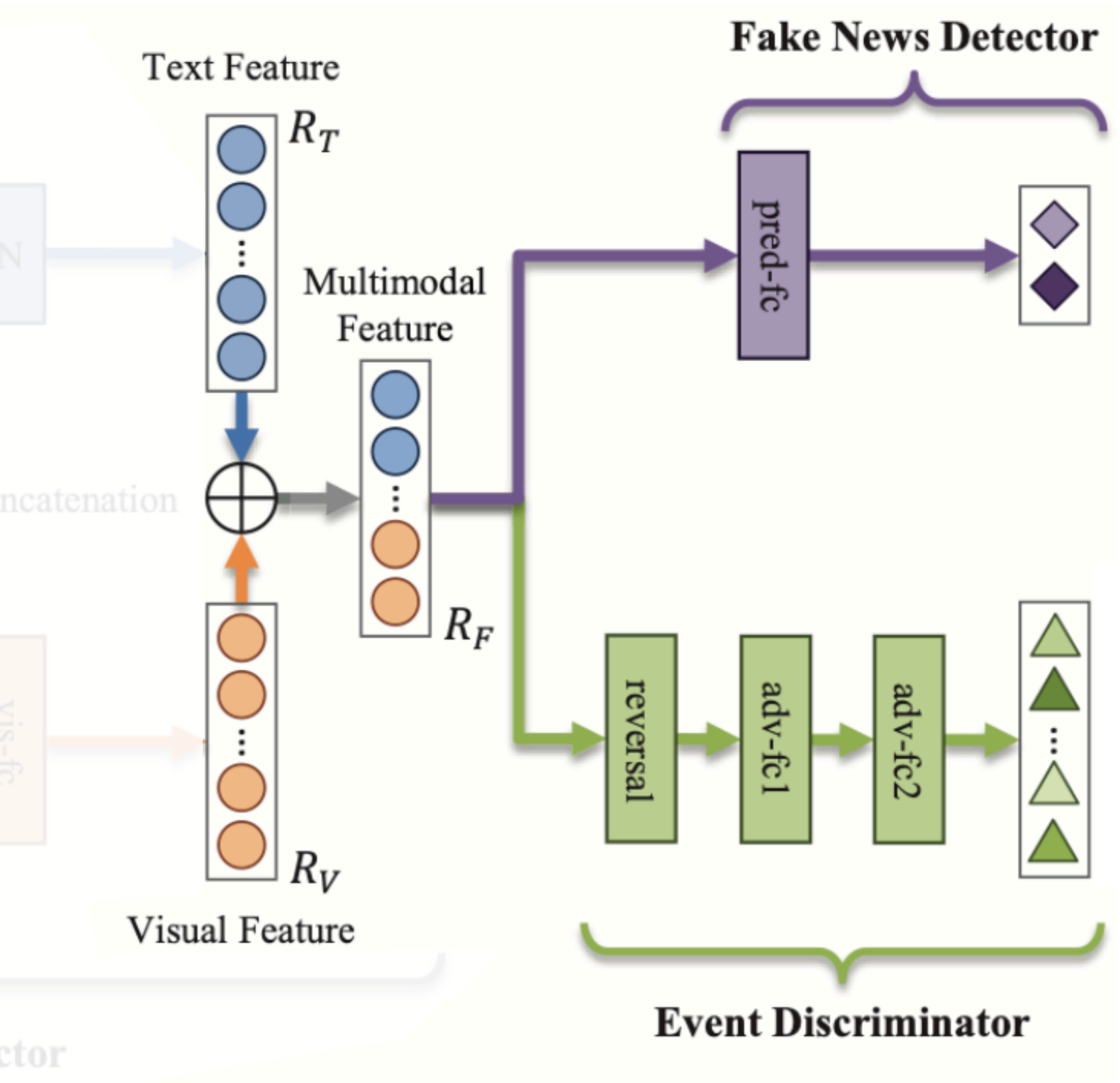
# Methodology.........

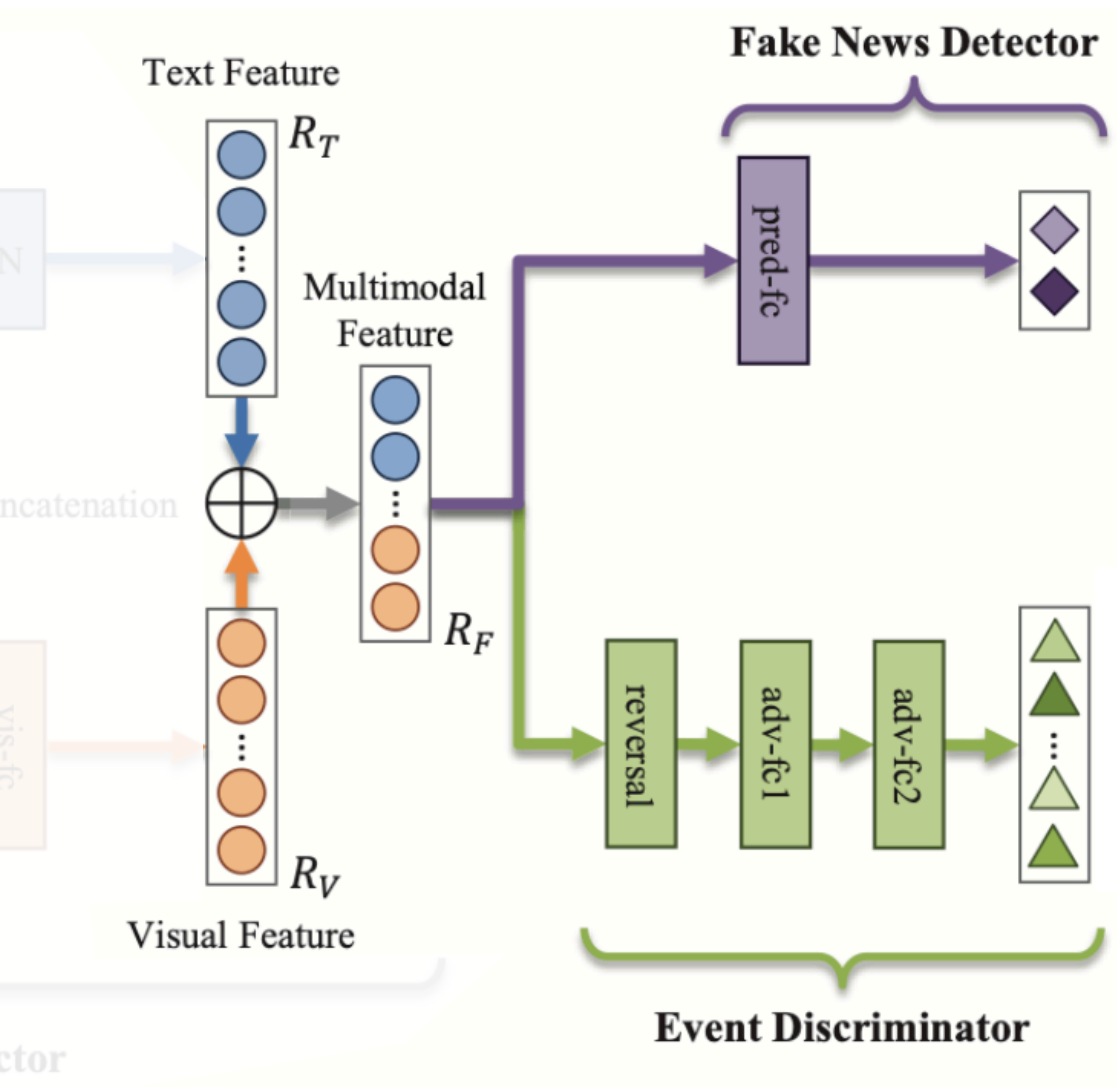
## Model Integration




- $G_f(\cdot\,;\theta_f)$ 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.

# 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 $G_e(\cdot\,; \hat{\theta}_e)$