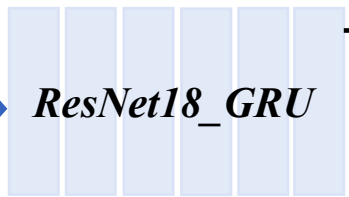
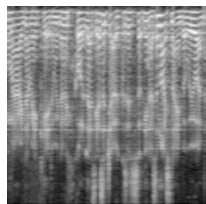
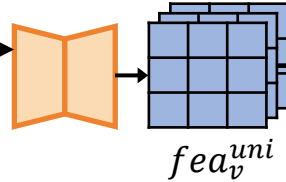
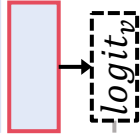




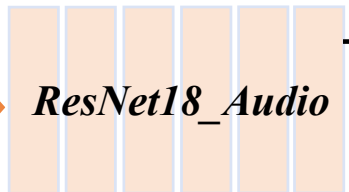
Video inputs



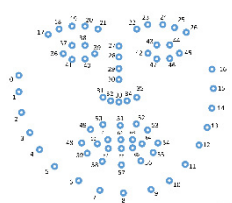
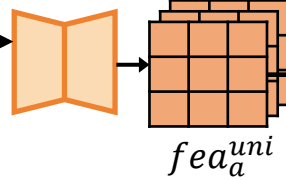
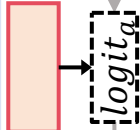
fea_v



Audio inputs



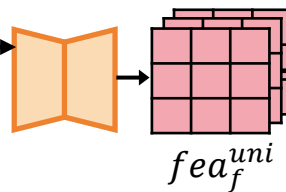
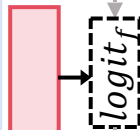
fea_a



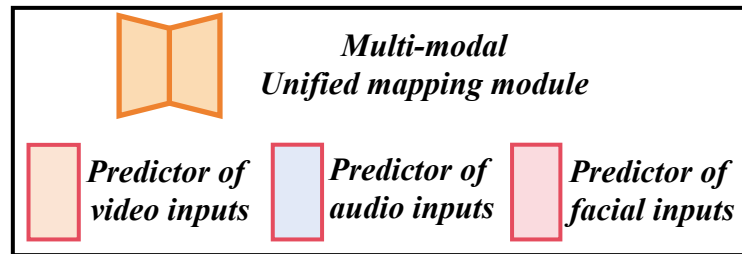
Facial inputs



fea_f



Logits of single modal



$$\mathcal{L}_{dm} = - \sum_n (Y \log(\text{logit}_n) + (1 - Y)(1 - \text{logit}_n))$$

Truth or Lies

