structure diagram

應名宥

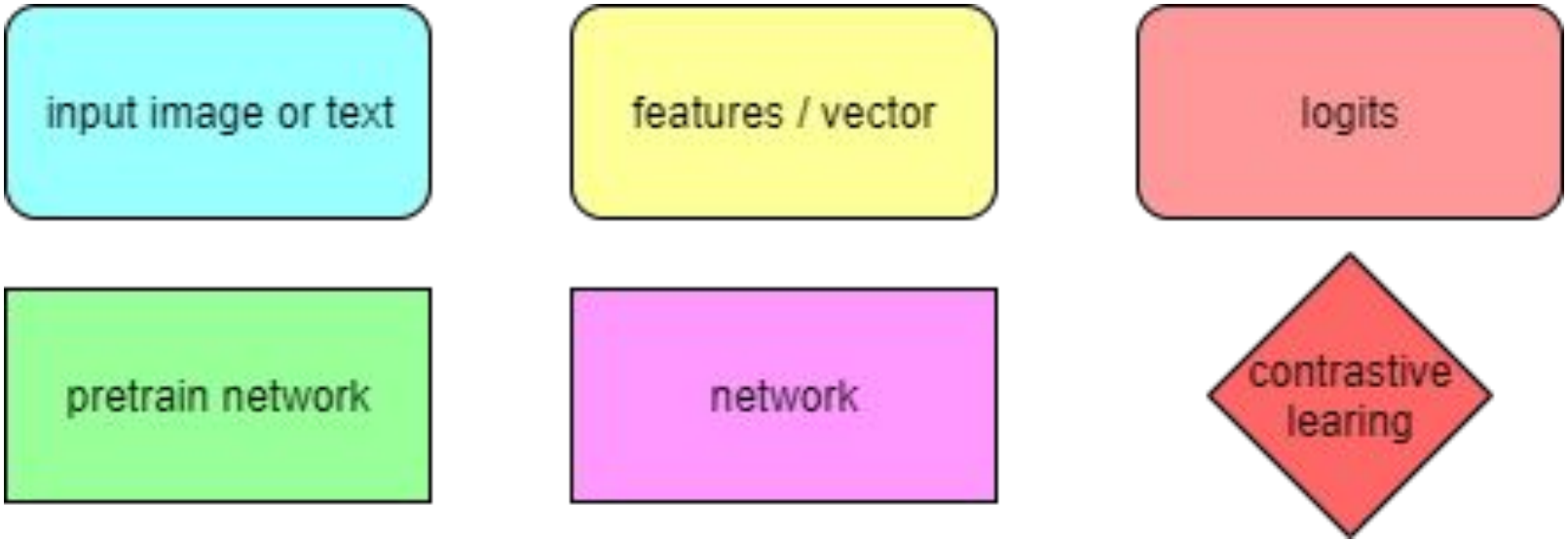
new stuff

- enable img-img loss
- separate img and txt features (discriminator)
- different generator layer (transform w)
- new contrastive learning structure
- mixing discriminator logits
- resnet guided discriminator

loss structure

- enable img-img loss
- new contrastive learning structure
- mixing discriminator logits
- resnet guided discriminator

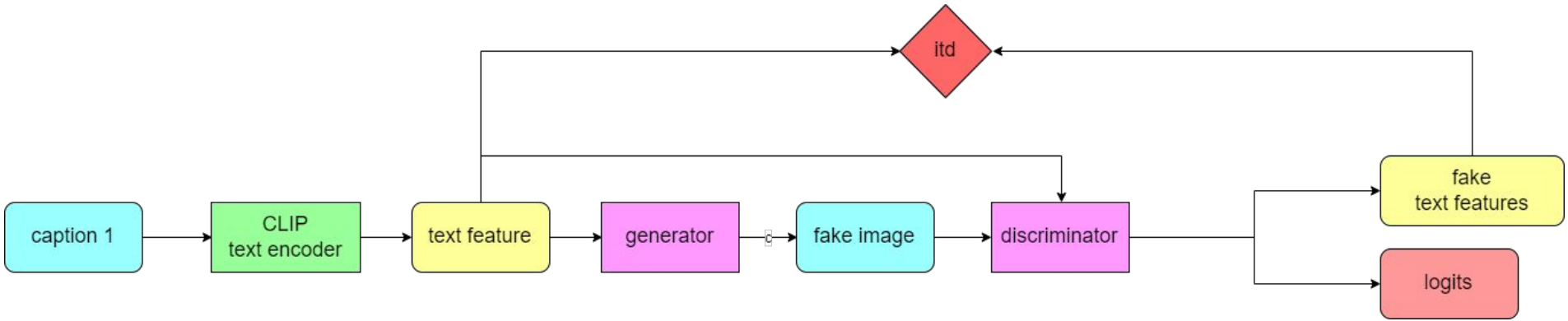
block type


input image or text
features / vector
logits
pretrain network
network
contrastive learing

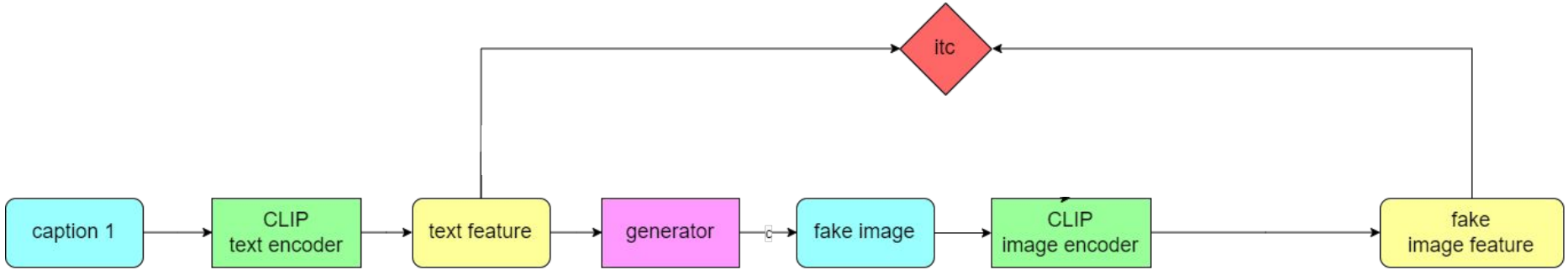
generator loss structure

origin

generator loss from discriminator (origin)



generator loss from clip (origin)



generator objective function (origin)

$$Sim(u, v) = exp(cos(u, v)/\tau)$$

$$L_G = -\sum_{i=1}^{n} log(\sigma(D(x_i, h_i)))$$

$$itd = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{img}(x_i), h_i))}{\sum_{j=1}^{n} exp(Sim(f_{img}(x_j), h_i))})$$

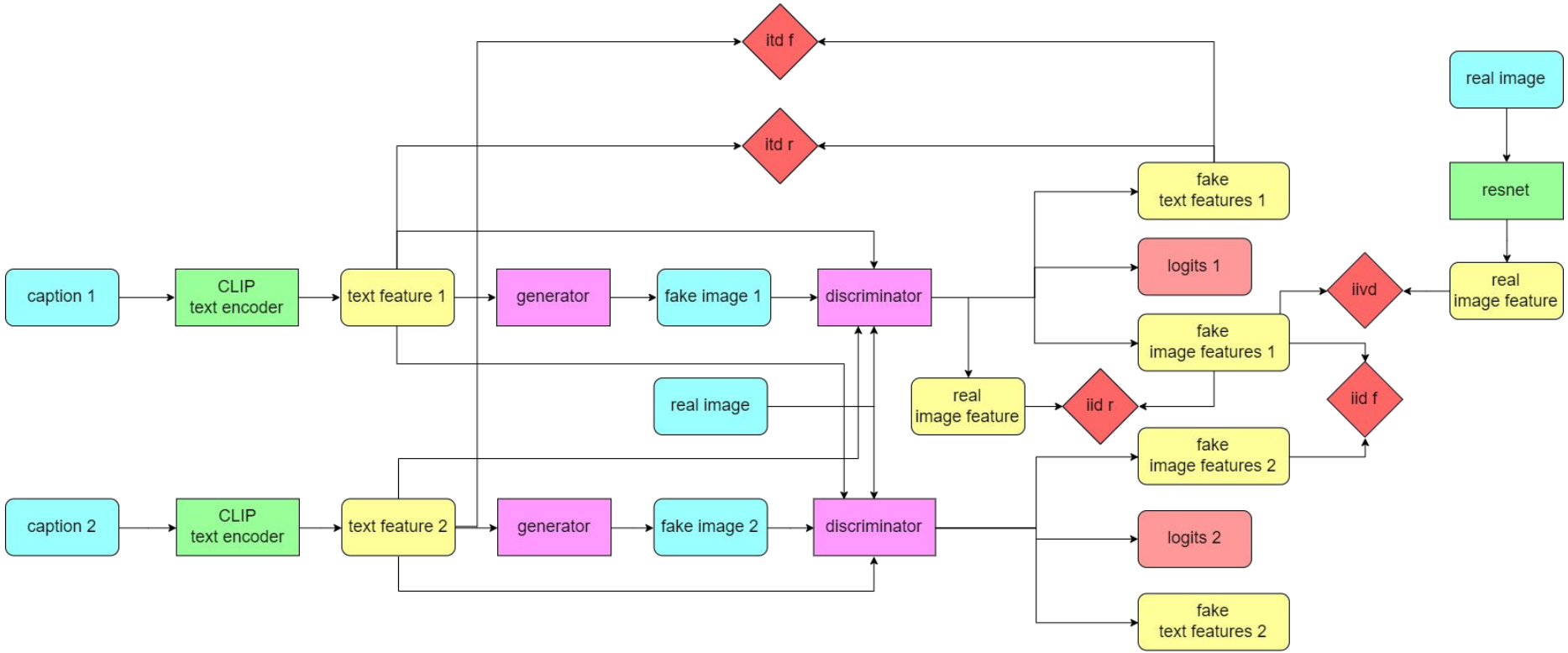
$$itc = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_c(x_i), h_i))}{\sum_{j=1}^{n} exp(Sim(f_c(x_i), h_i))})$$

$$L'_G = L_G + 5 \cdot itd + 10 \cdot itc$$

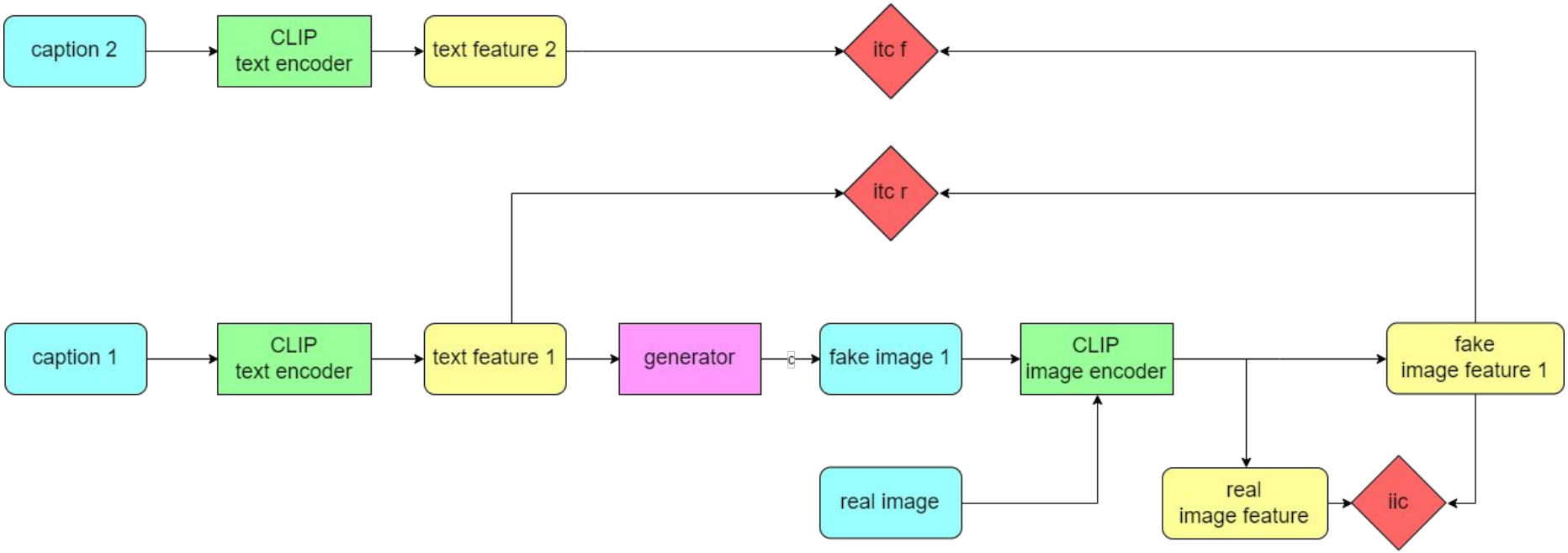
generator loss structure

modified

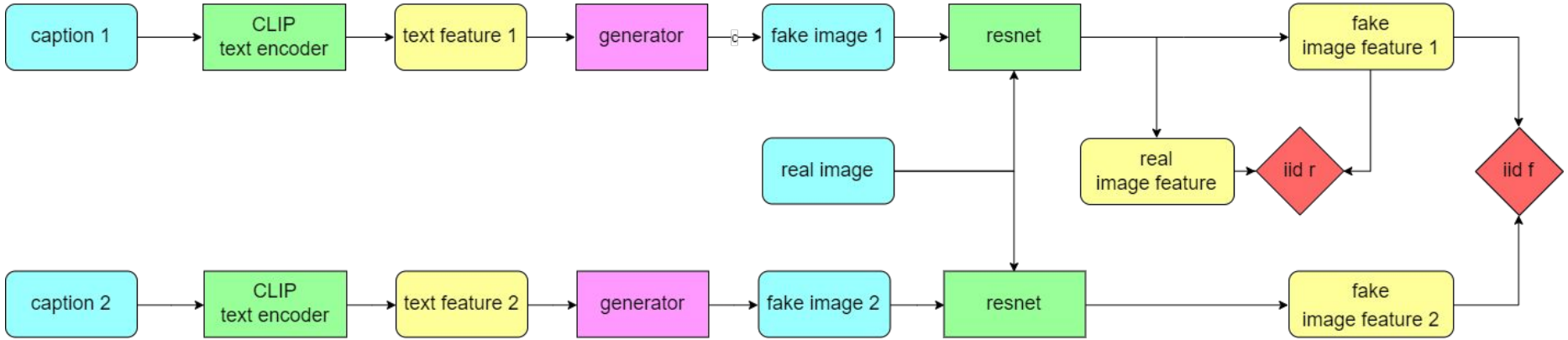
generator loss from discriminator (modified)



generator loss from clip (modified)



generator loss from pretrain resnet



generator objective function (modified)

$$Sim(u, v) = exp(cos(u, v)/\tau)$$

$$L_G = -\sum_{i=1}^{n} log(\sigma(D(x_i, h_i, h_i'))) + log(\sigma(D(x_i', h_i, h_i')))$$

$$itd_r = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{txt}(x_i), h_i))}{\sum_{j=1}^{n} exp(Sim(f_{txt}(x_j), h_i))})$$

$$itd_f = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{txt}(x_i), h_i'))}{\sum_{j=1}^{n} exp(Sim(f_{txt}(x_j), h_i'))})$$

$$iid_r = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{img}(x_i), f_{img}(r_i)))}{\sum_{j=1}^{n} exp(Sim(f_{img}(x_j), f_{img}(r_i)))})$$

$$iid_f = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{img}(x_i), f_{img}(x_i')))}{\sum_{j=1}^{n} exp(Sim(f_{img}(x_j), f_{img}(x_i')))})$$

cont.

$$iiv_r = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{res}(x_i), f_{res}(r_i)))}{\sum_{j=1}^{n} exp(Sim(f_{res}(x_j), f_{res}(r_i)))})$$

$$iiv_f = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{res}(x_i), f_{res}(x'_i)))}{\sum_{j=1}^{n} exp(Sim(f_{res}(x_j), f_{res}(x'_i)))})$$

$$iivd = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{img}(x_i), f_{res}(r_i)))}{\sum_{j=1}^{n} exp(Sim(f_{img}(x_j), f_{res}(r_i)))})$$

$$itc_r = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_c(x_i), h_i))}{\sum_{j=1}^{n} exp(Sim(f_c(x_i), h_i))})$$

$$itc_f = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_c(x_i), h'_i))}{\sum_{j=1}^{n} exp(Sim(f_c(x_i), h'_i))})$$

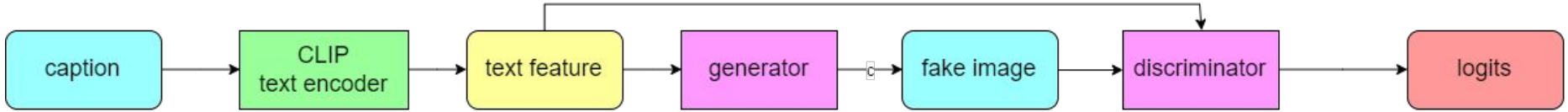
$$iic = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_c(x_i), f_c(x'_i)))}{\sum_{j=1}^{n} exp(Sim(f_c(x_i), f_c(x'_i)))})$$

$$L'_G = L_G + 5 \cdot itd_r + itd_f + 4 \cdot iid_r + 0.8 \cdot iid_f + 4 \cdot iiv_r + 0.8 \cdot iiv_f + 4 \cdot iivd + 10 \cdot itc_r + 2 \cdot itc_f + 3 \cdot iic$$

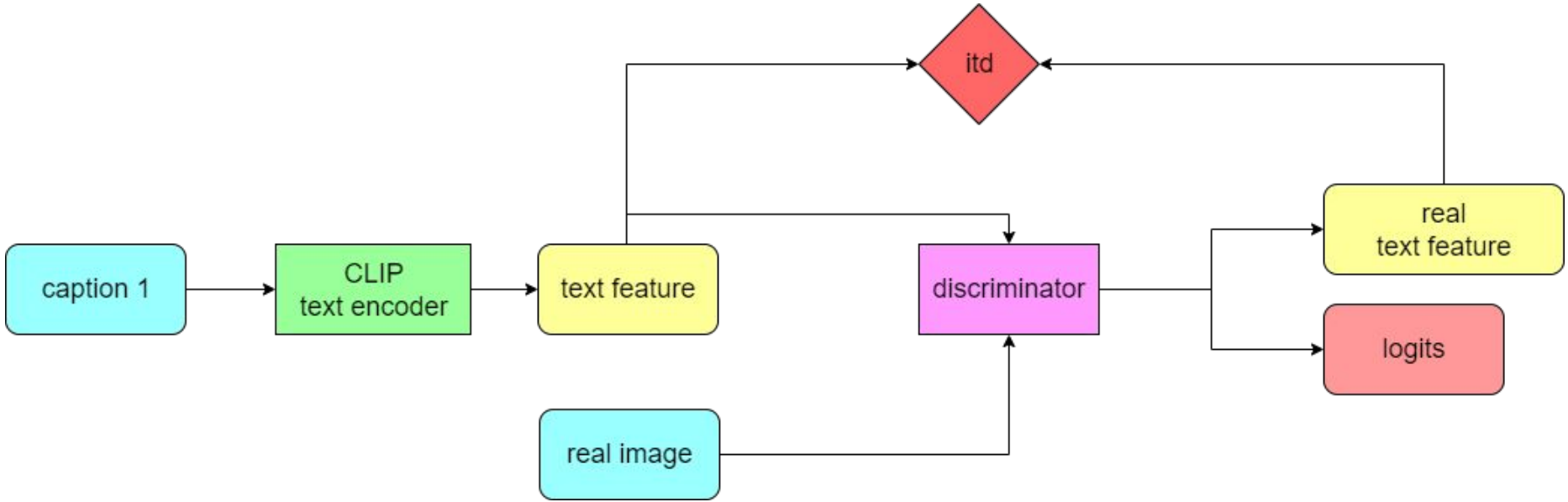
discriminator loss structure

origin

discriminator loss (fake) (origin)



discriminator loss (real) (origin)



discriminator objective function (original)

$$Sim(u, v) = exp(cos(u, v)/\tau)$$

$$L_D = -\sum_{i=1}^{n} log(\sigma(D(r_i, h_i))) - \sum_{i=1}^{n} log(1 - \sigma(D(x_i, h_i)))$$

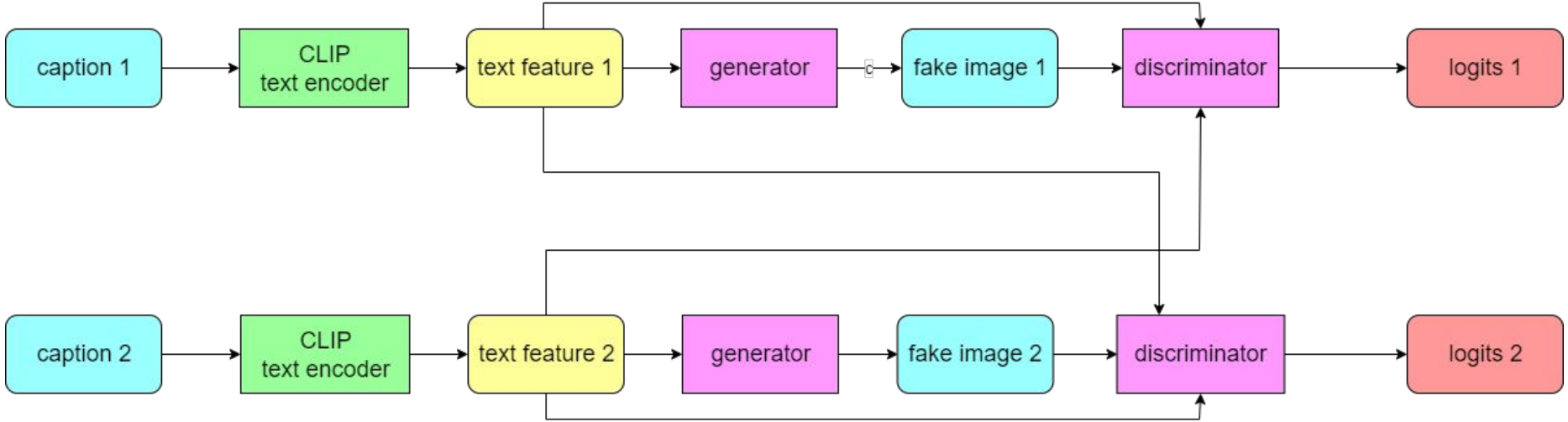
$$itd = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{txt}(x_i), h_i))}{\sum_{j=1}^{n} exp(Sim(f_{txt}(x_j), h_i))})$$

$$L'_D = L_D + 5 \cdot itd$$

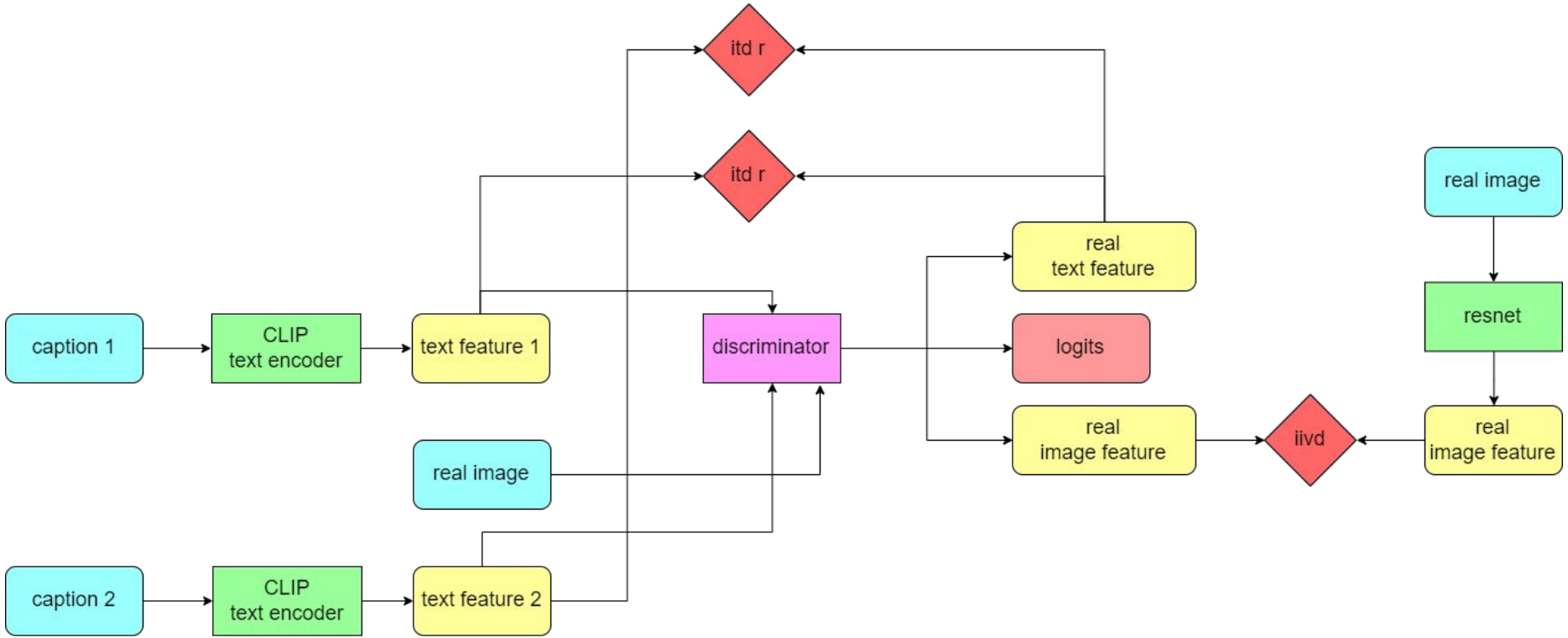
discriminator loss structure

modified

discriminator loss (fake) (modified)



discriminator loss (real) (modified)



discriminator objective function (modified)

$$Sim(u, v) = exp(cos(u, v)/\tau)$$

$$L_D = -\sum_{i=1}^{n} log(\sigma(D(r_i, h_i, h_i'))) - \sum_{i=1}^{n} log(1 - \sigma(D(x_i, h_i, h_i'))) - \sum_{i=1}^{n} log(1 - \sigma(D(x_i', h_i, h_i')))$$

$$itd = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{txt}(x_i), h_i))}{\sum_{j=1}^{n} exp(Sim(f_{txt}(x_j), h_i))})$$

$$iivd = -\tau \sum_{i=1}^{n} log(\frac{exp(Sim(f_{img}(r_i), f_{res}(r_i)))}{\sum_{j=1}^{n} exp(Sim(f_{img}(r_i), f_{res}(r_i)))})$$

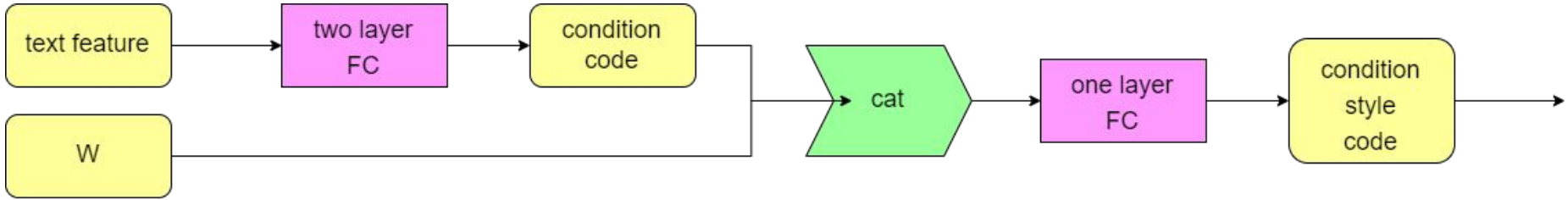
$$L_D' = L_D + 5 \cdot itd + 4 \cdot iivd$$

generator layer structure

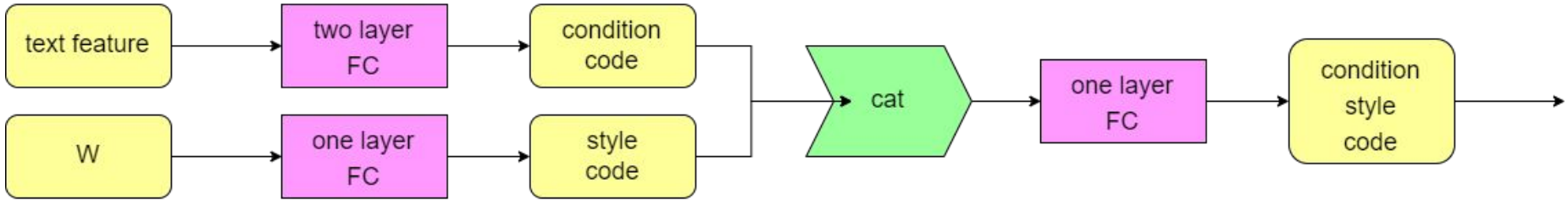
generator structure

- different generator layer (transform w)

different generator layer structure



origin generator layer



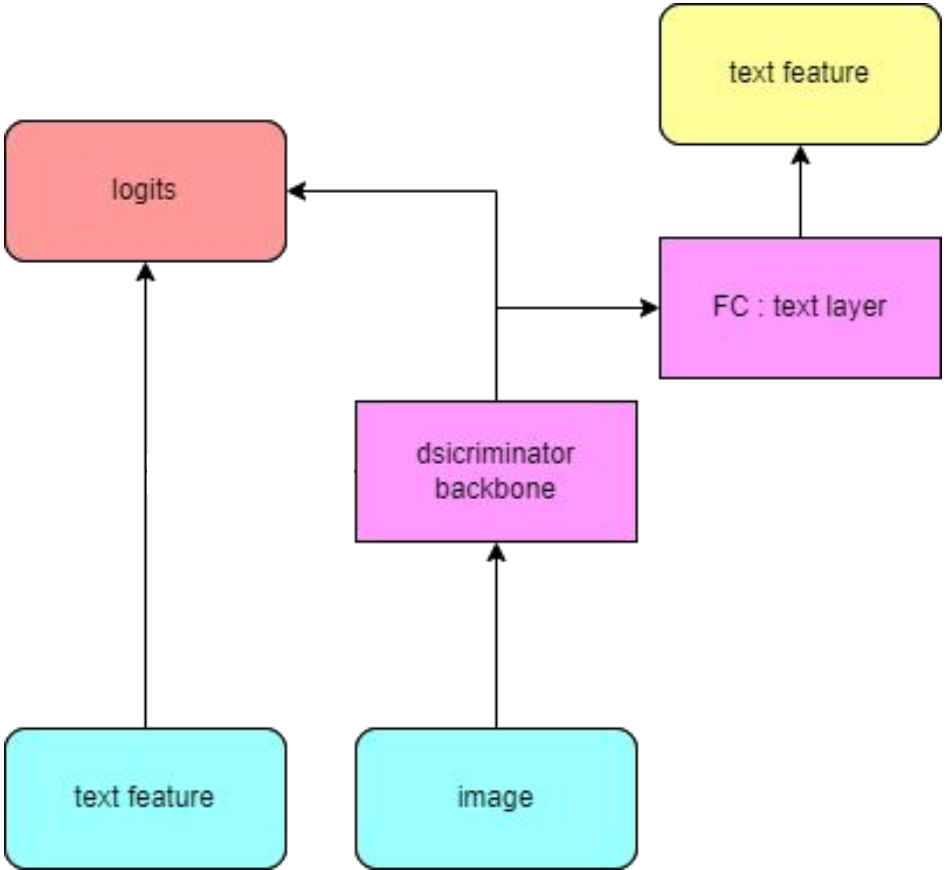
modified generator layer

discriminator feature extract structure

discriminator structure

- separate img and txt features (discriminator)

discriminator structure (origin)



discriminator structure (modified)