8/11

* Testing multi conditional guidance free image generation for pre-trained celeba model
* Testing clip + parse
  + Found that parse’s norm and grad norm is a lot bigger in terms of magnitude than clip
    - Thus, increase the weighing factor for clip, Nc (Set Nc to be 1000), so  
      Nc : Np = 1000:1
  + DEGRADED: Found that parse’s norm and grad norm continues to overpower clip’s
    - Have to add its contribution (grad norm) only at later stages, when time step > 200
      * So clip can guide the image generation at earlier stages
  + Found that if clip description is similar to the parse image (in terms of description), there will be better results. Parse gives directional/positional information while clip gives detailed facial descriptions
* Testing clip + faceid (arcface)
  + Nc:Na = 1:1 or 1:10 works well
  + Same, if clip description is similar to the parse image, there will be better results
* Testing clip + landmark
  + Nc:Nl = 1:1 or 1:10 works well
  + Same, if clip description is similar to the parse image, there will be better results
* Testing clip + sketch
  + Nc:Ns = 10:1 works well
  + Same, if clip description is similar to the parse image, there will be better results
* Final weights, A white background with black symbols

  Description automatically generated
  + Nc = 1, Np = 1/1000, Na = 1, Nl = 1, Ns = 1/10
* TODO: improve the weighted norm formula, with covariance information