# Application

Each dataset (Water Bird, CelebA, Color Minist, Jigsaw ):

1. Waterbird unbalanced (x); Waterbird almost balanced
2. Same class 1: class 2 ratio in minority and majority; Different class 1: class 2 ratio in minority and majority

**Water Bird**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Training: Minority: Majority Sample | Training: Class distribution within minority/majority | Test |
| Waterbird-Balanced-Same class ratio | 378:378 | - spurious=0: P(y=0):P(y=1) = 1:1(189:189)  - spurious=1: P(y=0):P(y=1) = 1:1(189:189)  - spurious=0:spurious=1 = 1:1(378:378) | spurious=0: P(y=0):P(y=1) = 2255:642  spurious=1: P(y=0):P(y=1) = 2255:642  (total=5794) |
| Waterbird-Balanced-Different class ratio | 756:756 | - spurious=0: P(y=0):P(y=1) = 1:3(189:567)  - spurious=1: P(y=0):P(y=1) = 3:1(567:189)  - spurious=0:spurious=1 = 1:1(756:756) |
| Waterbird-UnBalanced-Same class ratio | 378:2268 | - spurious=0: P(y=0):P(y=1) = 1:1(1134:1134)  - spurious=1: P(y=0):P(y=1) = 1:1(189:189)  - spurious=0:spurious=1 = 6:1(2268:378) |
| Waterbird-UnBalanced-Different class ratio | 264:1584 | - spurious=0: P(y=0):P(y=1) = 1:3(396:1188)  - spurious=1: P(y=0):P(y=1) = 3:1(198:66)  - spurious=0:spurious=1 = 6:1(1584:264) |

CelebA

|  |  |  |  |
| --- | --- | --- | --- |
|  | Training: Minority: Majority Sample | Training: Class distribution within minority/majority | Test |
| CelebA | 81733:100904 | - spurious=0: P(y=0):P(y=1) = 75150:25754(3:1)  - spurious=1: P(y=0):P(y=1) = 80164:1569(51:1)  - spurious=0:spurious=1 = 81733:100904(1:1.23) | spurious=0: P(y=0):P(y=1) = 7535:2480  spurious=1: P(y=0):P(y=1) = 9767:180  (total=19962) |
| CelebA-Balanced-Same class ratio | 3138:3138 | - spurious=0: P(y=0):P(y=1) = 1:1(1569:1569)  - spurious=1: P(y=0):P(y=1) = 1:1(1569:1569)  - spurious=0:spurious=1 = 1:1(3138:3138) |
| CelebA-Balanced-Different class ratio | 6276:6276 | - spurious=0: P(y=0):P(y=1) = 1:3(1569:4707)  - spurious=1: P(y=0):P(y=1) = 3:1(4707:1569)  - spurious=0:spurious=1 = 1:1(6276:6276) |
| CelebA-UnBalanced-Same class ratio | 18828:3138 | - spurious=0: P(y=0):P(y=1) = 1:1(9414:9414)  - spurious=1: P(y=0):P(y=1) = 1:1(1569:1569)  - spurious=0:spurious=1 = 6:1(18828:3138) |
| CelebA-UnBalanced-Different class ratio | 37656:6276 | - spurious=0: P(y=0):P(y=1) = 1:3(9414:28242)  - spurious=1: P(y=0):P(y=1) = 3:1(4707:1569)  - spurious=0:spurious=1 = 6:1(37656:6276) |

Needed results and comparisons

|  |  |  |  |
| --- | --- | --- | --- |
|  | S1= S^c\_{B,y} [complement of S2]  S2 = S\_{B, y} | S1=E^c\_{JTT}  S2=E\_{JTT} | S1=Union^c(S\_{B,y},  E\_{JTT})  S2 = Union(S\_{B,y},  E\_{JTT}) |
| WaterBird (%minority)  Test (training-provide if you believe insightful): (average accuracy using ERM on all samples, accuracy in majority/minority using ERM, ) | %S2  % of Minority in S2 |  |  |
| WaterBird-MixUp (Tuning rule) | (average accuracy on all, accuracy in majority/minority) |  |  |
| WaterBird-DRO |  |  |  |
| WaterBird-JTT |  |  |  |
| WaterBird-GIC |  |  |  |
| DRO (David) |  |  |  |
| JTT (Yiran) |  |  |  |
| GIC (with label) (Yiran) |  |  |  |

For WaterBird

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S1= S^c\_{B,y}  [complement of S2]  S2 = S\_{B, y}  Best Model (Non-Overtrained) | | | | S1= S^c\_{B,y} [complement of S2]  S2 = S\_{B, y}  300 epochs (Overtrained Model) | | | | S1=E^c\_{JTT}  S2=E\_{JTT}  (only training data) | | | | S1=Union^c(S\_{B,y},  E\_{JTT})  S2 = Union(S\_{B,y},  E\_{JTT}) | | | |
|  | Waterbird-Balanced-Same class ratio | Waterbird-Balanced-Different class ratio | Waterbird-UnBalanced-Same class ratio | Waterbird-UnBalanced-Different class ratio | Waterbird-Balanced-Same class ratio | Waterbird-Balanced-Different class ratio | Waterbird-UnBalanced-Same class ratio | Waterbird-UnBalanced-Different class ratio | Waterbird-Balanced-Same class ratio | Waterbird-Balanced-Different class ratio | Waterbird-UnBalanced-Same class ratio | Waterbird-UnBalanced-Different class ratio | Waterbird-Balanced-Same class ratio | Waterbird-Balanced-Different class ratio | Waterbird-UnBalanced-Same class ratio | Waterbird-UnBalanced-Different class ratio |
|  | 216 | 416 | 290 | 149 | 229 | 394 | 746 | 158 | 162 | 315 | 265 | 235 | 346 | 667 | 505 | 369 |
|  | 81.48%  (176/216) | 80.05%(333/416) | 41.38%(120/290) | 22.15%(33/149) | 82.10%(188/229) | 75.63%(298/394) | 4.69%(35/746) | 19.62%(31/158) | 83.95%(136/162) | 88.89%(280/315) | 80.38%(213/265) | 77.87%(183/235) | 82.08%(284/346) | 83.2%(555/667) | 59.40(300/505) | 56.91%(210/369) |
| WaterBird (%minority) | 50%  (378/756) | 50%  (756/1512) | 14.29%  (378/2646) | 14.29%  (264/1848) | 50%  (378/756) | 50%  (756/1512) | 14.29%  (378/2646) | 14.29%  (264/1848) | 50%  (378/756) | 50%  (756/1512) | 14.29%  (378/2646) | 14.29%  (264/1848) | 50%  (378/756) | 50%  (756/1512) | 14.29%  (378/2646) | 14.29%  (264/1848) |
| ERM avg train val acc on all samples | Train:  99.1%  Val:  78.2%  (best model is saved at the 273th epoch) | Train:  97.5%  Val:  82.8%  (best model is saved at the 259th epoch) | Train:  98.4%  Val:  75.2%  (best model is saved at the 195th epoch) | Train:  97.5%  Val:  70.8%  (best model is saved at the 203th epoch) | Train:  1.0  Val:  77.3%  (300 epoch) | Train:  98.77%  Val:  81.05%  (300 epoch) | Train:  87.32%  Val:  49.9%  (300 epoch) | Train:  98.95%  Val:  65.25%  (300 epoch) |  |  |  |  |  |  |  |  |
| ERM acc on all samples | 86.85% | 87.76% | 75.33% | 69.5% | 86.47% | 86.71% | 51.19% | 64.8% |  |  |  |  |  |  |  |  |
| ERM acc in majority | 94.27% | 93.86% | 98.24% | 96.76% | 94.93% | 93.2% | 90.92% | 94.93% |  |  |  |  |  |  |  |  |
| ERM acc in minority | 75.42% | 81.67% | 52.43% | 42.25% | 78.01% | 80.22% | 11.46% | 34.69% |  |  |  |  |  |  |  |  |

JTT results of waterbird dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Waterbird-Balanced-Same class ratio | | | Waterbird-Balanced-Different class ratio | | | Waterbird-UnBalanced-Same class ratio | | | Waterbird-UnBalanced-Different class ratio | | |
|  | Upweighting set = Error set | Upweighting set = S\_{B, y} | Upweighting set = Union(S\_{B,y}, E\_{JTT}) | Upweighting set = Error set | Upweighting set = S\_{B, y} | Upweighting set = Union(S\_{B,y}, E\_{JTT}) | Upweighting set = Error set | Upweighting set = S\_{B, y} | Upweighting set = Union(S\_{B,y}, E\_{JTT}) | Upweighting set = Error set | Upweighting set = S\_{B, y} | Upweighting set = Union(S\_{B,y}, E\_{JTT}) |
| Training Accuracy | 99.88%  (best model:227) | 97.89%(best model:63) | 98.83%(best model:44) | 97.81%(best model:60) | 98.29%(best model:50) | 99.06%(best model:51) | 99.71%(best model:96) | 99.01%(best model:54) | 97.08%(best model:20) | 94.29%(best model:25) | 98.28%(best model:62) | 98.70%(best model:35) |
| Validation Accuracymajority/ | 83.4% | 88.34% | 90.49% | 89.81% | 91.08% | 91.24% | 85.60% | 81.57% | 86.56% | 88.06% | 72.68% | 86.74% |
| Test Accuracy | 89.14% | 89.35% | 87.4% | 90.58% | 93.04% | 92.85% | 88.76% | 84.05% | 91.00% | 89.92% | 72.56% | 89.16% |

For CelebA(threshold=0.4)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S1= S^c\_{B,y}  [complement of S2]  S2 = S\_{B, y}  Best Model (Non-Overtrained) | | | | S1= S^c\_{B,y} [complement of S2]  S2 = S\_{B, y}  300 epochs (Overtrained Model) | | | | S1=E^c\_{JTT}  S2=E\_{JTT}  (only training data) | | | | S1=Union^c(S\_{B,y},  E\_{JTT})  S2 = Union(S\_{B,y},  E\_{JTT}) | | | |
|  | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio |
|  | 500 | 940 | 4636 | 17718 |  |  |  |  | 2018 | 2804 | 2918 | 4570 | 2436 | 3670 | 7442 | 21968 |
|  | 32.8%  (164/500) | 10.64%(100/940) | 3.47%(161/4636) | 2.52%(446/17718  )check! |  |  |  |  | 63.13%(1274/2018) | 64.73%(1815/2804) | 61.45%(1793/2918) | 67.61%(3090/4570) | 57.63%(1404/2436) | 51.14%(1877/3670) | 25.58%(1904/7442) | 15.19(3336/21968) |
| CelebA (%minority) | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) |  |  |  |  | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) | 50%  (378/756) | 50%(6276/12552) | 14.29%(3138/21966) | 14.29%(6276/17718) |
| ERM avg train val acc on all samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ERM acc on all samples |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ERM acc in majority |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ERM acc in minority |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

For CelebA(threshold=0.6)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S1= S^c\_{B,y}  [complement of S2]  S2 = S\_{B, y}  Best Model (Non-Overtrained) | | | | S1= S^c\_{B,y} [complement of S2]  S2 = S\_{B, y}  300 epochs (Overtrained Model) | | | | S1=E^c\_{JTT}  S2=E\_{JTT}  (only training data) | | | | S1=Union^c(S\_{B,y},  E\_{JTT})  S2 = Union(S\_{B,y},  E\_{JTT}) | | | |
|  | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio |
|  | 339 | 526 | 2522 | 10523 |  |  |  |  | 2018 | 2804 | 2918 | 4570 | 2265 | 3198 | 5283 | 14654 |
|  | 40.7%  (138/339) | 28.9%(152/526) | 6.42%(162/2522) | 4.37%(460/10523) |  |  |  |  | 63.13%(1274/2018) | 64.73%(1815/2804) | 61.45%(1793/2918) | 67.61%(3090/4570) | 60.1%(1361/2265) | 58.3%(1865/3198) | 35.2%(1860/5283) | 22.15%(3246/14654) |
| CelebA (%minority) | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) |  |  |  |  | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) | 50%  (378/756) | 50%(6276/12552) | 14.29%(3138/21966) | 14.29%(6276/17718) |

GIC(Trained for 100 epochs)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Comparison data = S\_{b,y}  Best Model (Non-Overtrained) | | | | Comparison data = original validation set | | | |
|  | Waterbird-Balanced-Same class ratio | Waterbird-Balanced-Different class ratio | Waterbird-UnBalanced-Same class ratio | Waterbird-UnBalanced-Different class ratio | Waterbird-Balanced-Same class ratio | Waterbird-Balanced-Different class ratio | Waterbird-UnBalanced-Same class ratio | Waterbird-UnBalanced-Different class ratio |
| GIC training accuracy | 0.7963 | 0.6761 | 0.6087 | 0.6981 | 0.8326 | 0.6538 | 0.5002 | 0.7260 |
| GIC test accuracy | 0.8100 | 0.4280 | 0.3735 | 0.2216 | 0.6943 | 0.2561 | 0.2216 | 0.2216 |

## Goal: Oct 23 - Oct 30

1. David: learn to use mixup and if possible, apply it to one waterbird data.
2. Yiran: Generate different training dataset and fill-in the downstream model agnostic information.

Citation:

@article{

zhang2018mixup,

title={mixup: Beyond Empirical Risk Minimization},

author={Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz},

journal={International Conference on Learning Representations},

year={2018},

url={https://openreview.net/forum?id=r1Ddp1-Rb},

}

1. Only negation process

CelebA Dataset(threshold=0.4)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S1= S^c\_{B,y}  [complement of S2]  S2 = S\_{B, y}  Best Model (Non-Overtrained) | | | | S1= S^c\_{B,y} [complement of S2]  S2 = S\_{B, y}  300 epochs (Overtrained Model) | | | | S1=E^c\_{JTT}  S2=E\_{JTT}  (only training data) | | | | S1=Union^c(S\_{B,y},  E\_{JTT})  S2 = Union(S\_{B,y},  E\_{JTT}) | | | |
|  | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio |
|  | 3537 | 6837 | 13859 | 26680 |  |  |  |  | 2018 | 2804 | 2918 | 4570 | 4620 | 8201 | 5283 | 27867 |
|  | 47.75%  (1689/3537) | 69.31.9%(4739/6837) | 11.262%(1611/13859) | 17.368%(4716/26680) |  |  |  |  | 63.13%(1274/2018) | 64.73%(1815/2804) | 61.45%(1793/2918) | 67.61%(3090/4570) | 51.99%(2402/4620) | 65.43%(5366/8201) | 16.80%(2611/15543) | 19.154%(5445/27867) |
| CelebA (%minority) | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) |  |  |  |  | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) | 50%  (378/756) | 50%(6276/12552) | 14.29%(3138/21966) | 14.29%(6276/17718) |

1. Only Masking process

CelebA Dataset(threshold=0.4)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S1= S^c\_{B,y}  [complement of S2]  S2 = S\_{B, y}  Best Model (Non-Overtrained) | | | | S1= S^c\_{B,y} [complement of S2]  S2 = S\_{B, y}  300 epochs (Overtrained Model) | | | | S1=E^c\_{JTT}  S2=E\_{JTT}  (only training data) | | | | S1=Union^c(S\_{B,y},  E\_{JTT})  S2 = Union(S\_{B,y},  E\_{JTT}) | | | |
|  | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio |
|  | 3050 | 6184 | 10547 | 25008 |  |  |  |  | 2018 | 2804 | 2918 | 4570 | 3903 | 7497 | 11633 | 27953 |
|  | 52.2%  (1592/3050) | 27.3%(1689/6184) | 15.97%(1684/10547) | 8.07%(2018/25008) |  |  |  |  | 63.13%(1274/2018) | 64.73%(1815/2804) | 61.45%(1793/2918) | 67.61%(3090/4570) | 53.9%(2104/3903) | 37.01%(2775/7497) | 20.53%(2388/11633) | 14.58%(4076/27953) |
| CelebA (%minority) | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) |  |  |  |  | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) | 50%  (378/756) | 50%(6276/12552) | 14.29%(3138/21966) | 14.29%(6276/17718) |

1. Masking process + Negation process

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S1= S^c\_{B,y}  [complement of S2]  S2 = S\_{B, y}  Best Model (Non-Overtrained) | | | | S1= S^c\_{B,y} [complement of S2]  S2 = S\_{B, y}  300 epochs (Overtrained Model) | | | | S1=E^c\_{JTT}  S2=E\_{JTT}  (only training data) | | | | S1=Union^c(S\_{B,y},  E\_{JTT})  S2 = Union(S\_{B,y},  E\_{JTT}) | | | |
|  | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio | CelebA-Balanced-Same class ratio | CelebA-Balanced-Different class ratio | CelebA-UnBalanced-Same class ratio | CelebA-UnBalanced-Different class ratio |
|  | 500 | 940 | 4636 | 17718 |  |  |  |  | 2018 | 2804 | 2918 | 4570 | 2436 | 3670 | 7442 | 21968 |
|  | 32.8%  (164/500) | 10.64%(100/940) | 3.47%(161/4636) | 2.52%(446/17718  )check! |  |  |  |  | 63.13%(1274/2018) | 64.73%(1815/2804) | 61.45%(1793/2918) | 67.61%(3090/4570) | 57.63%(1404/2436) | 51.14%(1877/3670) | 25.58%(1904/7442) | 15.19(3336/21968) |
| CelebA (%minority) | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) |  |  |  |  | 50%  (3138/6276) | 50%(6276/12552) | 14.29%  (3138/21966) | 14.29%  (6276/17718) | 50%  (378/756) | 50%(6276/12552) | 14.29%(3138/21966) | 14.29%(6276/17718) |

Prediction Accuracy(Remove GRAD-CAM identified features)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | Group Label | Total Samples | Accuracy  (Using spurious feature identified by GradCAM) | Accuracy  (Using core feature identified by GradCAM) | Prediction Accuracy  (Original Image) |
| CelebA-Balanced-Same class ratio | 0 | 3138 | 0.7779 | 0.5354 | 0.9398 |
| 1 | 3138 | 0.7345 | 0.4927 | 0.9031 |
| CelebA-Balanced-Different class ratio | 0 | 6276 | 0.6179 | 0.2838 | 0.9602 |
| 1 | 6276 | 0.8270 | 0.7309 | 0.9116 |
| CelebA-UnBalanced-Same class ratio | 0 | 18828 | 0.8670 | 0.5293 | 0.9816 |
| 1 | 3138 | 0.6746 | 0.4634 | 0.8333 |
| CelebA-UnBalanced-Different class ratio | 0 | 35168 | 0.8532 | 0.3463 | 0.9855 |
| 1 | 6276 | 0.8649 | 0.6785 | 0.8647 |