

Obiettivi Progetto Neuron Selectivity

1 Integrare meta-meter

fatto

1.1 Performance meta-meter

| | Baseline | Interpretable |
|--------------------|-------------|---------------|
| RMSE | 0.7054 | 0.7284 |
| MAE | 0.4667 | 0.4893 |
| REL | 0.1871 | 0.1975 |
| sigma1 | 71.747% | 69.892% |
| sigma2 | 92.3296% | 91.3019% |
| sigma3 | 97.823% | 97.4149% |
| selectivity | 0.4182 | 0.7688 |
| target selectivity | -0.2268 | 0.7208 |
| accuracy on target | 0% | 60% |
| Inactive units | 0 out of 16 | 6 out of 16 |

Table 1.

2 Misurare inference time

fatto

ho misurato la media dei tempi dell'inference time nelle immagini del test set, 5 test per modello e poi ho fatto la media tra i risultati dei test

ho usato la gpu rtx 4050

interpretable 2.946 ms

baseline 2.94 ms

cpu: 13th Gen Intel(R) Core(TM) i7-13620H (2.40 GHz)

interpretable 22,972 ms

baseline 22,982 ms

3 Misurare robustezza modello

in corso

3.1 FGSM

| | | | | |
|------------|-------------------|--------|--------|--------|
| epsilon -> | 0 (original pred) | 0,02 | 0,01 | 0,007 |
| MAE | 0,4893 | 1,0120 | 0,8934 | 0,8209 |
| RMSE | 0,7285 | 1,315 | 1,1774 | 1,0946 |

Table 2. Interpretable model

| | | | | |
|------------|-------------------|--------|--------|--------|
| epsilon -> | 0 (original pred) | 0,02 | 0,01 | 0,007 |
| MAE | 0,4667 | 0,9723 | 0,8685 | 0,8036 |
| RMSE | 0,7054 | 1,2648 | 1,1465 | 1,0736 |

Table 3. Baseline model

3.2 PGD

testato con 7 iterazioni, $\alpha = 0.007$ (il peso del passo del pixel nella direzione del gradiente),
 $\epsilon = 0.03$ (il bound della perturbazione)

| | Interpretable | Baseline |
|------|---------------|----------|
| MAE | 1,6605 | 1,6619 |
| RMSE | 2,0573 | 2,0558 |

Table 4.