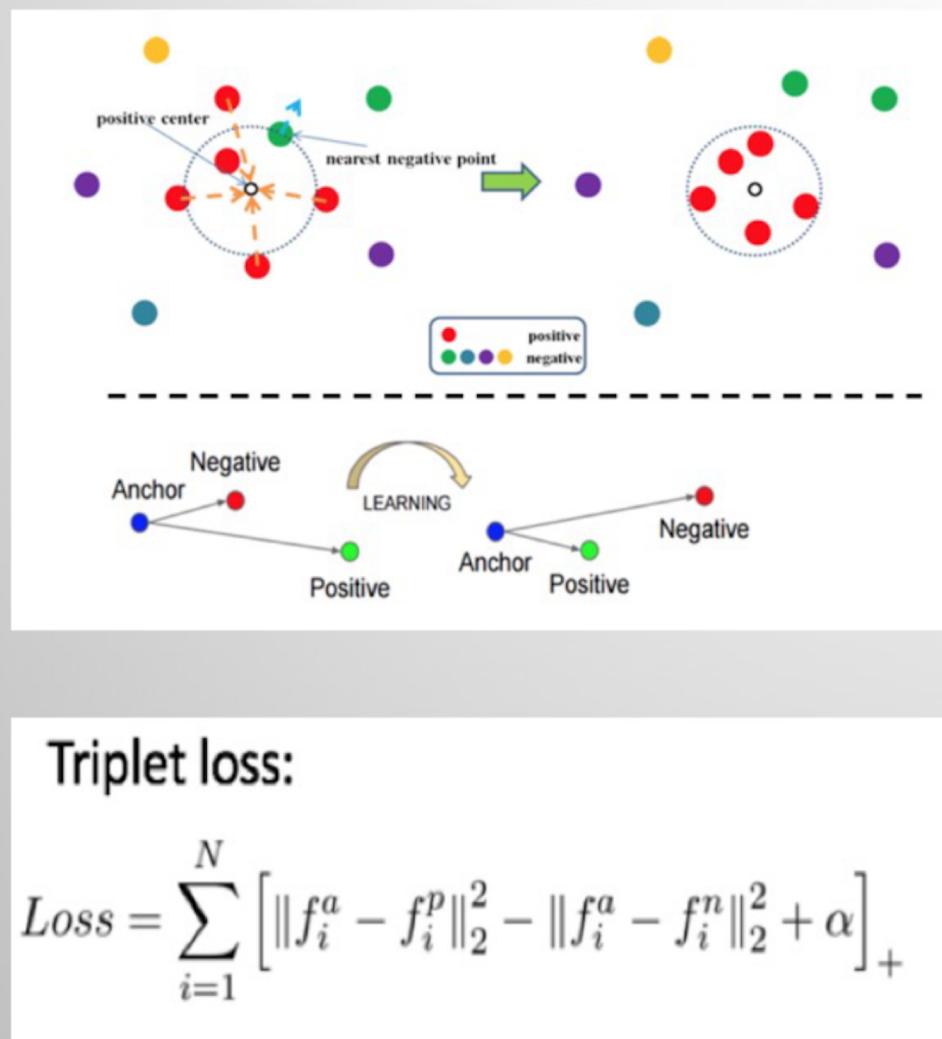


Similarity metrics for human tracking with triplet network

Thomas Liao, Jin Bai

Triplet encoding network of pedestrian feature similarity embedding for human re-identification with tri-hard triplet loss.



Triplet loss:

$$Loss = \sum_{i=1}^N \left[\|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha \right]_+$$

Batch hard triplet loss [1]

$$\mathcal{L}_{BH}(\theta; X) = \sum_{i=1}^P \sum_{a=1}^K \left[m + \underbrace{\max_{p=1 \dots K} D(f_\theta(x_a^i), f_\theta(x_p^i))}_{\text{hardest positive}} \right. \\ \left. - \underbrace{\min_{\substack{j=1 \dots P \\ n=1 \dots K \\ j \neq i}} D(f_\theta(x_a^i), f_\theta(x_n^j))}_{\text{hardest negative}} \right]_+, \quad (5)$$

Original Model

pretrained Resnet 50 (final layer removed) + 2fc(1024 – BN,relu - 128)

trained on Market 1501 dataset [2]

Fine-tuning on Diva Training Set

Training details:

- Pre-trained model: using previous model(trained on Market1501)
- Combined Diva Training data and market1501 data for training – total ~1.3k unique ids (market1501 train set ~750, diva ~550).
- Each batch: 18 unique ids each with 4 images (cropped into h-258 by w-128).
- Batch-hard loss.

Fine-tuning on Diva Training Dataset

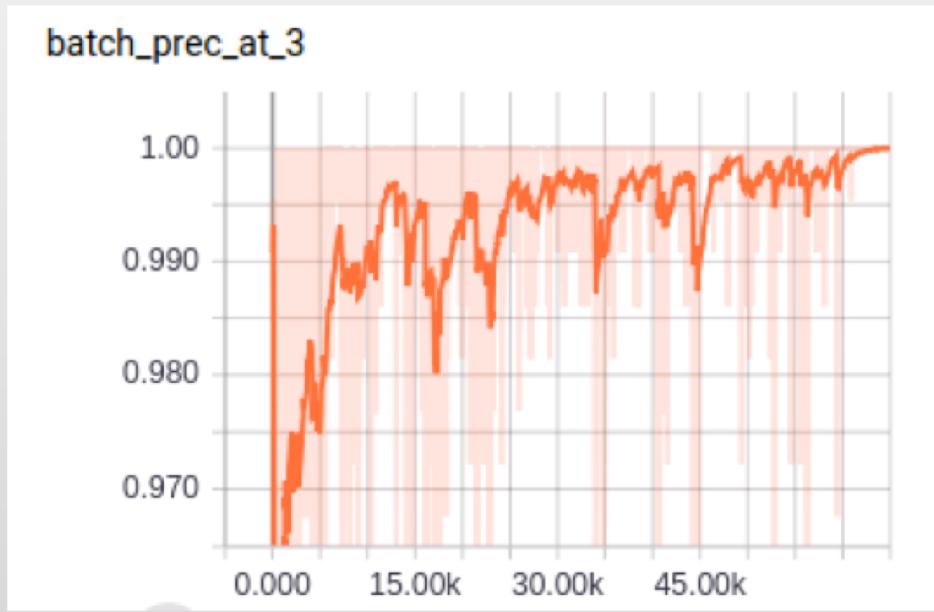


Fig 1. Accuracy on training set (recorded per iter, proportion of top-3 correct matching for each batch).

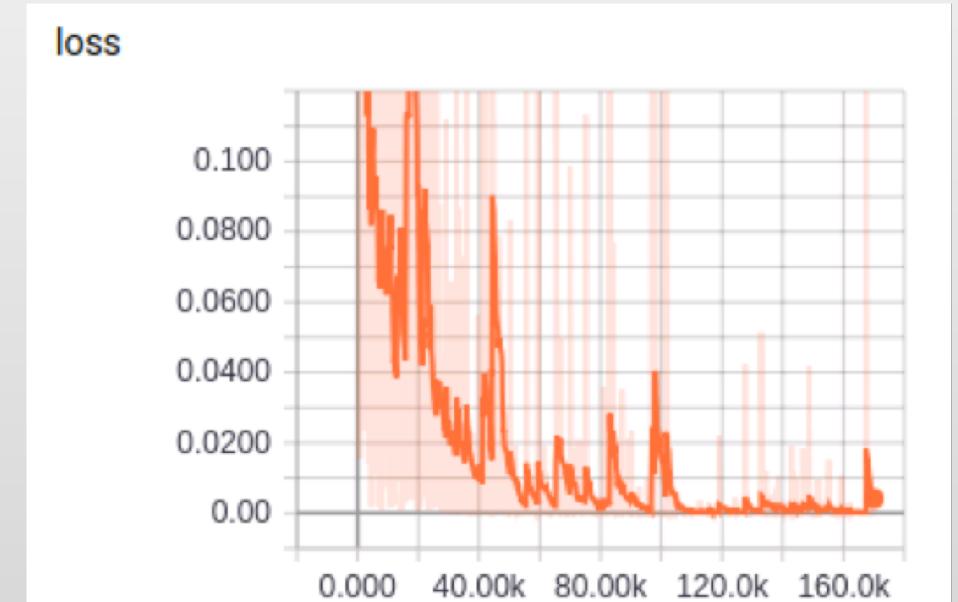


Fig 2. Training loss(batch hard triplet loss – see previous slides).

Manifold

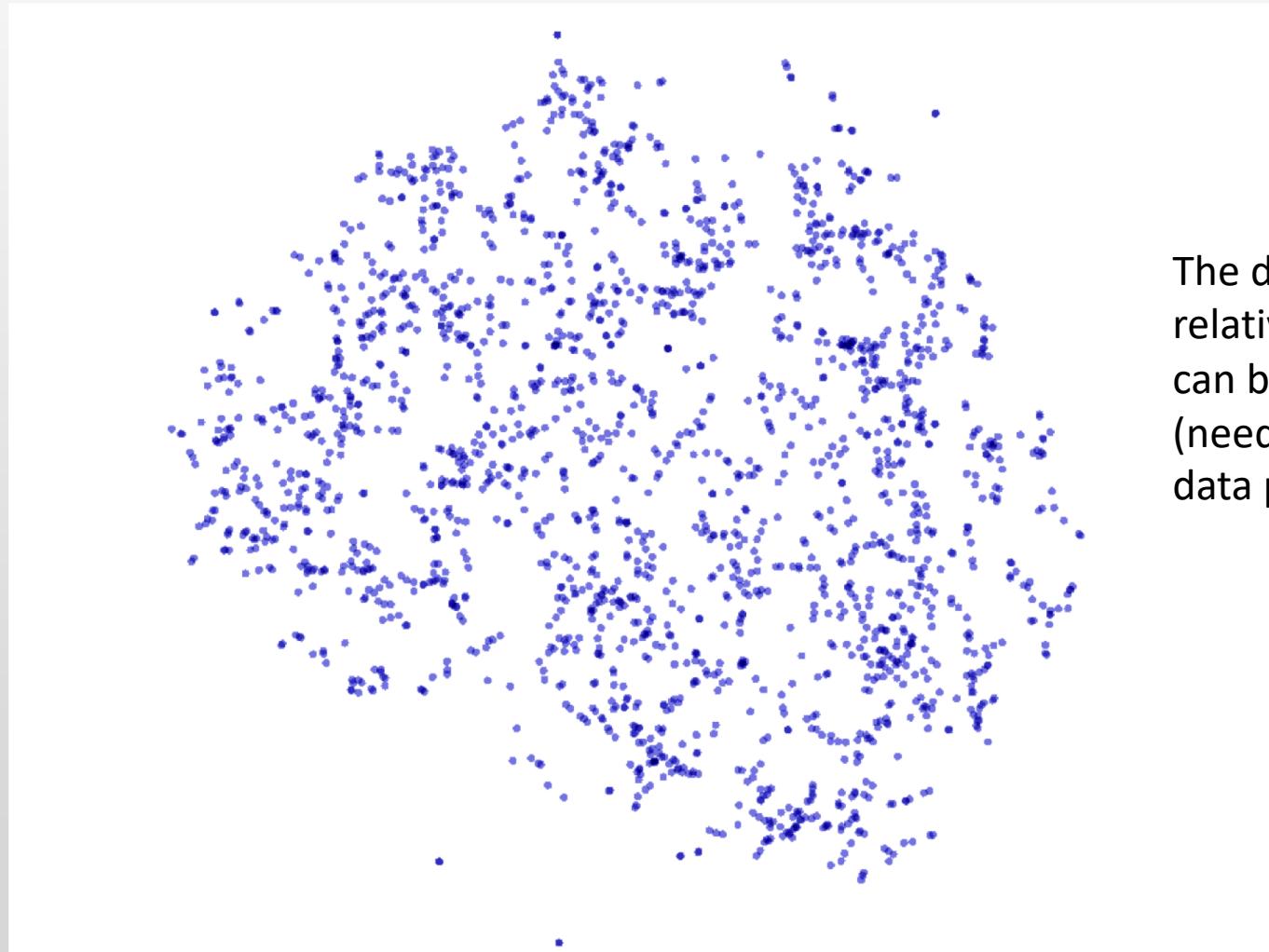


Fig 3. Penultimate layer output(FC, 1024 –d) of 2000 points(randomly chosen) projected onto 2d space (t-SNE).

The distribution looks reasonable, i.e. relatively even but some clustering patterns can be seen.
(need to embed each cropped image onto related data point to really see it.).

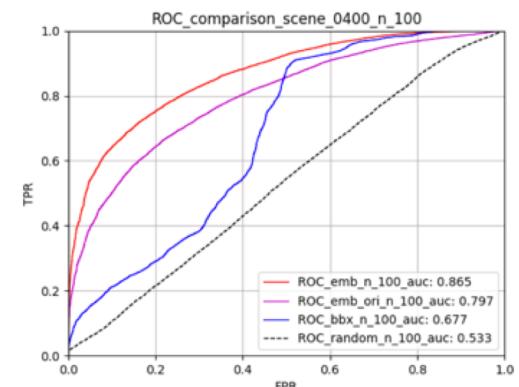
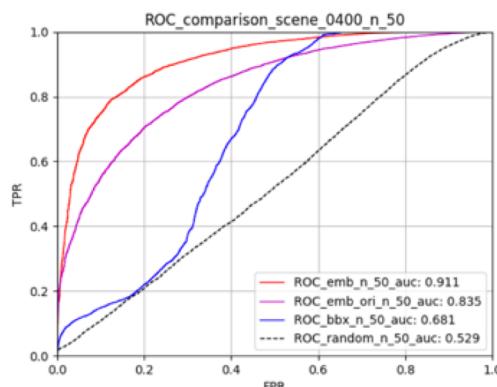
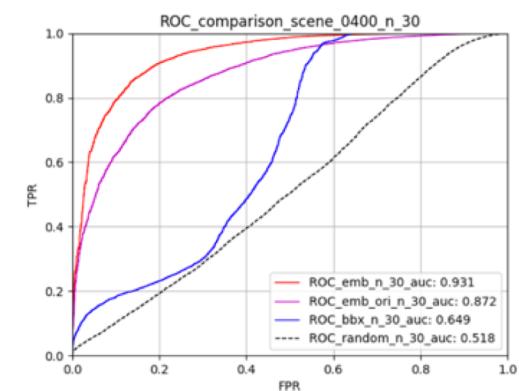
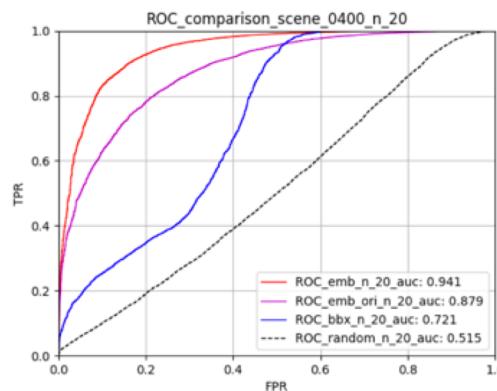
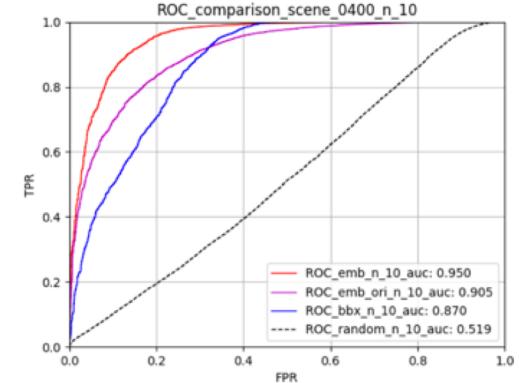
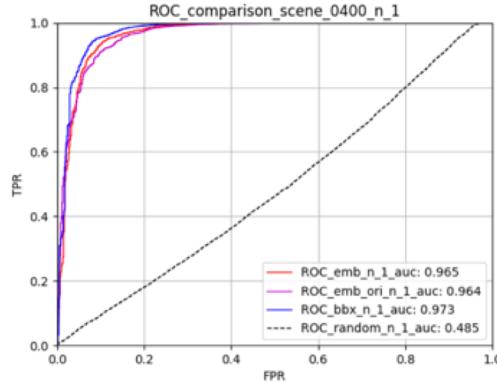
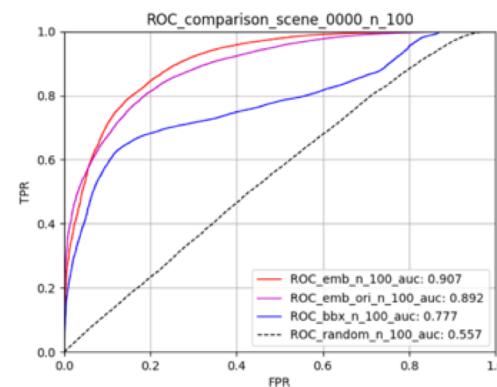
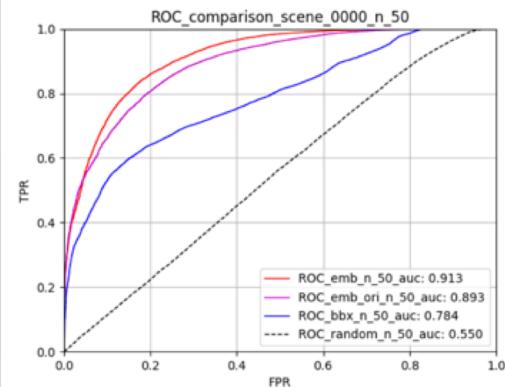
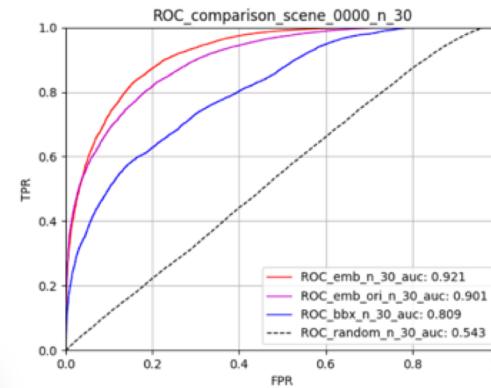
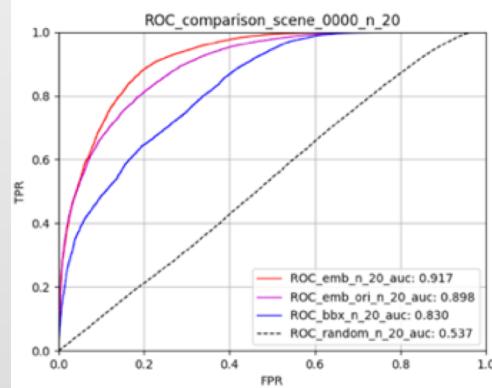
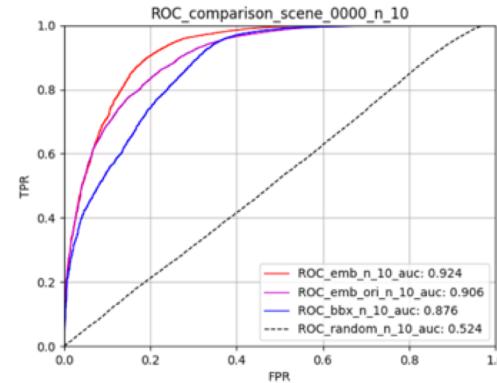
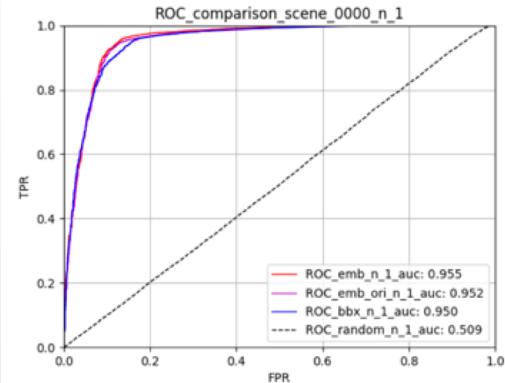


Fig 1. Comparison of ROC curve of scene 0000 for (red) triplet embedding fine-tuned (magenta) trip emb original (blue) bbx geometric distance and (black) random matching at different n.

Fig 2. Comparison of ROC curve of scene 0400 for (red) triplet embedding fine-tuned (magenta) trip emb original (blue) bbx geometric distance and (black) random matching at different n.

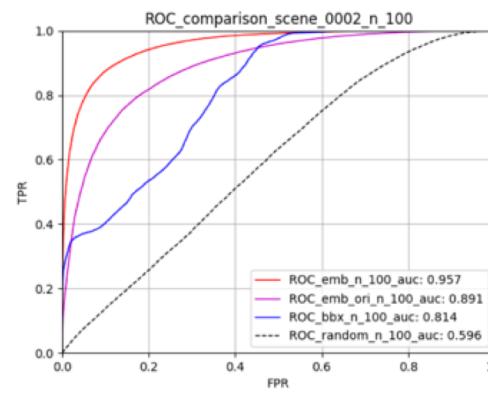
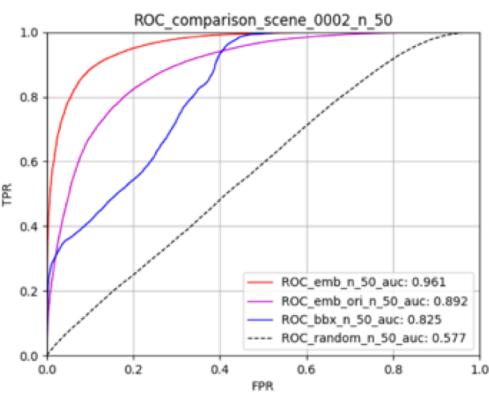
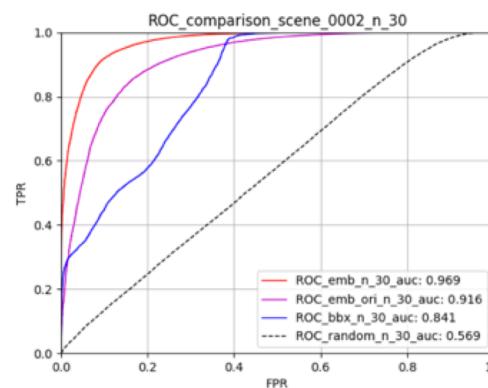
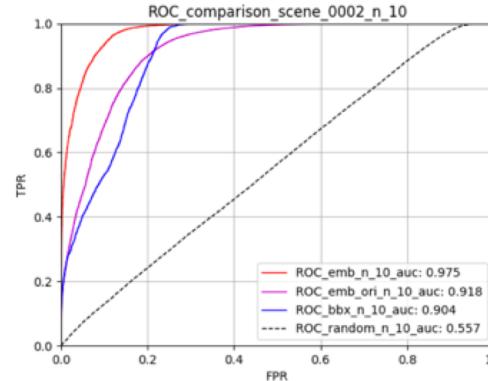
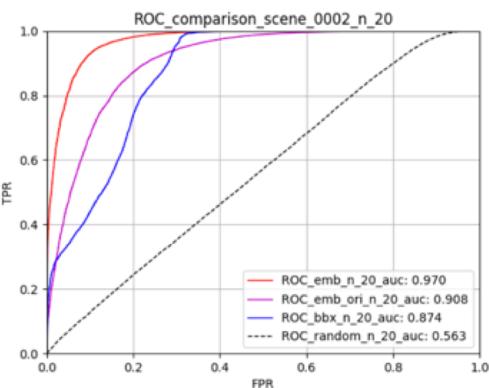
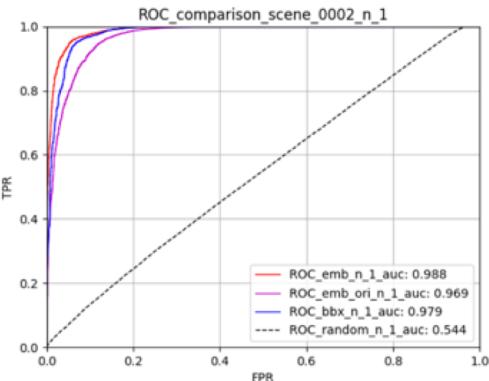
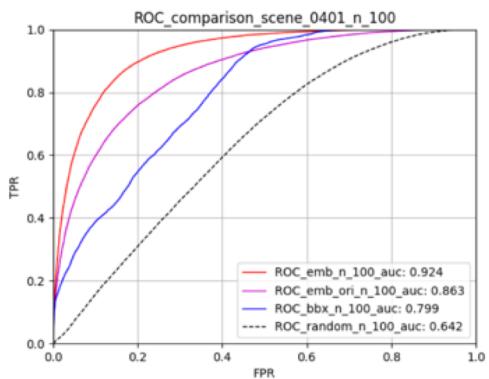
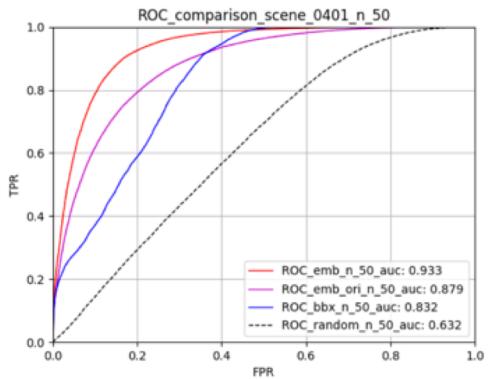
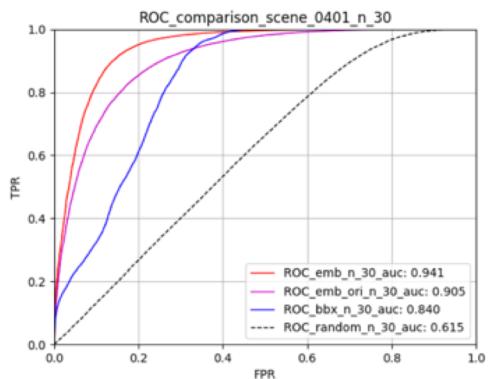
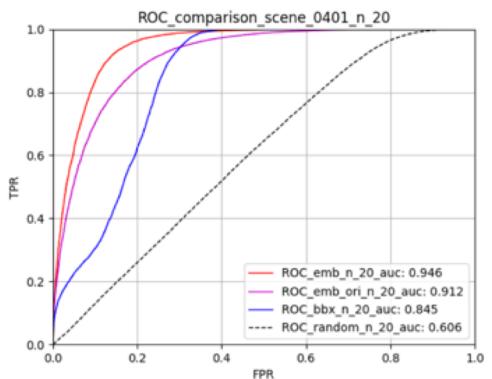
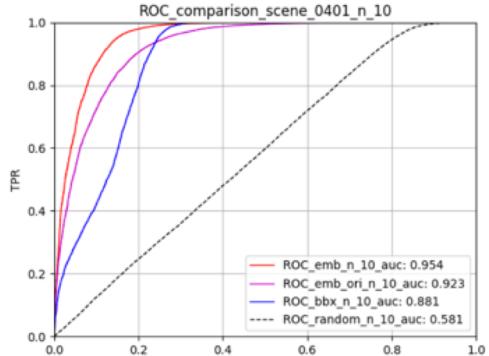
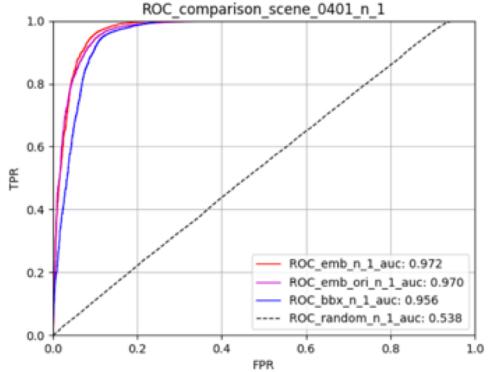


Fig 3. Comparison of ROC curve of scene 0401 for (red) triplet embedding fine-tuned (magenta) trip emb original (blue) bbx geometric distance and (black) random matching at different n.

Fig 4. Comparison of ROC curve of scene 0002 for (red) triplet embedding fine-tuned (magenta) trip emb original (blue) bbx geometric distance and (black) random matching at different n.

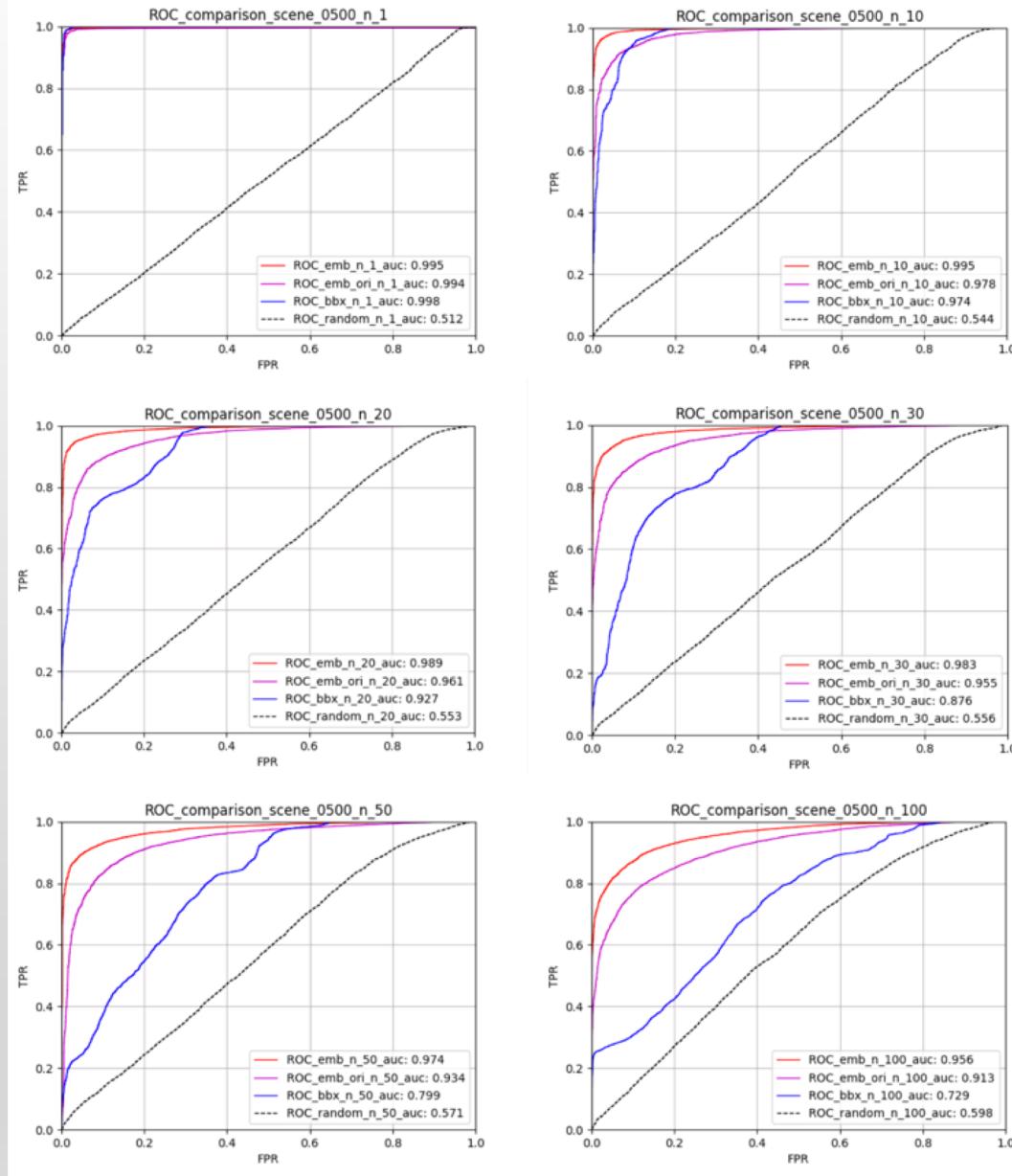


Fig 4. Comparison of ROC curve of scene 0002 for (red) triplet embedding fine-tuned (magenta) trip emb original (blue) bbx geometric distance and (black) random matching at different n .