

週間進捗報告

権藤陸

2022 年 6 月 17 日

1 今週の進捗

- いただいた ReID 論文の読み込み
- SRPNet に関連する論文の参照

2 SRPNet

2.1 PointNet[1]

PointNet の Classification Network をほぼそのまま採用。異なる点は、入力が $3 + 1$ 次元となっている部分である。点群は、以下の 3 つの重要な性質をもつ。

- 順序に依存しない。つまり、 N 個の 3 次元点群を処理する場合、データの順序が変わっても、出力は不変であることが望ましい。
- 点同士は孤立しているわけではなく、隣接する点が意味のある部分集合を形成している。そのため、モデルは局所構造や局所構造間の相互作用を捉えられる必要がある。
- 点群の回転や平行移動といった幾何学変換に対して不変である。

基本的には、MLP を重ね、Input transform と Feature transform という層で、アフィン変換を行い、最後に Max Pooling を行う構造になっている。アフィン変換により、回転・移動による不変性を獲得している。また、Max Pooling により、順序に依存しないモデルとしている。

2.2 Attention

4D-PointNet の出力特徴ベクトルを学習して重みを生成し、その重みを出力特徴ベクトルに乗じることで、意味のある特徴を強調している。SENet[2] と同様の構造だが、全結合層の次元と活性化関数に変更されている。本研究では、Attention Module の活性化関数にソフトマックス関数が用いられている。

2.3 Bi-LSTM

過去と未来のレーダ点群はどちらも有用で補完し合う関係にあると考えられるため、順方向の LSTM と逆方向の LSTM を組み合わせて使用する。2 つの LSTM の出力は、Mean Pooling により結合される。

3 classification vs ReID

論文内で分類タスクと ReID タスクの 2 つのタスクを行っているが、それら 2 つのタスクの違いはあまり言及されていない。分類タスクは交差エントロピー誤差を用いて、文字通り分類を行うタスクであるのに対し、ReID タスクは検索のタスクとみなすことができる。誤差関数には 2 つの画像の類似度を測るトリプレットロスを用いている。ReID のテストセットには、ギャラリーとクエリが存在し、与えられたクエリに対し、モデルがギャラリーから最適な候補を順番に k 個とってきて、上位 k 個の中にクエリと同じ人物がいれば Ok となる。

Classification のタスクでは、学習済みの人物だけ識別・分類できればよいが、ReID ではサンプルが少なく学習が十分でない人物についても識別できる必要がある。実用化を考慮すれば、ReID は識別対象は非協力的な場合が多いため、対象となる人物の人数は多いが、1 人あたりのサンプル数は少ないものとなる。また、ReID ではモデルが異なる時間、異なる環境、異なる視点で正確な識別を達成できることが要求される。

4 損失関数

4.1 交差エントロピー誤差

本研究における交差エントロピー誤差は、ソフトマックス関数との組み合わせで、以下のように記述される。

$$L_{cls} = -\frac{1}{N} \sum_{b=1}^N \sum_{u=1}^K y_{ub} \cdot \log \left(\frac{e^{Z_{ub}}}{\sum_{v=1}^K e^{Z_{vb}}} \right) \quad (1)$$

ここで、 u と v は出力が属するカテゴリ、 K はカテゴリの総数、 N はバッチ内のサンプルの総数、 y_{ub} はサンプル b に対するカテゴリ u のラベルである。これに正則化項を追加し、分類タスクで用いる損失関数となる。

$$L = L_{cls} + \omega \cdot L_{reg} = L_{cls} + \omega \cdot \|I - AA^T\|_F^2 \quad (2)$$

ここで、 L_{cls} は分類損失、 L_{reg} は正則化項損失、 I は単位行列、 A は特徴変換行列、 ω は重みで、0.001 に設定されている。

4.2 トリプレットロス

$$L_{triplet} = \max(d_p - d_n + \alpha, 0) \quad (3)$$

d_p と d_n はそれぞれ同一人物からの特徴量と異なる人物からの特徴量の L_2 距離である。 α はトリプレットロスのマージンである。トリプレットロスにより、同一人物からの特徴量は近くなり、異なる人物からの特徴量は遠くなる。

5 結果

- classification の精度

- SRPNet-15 モデルでは、15 人の個人に対して平均 98.16% の精度を達成。SRPNet-40 モデルでは、40 人の個人に対して平均 91.50% の精度を達成。
- ReID の CMC, mAP

| Method | mAP | CMC-1 | CMC-5 |
|----------------------|-------|-------|-------|
| SRP-Net (40*) | 72.5% | 90.1% | 98.8% |
| RGB-based ReID (20*) | 55.2% | 78.4% | 84.5% |

図 1 ReID タスクの結果

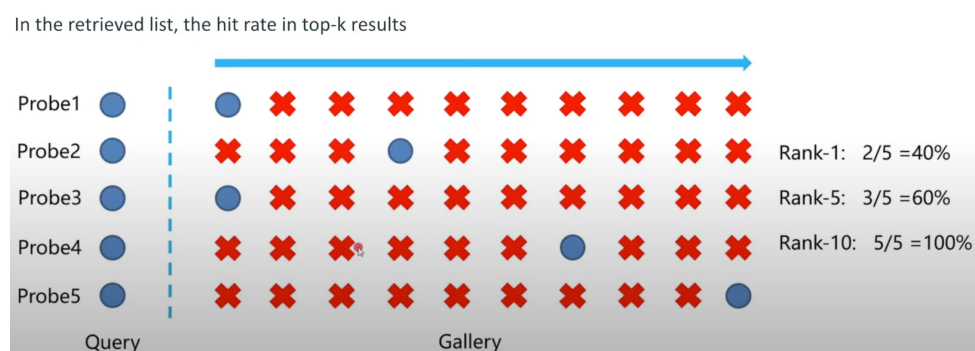


図 2 CMC の計算の模式図 [3]

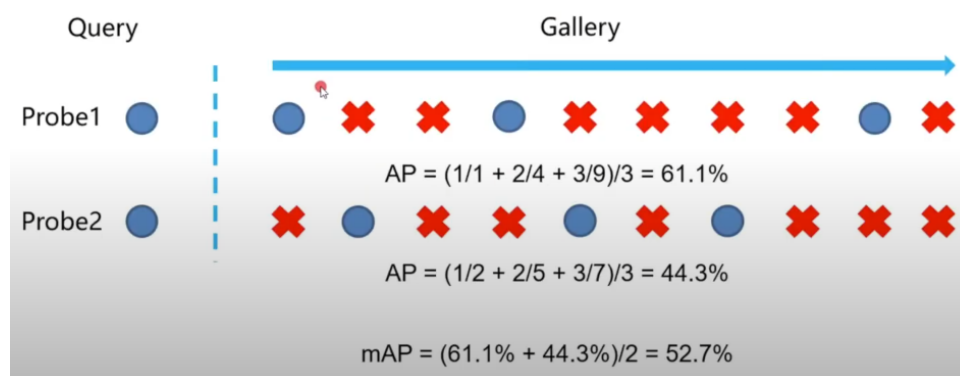


図 3 mAP の計算の模式図 [3]

6 計画

- ReID 論文の読み込み
- スライド作成

参考文献

- [1] i, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [2] u, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [3] Deep Person Re-identification Introduction, <https://www.youtube.com/watch?v=BPnQqHpLSGk>