

深層学習による人物再同定における損失関数への重み付け手法

清水 雄 哉[†](学生会員) 榎 田 修 一[†][†]九州工業大学

Multi-Loss Weighting for Person Re-Identification Based on Deep Learning

Yuya SHIMIZU[†] (*Student Member*), Shuichi ENOKIDA[†][†]Kyushu Institute of Technology

〈あらまし〉 視野を共有しない複数台カメラ間の人物追跡などにおいて、同一人物を判定する人物再同定が重要である。深層学習を用いた人物再同定は高い性能を達成しており、その損失関数として、一般的に Cross-Entropy Loss や Triplet Loss が用いられる。近年、人物再同定に対するアプローチとして、その両方の損失関数を線形和で用いる手法が注目されている。しかし、異なる性質を持つ損失関数を同時に用いる場合、他方への影響を考慮した重み付き線形和による統合が必要となる。そこで本研究では、損失関数の学習効率の違いとそのバランスを考慮し、学習中に損失の重みを自動調整する手法を提案する。

キーワード：深層学習, 人物再同定, 損失関数, cross-entropy loss, triplet loss

<Summary> Person re-identification is an important component to realize various image recognition systems, e.g., person tracking system by utilizing multiple cameras. Person re-identification based on deep learning has achieved high performance, and cross-entropy loss or triplet loss is generally used as the loss function. In recent years, a linear summation of both loss functions has been attracting attention as an approach to person re-identification. However, when loss functions with different properties are used at the same time, a method of synthesis by weighted linear summation that takes into account the effect of the loss function on the other loss function is necessary. To overcome above problems, in this paper, a method that automatically adjusts the weights of loss functions during learning is proposed.

Keywords: deep Learning, person re-identification, loss function, cross-entropy loss, triplet loss

1. はじめに

安心安全な社会のために、公共の様々な場所に設置された防犯カメラは犯罪認知件数の低下などに貢献している。防犯カメラは、単に犯罪の予防だけでなく、警察捜査における被疑者の特定・追跡や犯行の立証、迷子の発見などに活用される。人物再同定は防犯カメラからの人物特定や追跡において、その効率化や自動化のために重要である。人物再同定とは、複数台のカメラや、同一カメラの異なる場面で撮影された人物画像から同一人物を判別する問題である。一般的な人物検出や人物追跡では、一度 ID を付与した人物がカメラ画角から外れ、再び画角内に入った際に異なる ID が付与される。一方、人物再同定では同一の ID を付与することができる。近年、深層学習による人物再同定では、多クラス分類に用いられる Cross-Entropy Loss と、深層距離学習に用いられる Triplet

Loss¹⁾の2種類の損失関数の線形和により学習の損失を求める手法が注目されている。しかし、単純な線形和では、性質の異なる損失関数を常に同程度に評価するため、それぞれの性能を十分に活かせていない可能性がある。そこで本研究では、各損失関数に対し重み付けすることにより、他方の損失からの影響を抑え、効率的な損失の減少を目的とする。また、学習により得られる損失は、学習環境ごとに異なり、適切な重みを統一して用いることはできない。そこで、学習中に得られる損失から、損失関数への重みを計算する重み自動調整手法を提案する。本研究では提案手法を用いた損失関数への重み付けについて、学習モデルの精度へ与える影響について解析する。

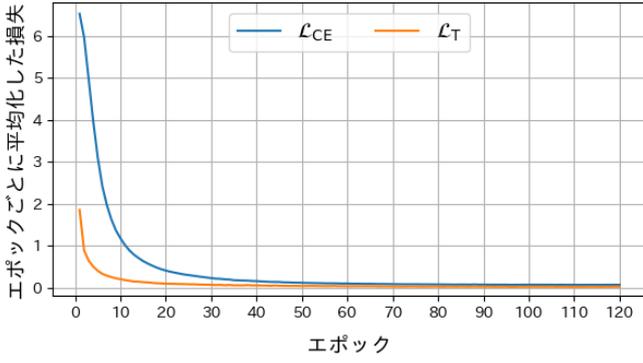


図 1: エポックごとに平均化した損失の推移
Fig. 1 Transition of average losses per epoch

2. 関連研究

人物再同定では、多クラス分類に用いられる Cross-Entropy Loss と、深層距離学習に用いられる Triplet Loss¹⁾の2種類の損失関数の線形和により学習の損失を求める手法が注目され、全体の損失関数 \mathcal{L} は、Cross-Entropy Loss を \mathcal{L}_{CE} 、Triplet Loss を \mathcal{L}_T として式 (1) で表される。

$$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_T \quad (1)$$

この手法を用いた人物再同定に対する学習モデルには、OS-Net²⁾、BDB Network³⁾、そして、TransReID⁴⁾などがあげられる。TransReID は、人物再同定において主流である Convolutional Neural Network ではなく、画像をパッチへ分解して扱う Vision Transformer⁵⁾をモデルベースとして採用した学習モデルである。モデル構造は、主に文献⁶⁾によるベースラインに基づき構成されている。また、新たに局所特徴量への注目を目的とする Jigsaw Patch Module と撮影カメラ環境の考慮を目的とする Side Information Embeddings の2つのモジュールが提案されている。本研究では TransReID を実験における学習モデルとして使用する。

3. 学習中の損失を用いた重み付け手法

3.1 人物再同定における損失関数

3.1.1 Cross-Entropy Loss

Cross-Entropy Loss は、主に多クラス分類問題で利用される損失関数である。損失の計算には確率分布が使用され、真の確率分布を p 、学習モデルにより推定された確率分布を q とし、Cross-Entropy Loss を \mathcal{L}_{CE} とすると、 \mathcal{L}_{CE} は式 (2) で表される。

$$\mathcal{L}_{CE} = - \sum_x p(x) \log(q(x)) \quad (2)$$

式 (2) は、確率分布が似ていると誤差が小さくなり、そうでなければ誤差が大きくなる関数であるため、損失を表すことができる。人物再同定においては各人物に割り当てられる ID を各クラスとみなすことで、多クラス分類問題として扱うこ

とができるため、Cross-Entropy Loss が損失関数として用いられる。

3.1.2 Triplet Loss

Triplet Loss¹⁾は、深層距離学習で用いられる損失関数である。深層距離学習では、データの類似度を表す距離に基づき、同じクラスの場合には距離が近く、異なるクラスの場合には距離が遠くなるように学習する。人物再同定や顔認識では各サンプル数も少なく、個人をクラスとして扱うため、実環境において、多くの場合サンプルが訓練に存在しないクラス、すなわち未知クラスとなる。深層距離学習の特徴として、訓練データにおける各クラスのサンプル数が少ない場合や、推論の際に未知のクラスがある場合にも有効な学習であることがあげられる。よって人物再同定には、深層距離学習の代表的な損失関数である Triplet Loss が用いられる。

基準画像と正解画像との距離を D_P 、基準画像と不正解画像との距離を D_N 、マージンを表すハイパーパラメータを τ とする。Triplet Loss の目的は、 $D_N \geq D_P + \tau$ を満たすように、 D_P を小さく、 D_N を大きくし、ニューラルネットワークにおけるパラメータを調整することである。Triplet Loss を \mathcal{L}_T とすると、 \mathcal{L}_T は式 (3) で表される。

$$\mathcal{L}_T = \max\{D_P - D_N + \tau, 0\} \quad (3)$$

3.2 損失関数への重み付け

図 1 に示すようにエポックごとに平均化した各損失の推移は、学習が進むに連れて同程度に収束しているものの、序盤には比較的大きな差が存在し、その減少傾向も異なっている。よって、式 (1) の損失を求める際に各損失関数に対し重み付けすることにより、他方の損失からの影響を抑え、効率的な損失の減少を検討する。そこで、 \mathcal{L}_{CE} 、 \mathcal{L}_T それぞれに対して重み W_{CE} 、 W_T を与えた損失関数を式 (4) に示す。

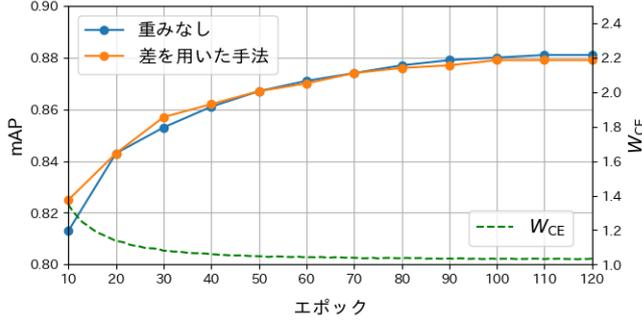
$$\mathcal{L} = W_{CE}\mathcal{L}_{CE} + W_T\mathcal{L}_T \quad (4)$$

深層学習では損失を最小化するように学習することで、学習精度を向上させる。式 (4) は、重要度が高い損失に対して重み付けすることで、効率的な学習の実現を目指す。しかし、あらゆる問題で精度を向上させる一定の重みは存在しないため、重みを適切に調整する必要がある。そこで、学習中に逐次求められる損失から、重みを自動調整する手法を提案する。

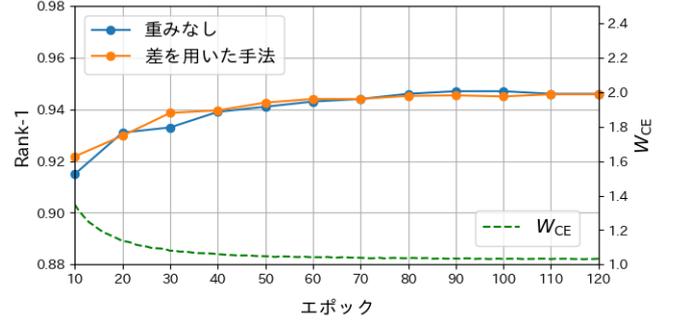
本研究では、 \mathcal{L}_{CE} と \mathcal{L}_T のうち、大きな損失を観測した損失関数の重みを大きくすることを基本方針とする。具体的には、 \mathcal{L}_{CE} と \mathcal{L}_T の差と比をそれぞれ用い、重みを自動調整する。提案手法において、重み調整の実施間隔はエポックごととする。式 (4) における重みの計算には、前エポック中に求められた損失の平均、 $\bar{\mathcal{L}}_{CE}$ および $\bar{\mathcal{L}}_T$ を用いる。また、最初のエポックにおける重み W_{CE} および W_T は 1.0 とした。

3.3 損失の差を用いた重み自動調整手法

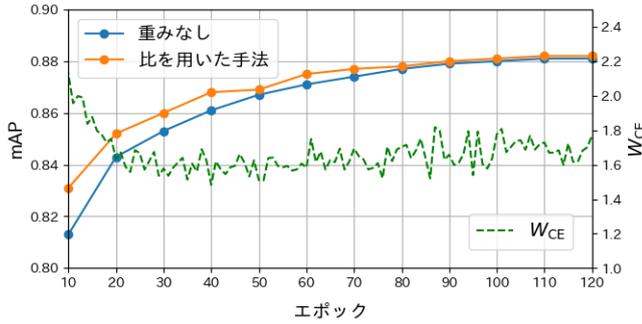
学習における損失について、 $\mathcal{L}_{CE} > 0$ および $\mathcal{L}_T > 0$ である。また、損失の大小比較を用いた説明のために、 $\bar{\mathcal{L}}_L =$



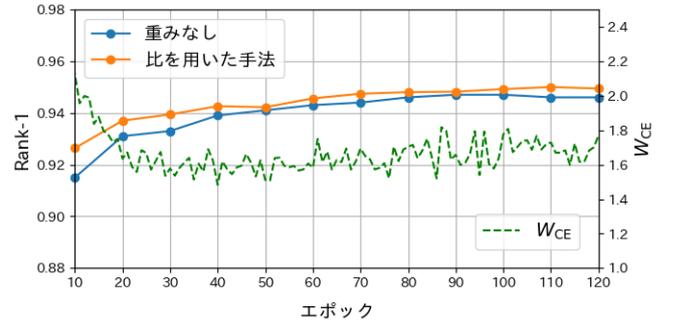
(a) mAP



(b) Rank-1

図 2: 損失の差を用いた重み付け手法における mAP, Rank-1, W_{CE} の推移Fig. 2 The progression of mAP, Rank-1, and W_{CE} in weighting methods using the difference of losses

(a) mAP



(b) Rank-1

図 3: 損失の比を用いた重み付け手法における mAP, Rank-1, W_{CE} の推移Fig. 3 The progression of mAP, Rank-1, and W_{CE} in weighting methods using the ratio of losses

$\max\{\bar{\mathcal{L}}_{CE}, \bar{\mathcal{L}}_T\}$, $\bar{\mathcal{L}}_S = \min\{\bar{\mathcal{L}}_{CE}, \bar{\mathcal{L}}_T\}$ とする.

$\bar{\mathcal{L}}_L - \bar{\mathcal{L}}_S$ を \mathcal{L}_L の重みとする手法を提案する. しかし, そのまま差を用いる場合, $\bar{\mathcal{L}}_L - \bar{\mathcal{L}}_S \leq 1.0$ の際に, \mathcal{L}_L に対する重みが 1.0 より小さくなるため, \mathcal{L}_L の影響が抑えられて学習することとなる. そこで, 求められた差に 1.0 を足すことでそれを防ぐ. ここで, 提案手法では各損失の大小によって W_{CE} , W_T の求め方が変化する. これを踏まえ,

式 (5), (6) により損失の差を用いて W_{CE} , W_T を求める.

$$W_{CE} = \begin{cases} \bar{\mathcal{L}}_{CE} - \bar{\mathcal{L}}_T + 1.0 & (\bar{\mathcal{L}}_{CE} \geq \bar{\mathcal{L}}_T) \\ 1.0 & (\bar{\mathcal{L}}_{CE} < \bar{\mathcal{L}}_T) \end{cases} \quad (5)$$

$$W_T = \begin{cases} 1.0 & (\bar{\mathcal{L}}_{CE} \geq \bar{\mathcal{L}}_T) \\ \bar{\mathcal{L}}_T - \bar{\mathcal{L}}_{CE} + 1.0 & (\bar{\mathcal{L}}_{CE} < \bar{\mathcal{L}}_T) \end{cases} \quad (6)$$

3.4 損失の比を用いた重み自動調整手法

$\bar{\mathcal{L}}_S$ に対する $\bar{\mathcal{L}}_L$ の比率を \mathcal{L}_L の重みとする手法を提案する. そこで, 式 (7), (8) により損失の比を用いて重み W_{CE} , W_T を求める.

$$W_{CE} = \begin{cases} \frac{\bar{\mathcal{L}}_{CE}}{\bar{\mathcal{L}}_T} & (\bar{\mathcal{L}}_{CE} \geq \bar{\mathcal{L}}_T) \\ 1.0 & (\bar{\mathcal{L}}_{CE} < \bar{\mathcal{L}}_T) \end{cases} \quad (7)$$

$$W_T = \begin{cases} 1.0 & (\bar{\mathcal{L}}_{CE} \geq \bar{\mathcal{L}}_T) \\ \frac{\bar{\mathcal{L}}_T}{\bar{\mathcal{L}}_{CE}} & (\bar{\mathcal{L}}_{CE} < \bar{\mathcal{L}}_T) \end{cases} \quad (8)$$

4. 評価実験

評価実験におけるデータセットとして, 人物再同定に広く用いられる Market1501⁷⁾を使用した. Market1501 は, 訓練用データに 12,936 枚, テスト用に 19,732 枚, 計 32,668 枚の人物の矩形領域画像で構成され, 各人物に対応する 1,501 個の ID が付与されている. TransReID モデルの学習において, 事前学習モデルには TransReID がベースラインと定める ViT-B/16 を使用した. 訓練データに対し 120 エポックの学習を 5 回行い, その平均を結果として用いる. 評価指標には, 人物再同定における慣例として, mean Average Precision(mAP) および Cumulative Matching Characteristic(CMC) 曲線における Rank-1 を用いる.

5. 結果

重みなしの場合と損失の差による手法の学習結果について mAP, Rank-1 および W_{CE} の推移を図 2 に示す. 同様に, 重みなしの場合と損失の比による手法の学習結果について

6. 結 論

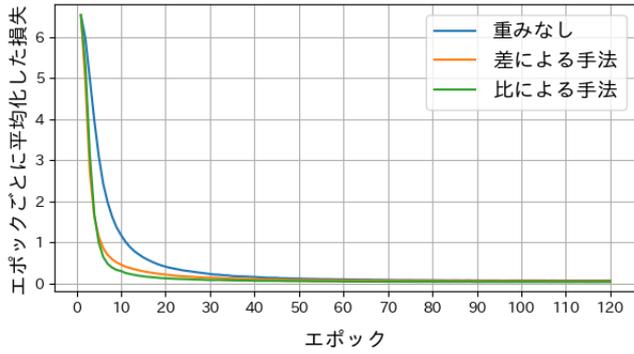
本研究では、深層学習による人物再同定において、Cross-Entropy Loss と Triplet Loss の 2 つの損失関数に対して動的に重み付けする学習手法を提案した。実験においては、学習精度に与える影響について解析し、特に損失の比を用いた手法について学習中いつれの時点においても精度の向上を確認した。なお、本論文では片方の損失に対してのみ重みを付ける手法に留まった。今後は、人物再同定における学習効率のさらなる向上のため、提案手法を様々な学習モデルやデータセットに適用して効果を多面的に調査していく。また、両方の損失に対し同時に重み付けする手法の検討が課題であると考えている。

参考文献

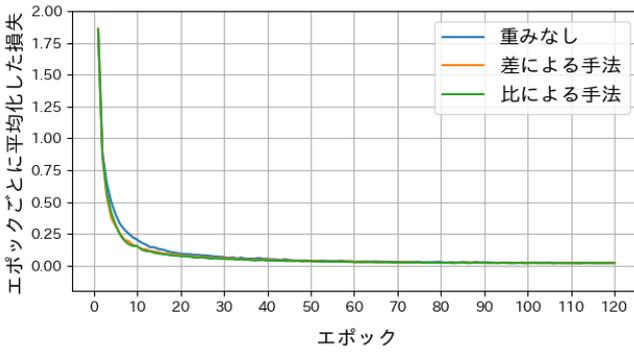
- 1) A. Hermans, L. Beyer, B. Leibe: “In Defense of the Triplet Loss for Person Re-Identification”, arXiv preprint arXiv:1703.07737(2017).
- 2) K. Zhou, Y. Yang, A. Cavallaro, T. Xiang: “Omni-Scale Feature Learning for Person Re-Identification”, Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.3702-3712(2019).
- 3) Z. Dai, M. Chen, X. Gu, S. Zhu, P. Tan: “Batch Dropblock Network for Person Re-Identification and Beyond”, Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.3691-3701(2019).
- 4) S. He, H. Luo, P. Wang, F. Wang, H. Li, W. Jiang: “TransReID: Transformer-Based Object Re-Identification”, Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.15013-15022 (2021).
- 5) D. Alexey, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, others: “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale”, International Conference on Learning Representations(2021).
- 6) H. Luo, Y. Gu, X. Liao, S. Lai, W. Jiang: “Bag of Tricks and a Strong Baseline for Deep Person Re-Identification”, Proc. of ICCV and pattern recognition workshops(2019).
- 7) L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian: “Scalable Person Re-identification: A Benchmark”, Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.1116-1124 (2015).

(2023年6月29日 受付)

(2023年10月13日 再受付)



(a) $\bar{\mathcal{L}}_{CE}$



(b) $\bar{\mathcal{L}}_T$

図 4: エポックごとに平均化した損失の各手法における推移比較

Fig. 4 Comparison of transitions in average losses per epoch among different methods.

mAP, Rank-1 および W_{CE} の推移を図 3 に示す。学習中の損失について、どちらの提案手法においても結果として常に $\bar{\mathcal{L}}_{CE} > \bar{\mathcal{L}}_T$ であり、Cross-Entropy Loss に対する重み W_{CE} のみが求められたため、 W_T は常に 1.0 であった。さらに、図 4 に各手法における損失の推移を示す。

図 2 より、学習前半において、損失の差による手法では mAP, Rank-1 とともに、10, 30 エポックでは重みなしの場合よりも向上し、その後は同程度で推移した。また、損失の比による手法では、差による手法では見られない学習中盤までの精度の向上が見られ、終盤では同程度で推移した。比による手法ではいずれの時点でも重みなしの場合以上の精度を示したことから、全体を通して収束が早まったことがわかる。

各手法における学習前半での精度向上について、図 4 より、重みなしの場合に比べて損失が早い段階で減少していることから、提案手法による重みが効果的であると考えられる。また、学習後半では、各手法において W_{CE} の振る舞いが異なり、差による手法では、学習が進むにつれて 1.0 に収束したことから重みの効果が薄かったと考えられる。比による手法では、学習後半において、 W_{CE} は 1.5 から 1.8 付近で変化し続け、損失の差が小さくなった場合でもその違いを捉えることができ、学習に対して差による手法より効果的であったと考えられる。



清水 雄哉 (学生会員)

2023 年 九州工業大学情報工学部卒. 現在, 同大学情報工学府情報創成工学専攻修士課程に在籍. 画像情報処理, パターン認識に関する研究に従事.



榎田 修一

2002 年, 九州工業大学大学院情報工学研究科博士後期課程修了, 同年九州工業大学情報工学部情報工学研究科助手. 2007 年, 同大学 助教, 2009 年, 九州工業大学大学院情報工学研究院准教授. 2017 年, 同大学院 教授. 現在に至る. 画像情報処理, パターン認識に関する研究に従事. 博士 (情報工学). 情報処理学会, 日本ロボット学会, 自動車技術会 各会員.