Adaptive Gradient Triplet Loss with Automatic Margin Learning for Forensic Medical Image Matching

Khanh Nguyen[*1] KHANH.NGUYEN@OULU.FI
Huy Hoang Nguyen[*1] HUY.NGUYEN@OULU.FI
Aleksei Tiulpin[1] ALEKSEI.TIULPIN@OULU.FI
[1] *University of Oulu, Oulu, Finland*

Editors: Under Review for MIDL 2022

Abstract

This paper tackles the challenge of forensic medical image matching (FMIM) using deep neural networks (DNNs). We investigate Triplet loss (TL), which is probably the most well-known loss for this problem. TL aims to enforce closeness between similar and enlarge the distance between dissimilar data points in the image representation space extracted by a DNN. Although TL has been shown to perform well, it still has limitations, which we identify and analyze in this work. Specifically, we first introduce AdaTriplet – an extension of TL that aims to adapt loss gradients according to the levels of difficulty of negative samples. Second, we also introduce AutoMargin – a technique to adjust hyperparameters of margin-based losses such as TL and AdaTriplet dynamically during training. The performance of our loss is evaluated on a new large-scale benchmark for FMIM, which we have constructed from the Osteoarthritis Initiative cohort. The codes allowing replication of our results have been made publicly available at https://github.com/Oulu-IMEDS/AdaTriplet.

Keywords: Deep Learning, Content-based Image Retrieval, Forensic Matching.

1. Introduction

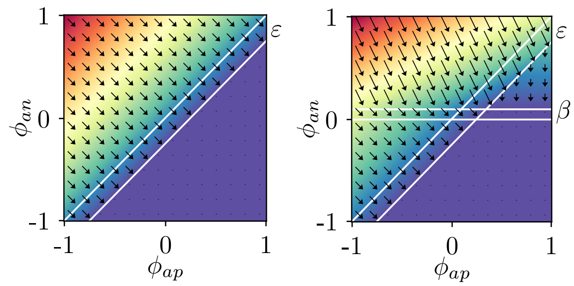
Forensic medical image matching (FMIM) is an instance of content-based image retrieval problem (CBIR), and may be of use in crime investigation and disaster victim identification. However, visual changes in medical data that appear due to aging or disease progression may pose significant challenges for FMIM to work well in practice.

The majority of literature in CBIR is based on deep metric learning (DML), the idea of which is to learn image representations having the following structure: the representations are close under some distance measure, if the images are semantically close, and far otherwise.

Triplet loss (TL) (Hoffer and Ailon, 2015) is one of the most common losses used in DML. Specifically, TL considers three data points: anchor, positive and negative. The case is named positive if it is of the same category as anchor, and negative otherwise. In FMIM, positive cases are labeled as images belonging to the same subject.

Despite being commonly used, TL has two following limitations: it ignores the magnitude of difference between hard negative samples and anchors, and it is heavily dependent on a margin hyperparameter that needs to be found empirically. In this work, we propose an improved version of TL – AdaTriplet, as well as an adaptive margin adjustment method AutoMargin, to address the aforementioned issues of TL.

[*] Contributed equally



(a) Triplet loss (b) AdaTriplet loss

Figure 1: 2D loss surfaces on the $\phi_{ap}\phi_{an}$ coordinate system. Each point represents a triplet. Loss magnitude is indicated by colors increasing from purple (0 loss) to red. Arrows are negative gradient vectors.

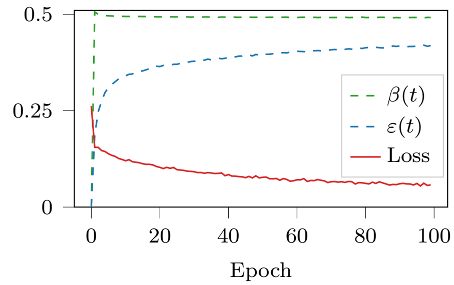


Figure 2: Margins and our loss during training phase.

2. Methodology

Adaptive Gradient Triplet Loss Triplet loss (TL) is an embedding loss that operates on the feature space of input data to enforce the closeness of similar objects in the embedding space, and the remoteness of the dissimilar ones. Precisely, given a triplet of embeddings with unit norm $\mathbf{f}_a$ (anchor), $\mathbf{f}_p$ (positive), and $\mathbf{f}_n$ (negative), triplet loss is formulated as $\mathcal{L}_{\text{Triplet}} = [\phi_{an} - \phi_{ap} + \varepsilon]_+$, where $[.]_+ = \max(0, .)$, ε is a non-negative margin variable, $\phi_{ap} = \mathbf{f}_a^\intercal \mathbf{f}_p$ and $\phi_{an} = \mathbf{f}_a^\intercal \mathbf{f}_n$. We observe that the TL fails to optimize easy triplets that have both positive and negative embeddings close to the anchor (i.e. $\phi_{ap} - \phi_{an} \geq \varepsilon$, $\phi_{ap} > 0$, and $\phi_{an} > 0$). Our empirical evidence shows that this issue cannot be resolved by simple adjustment of ε. To tackle the problem, we enforce a threshold on a cosine distance between $\mathbf{f}_a$ and $\mathbf{f}_n$. Formally, we propose an *adaptive gradient triplet loss*, named AdaTriplet, defined as

$$\mathcal{L}_{\text{AdaTriplet}} = [\phi_{an} - \phi_{ap} + \varepsilon]_+ + \lambda\,[\phi_{an} - \beta]_+\,, \tag{1}$$

where $\lambda \in \mathbb{R}_+$ is a coefficient, $\varepsilon \in [0, 2)$ is a strict margin, and $\beta \in [0, 1]$ is a relaxing margin. Whereas TL has a static gradient field, our AdaTriplet loss has an adaptive gradient field, as graphically illustrated in Fig. 1.

AutoMargin: Adaptive Hard Negative Mining Instead of performing an exhaustive grid search for fixed values of ε and β, we propose AutoMargin that dynamically adjusts the margins during the training phase. Let $\Delta = \phi_{ap} - \phi_{an}$, we rewrite (1) as $\mathcal{L}_{\text{AdaTriplet}} = [\varepsilon - \Delta]_+ + \lambda\,[\phi_{an} - \beta]_+$. ε and β are adaptively updated according to the summary statistics of the Δ and ϕ_{an} distributions:

$$\varepsilon(t) = \frac{\mu_\Delta(t)}{K_\Delta}, \tag{2}$$

$$\beta(t) = 1 + \frac{\mu_{an}(t) - 1}{K_{an}}, \tag{3}$$

where $\varepsilon(0) = \beta(0) = 1$, $K_\Delta, K_{an} \in \mathbb{Z}_+$ are hyperparameters, and $\mu_\Delta(t)$ and $\mu_{an}(t)$ are the means of Δ's and ϕ_{an}'s values at time step t, respectively. $\varepsilon(t)$ is enforced to reach the highest possible value. On the other hand, we aim to increase the threshold on virtual angles between anchors and negative samples, which leads to the decrease of $\beta(t)$.

(*a*) $\mathcal{L}_{\text{an}}$'s effect

λ	**mAP***	**mAP**
0	95.6	96.6
0.5	96.1	96.9
1	**96.3**	**97.0**
2	94.5	95.6

(*b*) AutoMargin

Method	N_s	**mAP**
Quantile Q1	1	94.3
Quantile Q2	1	88.9
Grid search	16	97.0
AutoMargin	4	**97.1**

Table 1: Ablation studies. * indicates the results when the query and the database are 6 years apart. N_s is the number of scanned times.

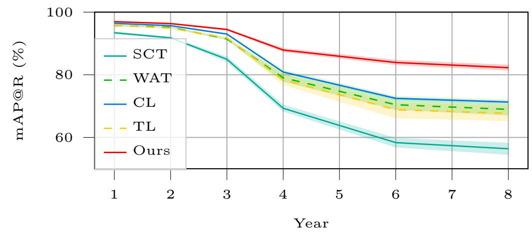


Figure 3: Performance comparisons (mean and standard errors over 5 runs).

3. Experiments and Results

Experimental setup Based on the Osteoarthritis Initiative (OAI) cohort (https://nda.nih.gov/oai/), we generated a novel FMIM benchmark comprising knee X-ray images from 4,796 subjects between 45 and 79 years old acquired at different visits (baseline, 12, 24, 48, 72, and 96 months). The total dataset contained 53,058 images, cropped to 140×140mm (resized to 256×256 pixels) and centered at the intercondylar notch using (Tiulpin et al., 2019). Data from the follow-up OAI visits were used as queries. We assessed the performance using mean average precision (mAP) and mAP@R. In all experiments, we used ResNet18.

Results Table 1(*a*) shows that the term $\mathcal{L}_{an}$ had positive impact on improving the matching performance, and we thus used AdaTriplet for the rest of the experiments. With a small size of AutoMargin grid search space – $K_\Delta \in \{2, 4\}$, $K_{an} \in \{2, 4\}$ (number of hyperparameters scans $N_s = 4$), we were able to achieve the same results as the extensive grid search for ε and β ($N_s = 16$). Quantiles Q1 or Q2 of empirical distribution of Δ were used as baselines in auto-tuning ε and β. Table 1(*b*) shows that AutoMargin yields a good trade-off between N_s and the performance. We found that $K_\Delta = 2$ and $K_{an} = 2$ yielded the best results. Moreover, Figure 2 demonstrates the convergences of the margins ε, β, together with our loss in the training phase. Furthermore, we compared the combination of the AdaTriplet with AutoMargin to other metric learning baselines such as SCT (Xuan et al., 2020), WAT (Zhao et al., 2019), contrastive loss (CL), and TL. In Figure 3, we show the proposed methodology outperformed the baseline approaches. Notably, the further the queries were away from the baseline, the more significant the performance difference was.

References

Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In *SIMBAD*, 2015.

Aleksei Tiulpin, Iaroslav Melekhov, and Simo Saarakkala. Kneel: knee anatomical landmark localization using hourglass networks. In *ICCV*, 2019.

Hong Xuan, Abby Stylianou, Xiaotong Liu, and Robert Pless. Hard negative examples are hard, but useful. In *ECCV*, 2020.

Xiaonan Zhao, Huan Qi, Rui Luo, and Larry Davis. A weakly supervised adaptive triplet loss for deep metric learning. In *ICCV*, 2019.