

Сферические ограничения в триплетной функции потерь*

К.К. Сулоев^{II}, А.В. Шешкус^{I,II}, В.Л. Арлазаров^I

^I Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

^{II} ООО «Smart Engines Service», г. Москва, Россия

Аннотация. Обучение с триплетной функцией потерь – один из самых распространенных подходов в метрическом обучении. Оно находит свое применение в задачах сравнения изображений, идентификации, кодировании и др. Однако триплетная функция потерь обладает рядом недостатков, которые могут негативно сказываться на качестве, например таких, как склонность сети к застреванию в локальных минимумах и формирование тривиальных триплетов. В данной работе предлагается геометрический подход по улучшению качества, основанный на введении дополнительного слагаемого в функцию потерь. Изменение траектории достигается за счет перенаправления решенных и нерешенных изображений к поверхностям двух концентрических гиперсфер разного радиуса. Применение данного метода способствует уменьшению расстояний между изображениями одного класса. Предложенный метод не препятствует использованию других модификаций функции потерь. Экспериментально показано, что предложенный подход позволяет уменьшить часть нерешенных триплетов и далеких пар изображений одного класса.

Ключевые слова: искусственные нейронные сети, метрическое обучение.

DOI: 10.14357/20790279230205

Введение

Метрическое обучение – популярный и перспективный в последнее время способ обучения искусственных нейронных сетей. При таком подходе модель глубокого обучения должна преобразовывать изображения в числовой вектор, содержащий информацию об исходном изображении.

Метрическое обучение находит свое применение во многих задачах компьютерного зрения, таких как идентификация [17], сегментация [7], распознавание текста [8], классификация [2, 11] и поиск структурных элементов [16]. Этот подход, в основном, может быть реализован двумя способами. Первый способ – использование *контрастной* функции потерь [6]. При ее использовании векторные представления изображений из одного класса должны притягиваться друг к другу, а векторные представления изображений из разных классов – отталкиваться. Недостаток данного подхода заключается в том, что при обучении с контрастной функцией потерь, векторы из разных классов долж-

ны отстоять друг от друга не меньше, чем на фиксированное расстояние, что накладывает сильные ограничения на процесс обучения. Естественным развитием данной технологии, исправляющей эти недостатки, стала *триплетная* функция потерь (1). При таком подходе формируются тройки изображений – *триплеты*, состоящие из якоря, позитивного и негативного изображений. Цель обучения – минимизация расстояния между векторными представлениями якоря и позитивного изображения, а так же максимизация расстояния между векторными представлениями якоря и негативного изображения [9]. Триплет будет считаться решенным, если расстояние между якорем и позитивным изображением не менее чем на значение порога m меньше, чем между якорем и негативным изображением:

$$L_{\text{triplet}} = \max(|f(A) - f(P)|_2^2 - |f(A) - f(N)|_2^2 + m, 0). \quad (1)$$

Во время обучения такой сети изначальные векторные представления изменяются: происходит их движение по сложным траекториям. Векторные

* Исследование выполнено за счёт гранта Российского научного фонда (проект номер 23-21-00524).

представления изображений одного класса двигаются ближе друг к другу, формируя кластеры, а векторы других классов отдаляются от них. Разброс расстояний между элементами одного класса характеризуется *внутриклассовой* дисперсией, а разброс расстояний между элементами разных классов – *межклассовой* дисперсией. Показатель качества сформировавшихся структур векторных представлений – низкие значения внутриклассовой и высокие значения межклассовой дисперсий.

Однако [1, 17, 9] показывают, что триплетная функция потерь обладает рядом недостатков. Известно [17, 19], что применение триплетной функции потерь без дополнительных модификаций не позволяет получить максимально возможное качество, так как при обучении сеть склонна к застреванию в локальных минимумах. Одна из причин такого поведения заключается в том, что сеть очень быстро обучается корректно решать *тривиальные* триплеты, отчего большая часть данных становится неинформативной [1, 4, 9, 15].

Еще одна из проблем метрического обучения заключается в том, что расстояние между *центроидами* (усредненные векторы всех элементов класса) соседних классов в результирующем пространстве оказывается меньше, чем расстояние между элементами одного класса. Большие значения внутриклассовой дисперсии и малые расстояния между центроидами разных классов вредят точности дальнейшей классификации [12].

В силу того, что не существует однозначно эффективного способа улучшения качества сети с триплетной функцией потерь, это улучшение достигается путем введения дополнительных модификаций и комбинаций уже имеющихся. Одно из направлений для модифицирования – выбор стратегии по генерации триплетов. Рассмотрим некоторые из них.

В работе [9] авторы одними из первых применили триплетную функцию потерь для задачи идентификации лиц. Ими также был предложен подход по формированию триплетов со сложным негативным примером. Предложенный ими метод позволяет исключить из обучения неинформативные *тривиальные* триплеты.

В той же работе отмечено, что формирование самых *сложных* триплетов приводит к застреванию сети в локальном минимуме, или *коллапсу*, поэтому был предложен подход по добыче триплетов умеренной сложности (рис. 1). В результате было получено ускорение процесса обучения, а также улучшение качества идентификации лиц.

Авторы [13] предлагают более эффективный способ формирования триплетов, основанный на

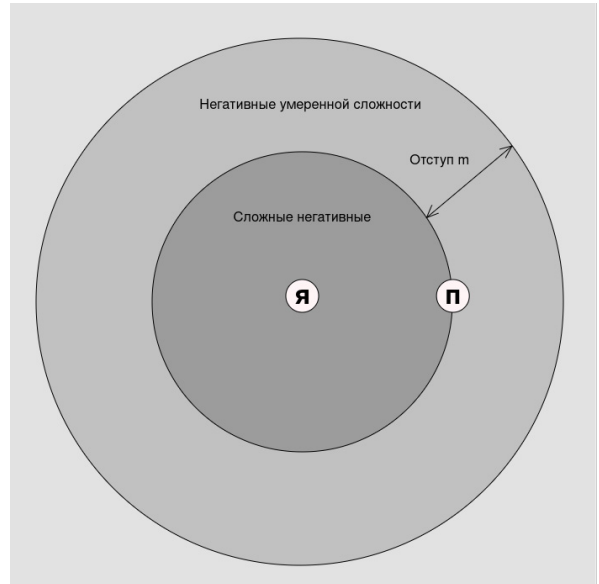


Рис. 1. Типы триплетов

алгоритме ранжирования и вычисления релевантности изображения. Релевантность конкретного изображения вычисляется, как сумма расстояний до всех изображений данного класса. Предложенный авторами подход также вводит концепцию *внеклассового* негативного примера (пример из отличного от якоря класса) и *внутриклассового* негативного примера (пример из того же класса, что и якорь, но обладающий низким значением релевантности). Хотя предложенный подход и помогает увеличить точность ранжирования изображений, он оказывается неприменим для других задач, в частности, в задаче идентификации. При таком подходе возможна ситуация, когда расстояние между изображениями одного класса в процессе обучения будет увеличено. Это сильно затрудняет последующую кластеризацию таких векторных представлений.

В работе [4] при обучении был применен итерационный процесс, на каждой итерации которого человек выбирал неверно классифицированные изображения. Таким образом формировалась выборка сложных негативных примеров. В результате был получен большой прирост в качестве итоговой классификации. К недостаткам данного подхода можно отнести невозможность разметки данных человеком, не имеющим опыта в данной области, а также ограниченную производительность человека и большие временные затраты.

Второе направление для модификаций – изменение непосредственно функции потерь. При введении модификаций, аналогично генерации триплетов, борются с тривиальными триплетными, порождающими нулевой градиент, а также с застреваниями сети в локальных минимумах.

Авторы [18] пытаются решить проблему большой внутриклассовой дисперсии. Предлагается ввести адаптивный порог $m_{adaptive}$, зависящий от входных данных. Таким образом при обучении учитывается влияние тех триплетов, которые при фиксированном пороге являлись бы решенными. Данный метод позволяет снизить внутриклассовую дисперсию.

Решение проблемы малых расстояний между центрами соседних классов предлагают в работе [3]. При обучении к триплету добавляется еще один негативный пример, что формирует *квадруplet*. Авторы показывают, что такой подход позволяет увеличить расстояние между центроидами разных классов.

В работе [14] авторы предлагают использовать введенную ими *центроидную* триплетную функцию потерь. Вместо вычисления расстояния между якорем, позитивным и негативным примерами, вычисляется расстояние между якорем и центроидами соответствующих классов.

Важно также отметить, что при изменении векторного представления изображения в процессе обучения возможна ситуация, когда его траектория движения лежит через уже сформированные структуры, относящиеся к другим классам. Из-за отталкивания между элементами различных классов эти структуры могут быть разрушены, либо препятствовать перемещению рассматриваемого примера в нужную позицию. Это повышает вероятность застревания сети в локальном минимуме.

В данной работе предложен геометрический метод по стабилизации обучения нейронной сети с триплетной функцией потерь путем введения сферических ограничений. Суть предложенного метода заключается в корректировании траекторий движений векторных представлений изображений при обновлении весов нейронной сети. Предлагаемые ограничения разделяют изображения, векторные представления которых уже разведены на необходимую дистанцию от тех, которые еще не являются решенными, по поверхностям двух гипербол разных радиусов. Предложенный подход реализован путем введения дополнительного слагаемого и не противоречит другим модификациям триплетной функции потерь, описанным выше.

1. Постановка задачи

Рассматривается обучение сети с триплетной функцией потерь. На рис. 2 схематически изображена траектория движения векторного представления изображения класса “1” в процессе обучения. Цель обучения – перенести объект класса “1” в об-

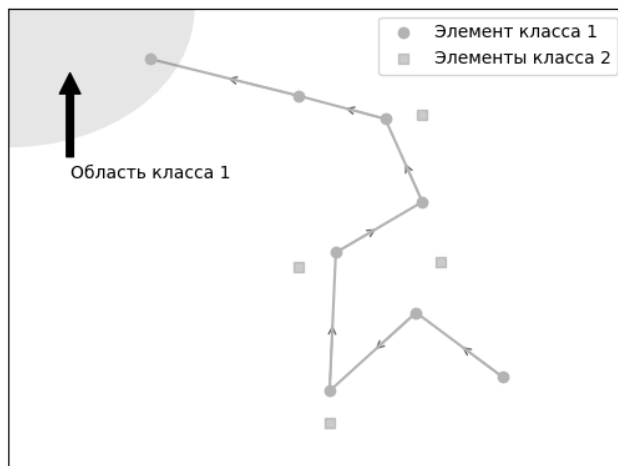


Рис. 2. Движение векторного представления изображения при использовании триплетной функции потерь

ласть, где содержатся другие объекты класса “1”. Можно видеть, что траектория движения объекта протекает через область с объектами класса “2”. При прохождении через эту область объект класса “1” может получить градиент, направленный противоположно этой области. Это может замедлить процесс обучения и привести к застреванию сети в локальных минимумах.

Таким образом, задача разрабатываемого метода заключается в построении такой траектории движения векторного представления изображения, которая бы не пролегла (минимизировала бы путь) через уже устоявшиеся кластеры решенных изображений. Предложенный в работе подход будет протестирован при обучении сети. Ожидается, что итоговые и промежуточные (в процессе обучения) векторные представления изображений будут удовлетворять предложенным в работе геометрическим ограничениям, что приведет к улучшению качества сети.

2. Описание подхода

Будем называть триплет из якоря, позитивного и негативного примеров *решенным*, если для него оказался выполнен критерий, то есть расстояние до позитивного примера не менее, чем на t меньше, чем расстояние до негативного примера, иначе триплет становится *нерешенным*. Зафиксируем в результирующем пространстве две гиперсферы с радиусами r_0 и r_1 . Будем сдвигать решенные триплеты к гиперсфере с большим радиусом, а нерешенные триплеты – с меньшим радиусом. Такой порядок выбран в силу того, что для решенных примеров предполагается больший объем пространства, в котором они могут находиться. Та-

кого эффекта можно достичь, добавив *позиционное* слагаемое в функцию потерь:

$$L = L_{triplet} + q \cdot |f(A)|_2(|f(A)|_2 - r) + q \cdot |f(P)|_2(|f(P)|_2 - r) \quad (2)$$

$$(q, r) = \begin{cases} (q_0, r_0), & \text{если триплет решен} \\ (q_1, r_1), & \text{если триплет не решен} \end{cases} \quad (3)$$

В данном выражении r принимается равным r_0 или r_1 в зависимости от того, решен триплет или нет. Константа q , отвечающая силе притяжения к одной из двух сфер, выбирается, соответственно, равной q_0 или q_1 . Соответственно, если триплет является решенным, то он получает штраф за отдаленность от сферы r_0 . Проанализируем прибавку к градиентам, возникающую вследствие добавления дополнительного слагаемого (4):

$$\begin{aligned} \frac{\partial L}{\partial f(A)} &= (f(N) - f(P)) + q \left(f(A) - \frac{f(A) \cdot r}{|f(A)|_2} \right); \\ \frac{\partial L}{\partial f(P)} &= (f(P) - f(A)) + q \left(f(P) - \frac{f(P) \cdot r}{|f(P)|_2} \right); \\ \frac{\partial L}{\partial f(N)} &= f(A) - f(N). \end{aligned} \quad (4)$$

Можно видеть, что каждый элемент получает дополнительный градиент по направлению к одной из двух гиперсфер в зависимости от того, решен ли триплет, в котором он находится.

Желаемая траектория движения элемента при обучении упрощенно проиллюстрирована на рис. 3. Предлагаемые сферические ограничения изменяют путь векторного представления изображения так, чтобы он не пролегал через устоявшиеся области других классов.

3. Эксперименты

3.1 Данные

Используется набор данных, содержащий в себе изображения различной внутренней природы

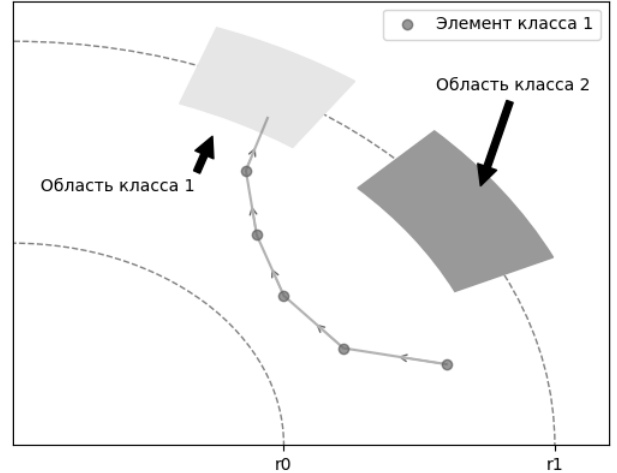


Рис. 3. Траектория движения векторного представления изображения при обучении со сферическими ограничениями



Рис. 4. Пример генерируемого триплета

[10]. Основной массив данных – геометрические структуры и отрывки текста под разными углами. Для формирования обучающей выборки из набора изображений были выбраны 2000 классов с наибольшим числом представителей. Итоговая выборка состоит из 8400 изображений. Размер всех изображений фиксирован – одноканальные изображения размером 32 на 32. Во время обучения используется стратегия генерации случайных триплетов на лету. Пример получаемого триплета изображен на рис. 4. Ввиду малого количества

Табл. 1

Архитектура нейронной сети

#	Слой	Параметры	Функция активации	Размер выхода
	Input	-	-	32x32x1
1	Conv	8 фильтров 4x4, шаг 2x2, без отступа	symrelu[1]	15x15x8
2	Conv	8 фильтров 3x1, шаг 1x1, без отступа	symrelu[1]	13x15x8
3	Conv	8 фильтров 1x3, шаг 1x1, без отступа	symrelu[1]	13x13x8
4	Conv	20 фильтров 3x3, шаг 2x2, без отступа	symrelu[1]	6x6x20
5	Conv	16 фильтров 4x4, шаг 1x1, без отступа	symrelu[1]	6x6x16
6	Conv	12 фильтров 4x4, шаг 1x1, без отступа	symrelu[1]	6x6x12
7	Conv	20 фильтров 4x4, шаг 1x1, без отступа	symrelu[1]	5x5x20
8	Conv	48 фильтров 3x3, шаг 2x2, без отступа	symrelu[1]	2x2x48
9	FC	Выход 128	symrelu[1]	1x1x128
10	FC	Выход d	-	1x1xd

данных, а также для расширения вариативности в ходе обучения к изображениям применяется аугментация *на лету* [5], включающая в себя шум, изменение интенсивности, размытие и проективные искажения.

3.2 Архитектура

В качестве архитектуры выбрана сверточная нейронная сеть. Архитектура нейронной сети аналогична [10] и представлена на табл. 1. В качестве функции активации используется пороговая функция:

$$\text{symrelu}(x) = \min(u, \max(l, x))$$

Размер векторного представления регулируется канальностью последнего слоя.

Выбранная архитектура отличается крайне малым числом параметров – $3.9 \cdot 10^4$. Она была применена в работе [10] для получения векторных представлений изображений.

Для проведения экспериментов используется сеть с размером векторного представления $d = 16$. Также, для возможности визуализации работы сферических ограничений дополнительно используется сеть с размером представления $d = 3$. Рассматривается сеть с триплетной функцией потерь, а также сеть с предложенной модификацией. В каждом из экспериментов был зафиксирован порог $m = 2.25$. Для модифицированной функции потерь также выбраны параметры $r_0 = 10, r_1 = 1, q_0 = 0.1, q_1 = 0.1$. Обучение производится на протяжении 150 эпох методом стохастического градиентного спуска со скоростью обучения и моментом $\text{lr} = 0.0004, \text{momentum} = 0.99$.

3.3 Влияние предложенной модификации на обучение

Рассмотрим влияние выбранного подхода на процесс обучения. Рис. 5 иллюстрирует характеристики сети на тестовом наборе данных в зависимости от эпохи для сети с размером выхода равным 16.

Верхний график на рисунке показывает значение функции ошибки на тестовом множестве для сети с модификацией и без. Можно видеть, что глубина и скорость сходимости при добавлении модификации не изменяются.

Средний график на рис. 5 позволяет увидеть, что часть неразрешенных триплетов, для которых значение функции потерь $L_{triplet}$ оказалось не нулевым, уменьшилось.

Проанализируем нижний график на рисунке. Назовем пару элементов одного класса *далекими*, если расстояние между их векторными представлениями больше $m/2$. Далекие пары увеличивают

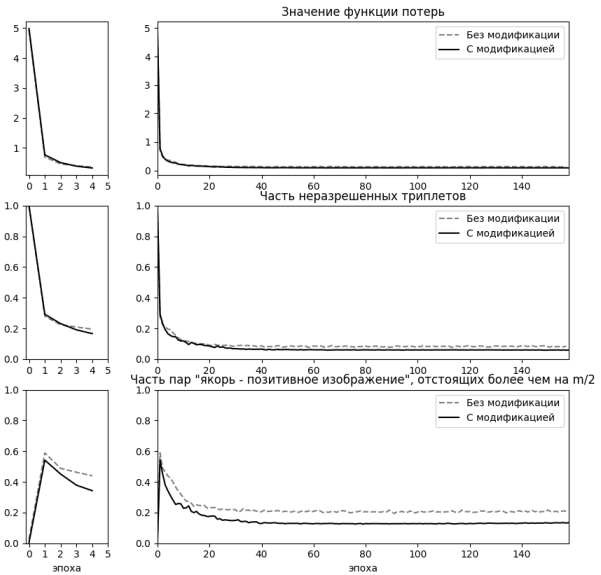


Рис. 5. Характеристики сети с размерностью выхода d=16

внутриклассовую дисперсию. Изначально данный показатель равняется нулю в силу того, что векторные представления всех изображений находятся в небольшой окрестности одной точки. Можно заметить, что при использовании сферических ограничений значения этого показателя оказываются ниже, чем при обучении без модификации. Это значит, что элементы одного класса при таком подходе будут расположены ближе, что положительно скажется на уменьшении внутриклассовой дисперсии.

Итоговые результаты занесены в табл. 2.

Табл. 2

Замеры итоговых метрик		
Показатель	Часть нерешенных триплетов	Часть далеких пар
Без модификации	0.075	0.193
С модификацией	0.055	0.124

Эксперимент показывает, что использование предложенной модификации помогает снизить часть нерешенных триплетов на 27%, а часть далеких пар – на 35%.

3.4 Распределение центроидов

Центроид класса – усреднение векторных представлений всех его представителей с учетом аугментации. Центроиды классов используются при дальнейшей классификации. На Рис. 6 приведены распределения центроидов для каждого из 2000 классов для триплетной функции потерь без модификации (верхняя линия) и с предложенными сферическими ограничениями (нижняя линия). Из графиков можно видеть процесс распро-

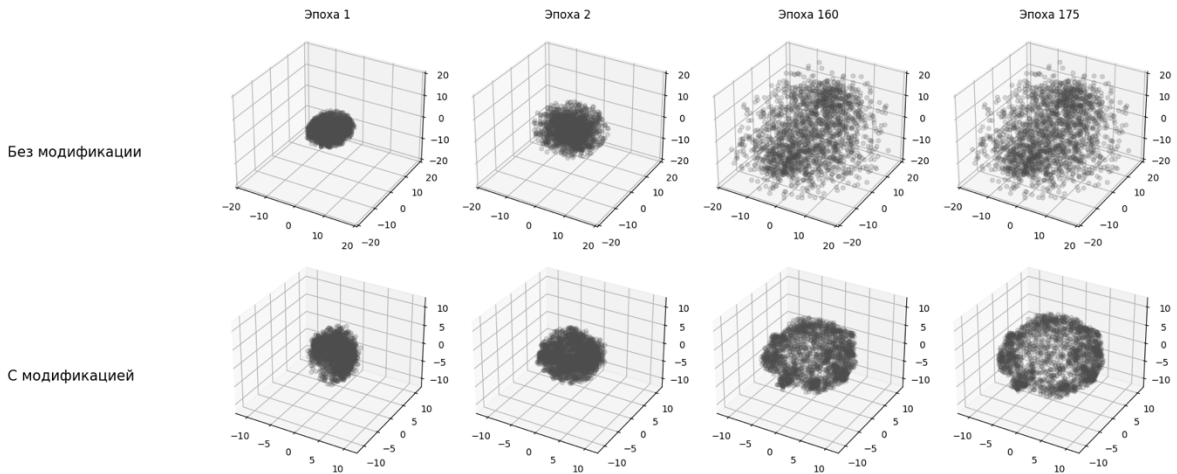


Рис. 6. Распределение центроидов в процессе обучения

странения центроидов по пространству в разные этапы обучения. При обучении со сферическими ограничениями сфера формируется постепенно: первоначально центроиды собираются около сферы меньшего радиуса. Далее в процессе обучения центроиды постепенно движутся по направлению большей сферы в силу того, что элементы соответствующих классов оказываются решенными. Данный эксперимент позволяет подтвердить, что траектории элементов на пути к своему финальному положению пролегают через меньшую сферу, не затрагивая уже сформировавшиеся структуры на большей сфере.

3.5 Траектории движения центроидов

Для подтверждения выдвинутой гипотезы о траекториях, исследуем положения центроидов относительно центра координат в процессе обучения. График рис. 7 иллюстрирует сравнение максимального, минимального и среднего расстояния от центра на каждой эпохе обучения для сети с модификацией и без. Из верхнего графика можно видеть, что при обучении сети без модификации, изменение положений элементов достигается, в основном, за счет отдаления от центра.

При обучении с модификацией (нижний график на рис. 7), простые для решения классы быстро выталкиваются на поверхность сферы. Можно заметить, что при обучении с модификацией, минимальное расстояние имеет гораздо более заметные локальные минимумы. Это связано с тем, что при необходимости переместить элементы некоторого класса, они получают градиент, направленный внутрь, и сдвигаются ближе к центру. Таким образом, элементы занимают свои позиции, не разрушая уже сформировавшиеся структуры для других классов.

3.6 Движение между сферами

Для демонстрации работы предложенного подхода зафиксируем эпоху обучения, когда сфера уже сформирована. Увеличим порог m до 3.5. При таком изменении часть решенных триплетов станут нерешенными, а следовательно, получат градиент по направлению к сфере радиуса r_1 . При дальнейшем обучении сети с таким параметром m часть триплетов снова становится решенными и получает градиент по направлению к сфере r_0 . Через некоторое количество эпох преобладающее количество триплетов будут решены и можно наблюдать исходную картину. Этот процесс продемонстрирован на рис. 8

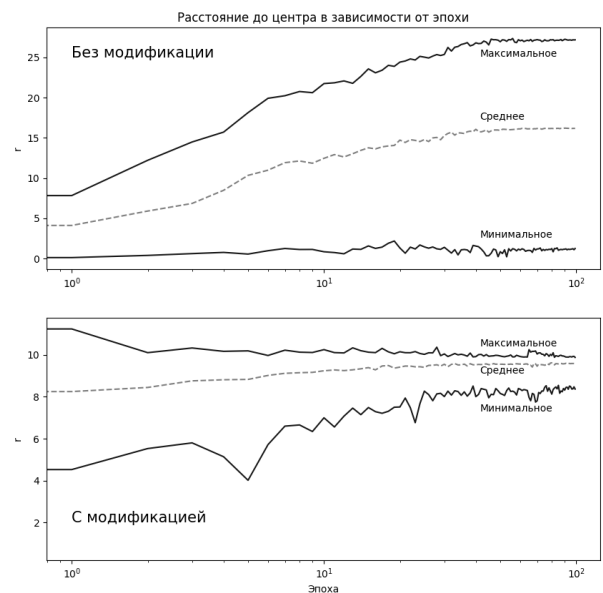


Рис. 7. Зависимость расстояния до центра от этапа обучения

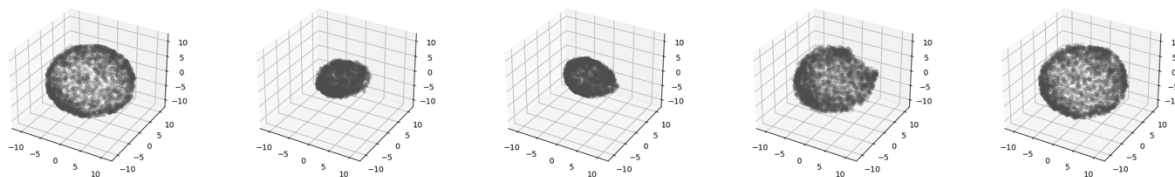


Рис. 8. Формирование малой сферы и ее эволюция

Данный эксперимент позволяет подтвердить выдвинутую ранее гипотезу о траекториях векторных представлений изображений.

Заключение

В данной работе предложен подход по модификации триплетной функции потерь, который корректирует траектории движения векторных представлений изображений в процессе обучения. Данный метод допускает его использование вкупе с другими модификациями, основанными на модификации функции потерь и стратегии генерации триплетов. Замеры показывают, что использование сферических ограничений приводит к уменьшению части нерешенных триплетов, а также к уменьшению части далеких пар изображений.

В дальнейшей работе предполагается исследовать зависимость между параметрами задачи и радиусами выбираемых гиперсфер.

Литература

1. Alnissany Alaa and Yazan Dayoub. *Modified Centroid Triplet Loss for Person Re-Identification*. Research Square Platform LLC, 31 Mar. 2022, <http://dx.doi.org/10.21203/rs.3.rs-1501673/v1>. Accessed 21 Feb. 2023.
2. Bucher Maxime, et al. "Improving Semantic Embedding Consistency by Metric Learning for Zero-Shot Classification." *Computer Vision – ECCV 2016*, Springer International Publishing, 2016, pp. 730–46, http://dx.doi.org/10.1007/978-3-319-46454-1_44. Accessed 9 Mar. 2023.
3. Chen Weihua et al. "Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-Identification." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017, <http://dx.doi.org/10.1109/cvpr.2017.145>. Accessed 22 Feb. 2023.
4. Cui Yin et al. "Fine-Grained Categorization and Dataset Bootstrapping Using Deep Metric Learning with Humans in the Loop." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016. <http://dx.doi.org/10.1109/cvpr.2016.130>. Accessed 27 Feb. 2023.
5. Gayer Alexander et al. "Effective Real-Time Augmentation of Training Dataset for the Neural Networks Learning." *Eleventh International Conference on Machine Vision (ICMV 2018)*, SPIE, 2019, <http://dx.doi.org/10.1117/12.2522969>. Accessed 21 Feb. 2023.
6. Hadsell R. et al. "Dimensionality Reduction by Learning an Invariant Mapping." *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Volume 2 (CVPR'06)*, IEEE, <http://dx.doi.org/10.1109/cvpr.2006.100>. Accessed 24 Mar. 2023.
7. Min Zhu and Zhang Chongyang. "Few-Shot Object Detection via Metric Learning." *Fourteenth International Conference on Machine Vision (ICMV 2021)*, SPIE, 2022, <http://dx.doi.org/10.1117/12.2622909>. Accessed 24 Mar. 2023.
8. Mokin Arseniy et al. "Auto-Clustering Pairs Generation Method for Siamese Neural Networks Training." *Fourteenth International Conference on Machine Vision (ICMV 2021)*, SPIE, 2022, <http://dx.doi.org/10.1117/12.2623139>. Accessed 24 Mar. 2023.
9. Schroff Florian et al. "FaceNet: A Unified Embedding for Face Recognition and Clustering." *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2015, <http://dx.doi.org/10.1109/cvpr.2015.7298682>. Accessed 21 Feb. 2023.
10. Sheshkus A. et al. "Tiny CNN for Feature Point Description for Document Analysis: Approach and Dataset." *Computer Optics*, vol. 46, no. 3, June 2022, <https://doi.org/10.18287/2412-6179-co-1016>.
11. Uzhinskiy A.V. et al. "One-Shot Learning with Triplet Loss for Vegetation Classification Tasks." *Computer Optics*, vol. 45, no. 4, Aug. 2021, <https://doi.org/10.18287/2412-6179-co-856>.
12. Venkataramanan Aishwarya et al. "Tackling Inter-Class Similarity and Intra-Class Variance for Microscopic Image-Based Classification." *Lecture Notes in Computer Science*, Springer International Publishing, 2021, pp. 93–103, http://dx.doi.org/10.1007/978-3-030-87156-7_8. Accessed 21 Feb. 2023.

13. Wang Jiang et al. "Learning Fine-Grained Image Similarity with Deep Ranking." *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2014, <http://dx.doi.org/10.1109/cvpr.2014.180>. Accessed 27 Feb. 2023.
14. Wiecek Mikolaj et al. "On the Unreasonable Effectiveness of Centroids in Image Retrieval." *Neural Information Processing*, Springer International Publishing, 2021, pp. 212–23, http://dx.doi.org/10.1007/978-3-030-92273-3_18. Accessed 22 Feb. 2023.
15. Wu Chao-Yuan et al. "Sampling Matters in Deep Embedding Learning." *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2017, <http://dx.doi.org/10.1109/iccv.2017.309>. Accessed 22 Feb. 2023.
16. Славин О.А. Применение дескрипторов объектов для привязки структурных элементов зашумленных образов деловых документов // ИТиВС. 2022. № 4. С. 13–24. DOI: 10.14357/20718632220402.
17. Hermans Alexander, Lucas Beyer and Bastian Leibe. "In defense of the triplet loss for person re-identification." *arXiv preprint arXiv:1703.07737* (2017).
18. Ha Mai Lan and Volker Blanz. "Deep ranking with adaptive margin triplet loss." *arXiv preprint arXiv:2107.06187* (2021).
19. Xu Albert et al. "Mathematical Justification of Hard Negative Mining via Isometric Approximation Theorem." *arXiv preprint arXiv:2210.11173*. 2022.

Сулоев Константин Константинович. ООО «Smart Engines Service», г. Москва, Россия. Программист-лаборант. Область научных интересов: глубокие нейронные сети, компьютерное зрение, генеративные нейронные сети. E-mail: k.suloev@smartengines.com (Ответственный за переписку).

Шешкус Александр Владимирович. Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Научный сотрудник. Количество печатных работ: более 50. Область научных интересов: глубокие нейронные сети, компьютерное зрение, проективно инвариантная сегментация изображений. E-mail: astdcall@gmail.com

Арлазаров Владимир Львович. Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Заведующий отделением. Чл.–корр. РАН, профессор. Количество печатных работ: более 200. Область научных интересов: теория графов, распознавания образов, программирование. E-mail: arl@isa.ru

Spherical constraints in the triplet loss function

K.K. Suloev^I, A.V. Sheshkus^{I,II}, V.L. Arlazarov^I

^I Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

^{II} Smart Engines Service, LLC, Moscow, Russia

Abstract. Learning with a triplet loss function is one of the most common approaches in metric learning. It finds its application in the tasks of image comparison, identification, coding, etc. However, the triplet loss function has a number of disadvantages that can negatively affect quality, such as the tendency of the network to get stuck in local minima and the formation of trivial triplets. This paper proposes a geometric approach to improve quality based on the introduction of an additional term in the loss function. The trajectory change is achieved by redirecting solved and unsolved images to the surfaces of two concentric hyperspheres of different radii. The use of this method helps to reduce the distances between images of the same class. The proposed method does not prevent the use of other modifications of the loss function. It is experimentally shown that the proposed approach makes it possible to reduce the number of unsolved triplets and the number of distant pairs of images of the same class.

Keywords: artificial neural networks, metric learning.

DOI: 10.14357/20790279230205

References

1. Alnissany Alaa and Yazan Dayoub. Modified Centroid Triplet Loss for Person Re-Identification. Research Square Platform LLC, 31 Mar. 2022, <http://dx.doi.org/10.21203/rs.3.rs-1501673/v1>. Accessed 21 Feb. 2023.
2. Bucher Maxime, et al. "Improving Semantic Embedding Consistency by Metric Learning for Zero-Shot Classification." *Computer Vision – ECCV 2016*, Springer International Publishing, 2016, pp. 730–46.

- http://dx.doi.org/10.1007/978-3-319-46454-1_44. Accessed 9 Mar. 2023.
3. *Chen Weihua et al.* “Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-Identification.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, <http://dx.doi.org/10.1109/cvpr.2017.145>. Accessed 22 Feb. 2023.
 4. *Cui Yin et al.* “Fine-Grained Categorization and Dataset Bootstrapping Using Deep Metric Learning with Humans in the Loop.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, <http://dx.doi.org/10.1109/cvpr.2016.130>. Accessed 27 Feb. 2023.
 5. *Gayer Alexander et al.* “Effective Real-Time Augmentation of Training Dataset for the Neural Networks Learning.” Eleventh International Conference on Machine Vision (ICMV 2018), SPIE, 2019, <http://dx.doi.org/10.1117/12.2522969>. Accessed 21 Feb. 2023.
 6. *Hadsell R. et al.* “Dimensionality Reduction by Learning an Invariant Mapping.” 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Volume 2 (CVPR’06), IEEE, <http://dx.doi.org/10.1109/cvpr.2006.100>. Accessed 24 Mar. 2023.
 7. *Min Zhu and Zhang Chongyang.* “Few-Shot Object Detection via Metric Learning.” Fourteenth International Conference on Machine Vision (ICMV 2021), SPIE, 2022, <http://dx.doi.org/10.1117/12.2622909>. Accessed 24 Mar. 2023.
 8. *Mokin Arseniy et al.* “Auto-Clustering Pairs Generation Method for Siamese Neural Networks Training.” Fourteenth International Conference on Machine Vision (ICMV 2021), SPIE, 2022, <http://dx.doi.org/10.1117/12.2623139>. Accessed 24 Mar. 2023.
 9. *Schroff Florian et al.* “FaceNet: A Unified Embedding for Face Recognition and Clustering.” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015, <http://dx.doi.org/10.1109/cvpr.2015.7298682>. Accessed 21 Feb. 2023.
 10. *Sheshkus A. et al.* “Tiny CNN for Feature Point Description for Document Analysis: Approach and Dataset.” *Computer Optics*, vol. 46, no. 3, June 2022, <https://doi.org/10.18287/2412-6179-co-1016>.
 11. *Uzhinskiy A.V. et al.* “One-Shot Learning with Triplet Loss for Vegetation Classification Tasks.” *Computer Optics*, vol. 45, no. 4, Aug. 2021, <https://doi.org/10.18287/2412-6179-co-856>.
 12. *Venkataramanan Aishwarya et al.* “Tackling Inter-Class Similarity and Intra-Class Variance for Microscopic Image-Based Classification.” *Lecture Notes in Computer Science*, Springer International Publishing, 2021, pp. 93–103, http://dx.doi.org/10.1007/978-3-030-87156-7_8. Accessed 21 Feb. 2023.
 13. *Wang Jiang et al.* “Learning Fine-Grained Image Similarity with Deep Ranking.” 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014, <http://dx.doi.org/10.1109/cvpr.2014.180>. Accessed 27 Feb. 2023.
 14. *Wieczorek Mikołaj et al.* “On the Unreasonable Effectiveness of Centroids in Image Retrieval.” *Neural Information Processing*, Springer International Publishing, 2021, pp. 212–23, http://dx.doi.org/10.1007/978-3-030-92273-3_18. Accessed 22 Feb. 2023.
 15. *Wu Chao-Yuan et al.* “Sampling Matters in Deep Embedding Learning.” 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, <http://dx.doi.org/10.1109/iccv.2017.309>. Accessed 22 Feb. 2023.
 16. *O.A. Slavin.* “Object Descriptors for Linking Structural Elements of Noisy Document Images,” *ITiVS*, no 4, pp. 13-24, 2022, DOI: 10.14357/20718632220402.
 17. *Hermans Alexander, Lucas Beyer and Bastian Leibe.* “In defense of the triplet loss for person re-identification.” *arXiv preprint arXiv:1703.07737* (2017).
 18. *Ha Mai Lan and Volker Blanz.* “Deep ranking with adaptive margin triplet loss.” *arXiv preprint arXiv:2107.06187* (2021).
 19. *Xu Albert et al.* “Mathematical Justification of Hard Negative Mining via Isometric Approximation Theorem.” *arXiv preprint arXiv:2210.11173*. 2022.

Suloev K.K. LLC “Smart Engines Service”, Moscow, Russia. Researcher-Programmer. Scientific interests: deep neural networks, computer vision, generative neural networks. E-mail: k.suloev@smartengines.com

Sheshkus A.V. Researcher, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, 44/2 Vavilova str., Moscow, 119333, Russia. He has more than 50 published papers. His research interests include deep neural networks, computer vision, and projective invariant image segmentation. E-mail: asheshkus@smartengines.com

Arslazarov V.L. Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, 44/2 Vavilova str., Moscow, 119333, Russia. Head of the department, Corresponding member. RAS, Professor. Number of publications: great than 100 papers and books. Scientific interests: artificial intelligence, recognition systems, information technology, programming. E-mail: arl@isa.ru