 © Авторы статьи / Authors of the article

МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ, КОМПЛЕКСОВ И КОМПЬЮТЕРНЫХ СЕТЕЙ/MATHEMATICAL SOFTWARE FOR COMPUTERS, COMPLEXES AND COMPUTER NETWORKS

DOI: https://doi.org/10.60797/IRJ.2026.167.89 EDN: NOQTWI

АНАЛИЗ ВЛИЯНИЯ МОДИФИКАЦИЙ ТРИПЛЕТНОЙ ФУНКЦИИ ПОТЕРЬ НА ВЫЯВЛЕНИЕ ПРИЗНАКОВ КЛАССА

Научная статья

Вольнова Я.С.[1, *], Сущинский А.П.[2]
[1] ORCID : 0009-0005-4377-7981;
[1] Московский государственный технический университет имени Н. Э. Баумана, Москва, Российская Федерация
[2] Группа Компаний «СНС», Москва, Российская Федерация

* Корреспондирующий автор (ypetrova[at]bmstu.ru)

Предложена: 04.03.2026; Принята: 30.03.2026; Опубликована: 18.05.2026

Аннотация

Сверточные нейронные сети имеют склонность считать признаком класса фон, на котором обычно находятся его объекты, что является нежелательным поведением. Статья посвящена анализу различий сверточных нейронных сетей, которые обучены с помощью разных модицификаций триплетной функции потерь, с точки зрения выделяемых признаков классов. Для сравнения использованы оригинальная, фокальная и триплетная функция потерь с гибкой границей, в которой используются дополнительные метки. Различия в выделяемых признаках классов были проанализированы с помощью Grad-CAM, адаптированного для векторных представлений. Для экспериментов использовался датасет продуктов. Результаты показали, что наиболее корректно признаки объектов выделяет модель, обученная с помощью триплетной функции потерь с гибкой границей. Наибольший вклад в выделение корректных признаков вносит фокальный расчет расстояний. При внедрении фокального подсчета расстояний количество изображений, на которых корректно распознаны класс и признаки объекта, увеличивается на 26% относительно оригинальной триплетной функции потерь.

Ключевые слова: триплетная функция потерь, Grad-CAM, признаки, классификация, векторное представление, метрическое обучение, объяснимость.

ANALYSIS OF THE INFLUENCE OF TRIPLET LOSS FUNCTION MODIFICATIONS ON THE CLASS FEATURES DETECTION

Research article

Volnova Y.S.[1, *], Suschinskiy A.P.[2]
[1] ORCID : 0009-0005-4377-7981;
[1] Bauman Moscow State Technical University, Moscow, Russian Federation
[2] SNS - Group of companies, Moscow, Russian Federation

* Corresponding author (ypetrova[at]bmstu.ru)

Suggested: 04.03.2026; Accepted: 30.03.2026; Published: 18.05.2026

Abstract

Convolutional neural networks tend to treat the class background — on which its objects are typically located — as a class feature, which is undesirable behaviour. The article analyses the differences between convolutional neural networks trained using various modifications of the triplet loss function, in terms of the class features they extract. For comparison, the original, focal and triplet loss functions with a flexible boundary, in which additional labels are used, were applied. Differences in the class features extracted were analysed using Grad-CAM, adapted for vector representations. A product dataset was used for the experiments. The results showed that the model trained using the triplet loss function with a flexible boundary identifies object features most accurately. Focal distance calculation makes the greatest contribution to the extraction of correct features. When focal distance calculation is implemented, the number of images in which the class and object features are correctly recognised increases by 26% compared to the original triplet loss function.

Keywords: triplet loss function, Grad-CAM, features, classification, vector representation, metric learning, solvability.

Введение

В связи с расширением вариантов и сфер применения нейронных сетей возрастает потребность в повышении доверия к их решениям. Доверие тесно связано с возможностью объяснения полученного результата. В компьютерном зрении для интерпретации результата нейронных сетей используются тепловые карты. На них выделяются фрагменты изображения, содержащие признаки, которые внесли наибольший вклад в итоговое предсказание. Сопоставив такие фрагменты с полученным результатом, можно приблизиться к пониманию, с чем связаны ошибки модели и какие признаки она выявляет как свойственные классу, то есть влияющие на решение о классификации объекта. К примеру, авторы статей [1], [2] показали, что для сверточных нейронных сетей признаки фона могут превалировать над признаками основного объекта. Такие нейронные сети могут показывать высокую точность при обучении, но обладать

 © Авторы статьи / Authors of the article

низкой обобщающей способностью и в связи с этим плохо адаптироваться к изменяющимся реальным условиям, совершать грубые ошибки.

В исследовании [3] предлагается метод обучения с использованием модифицированной триплетной функции потерь с гибкой границей, который снижает количество грубых ошибок классификации за счет использования дополнительной информации о классах. Данные о классах в виде дополнительных меток используются для коррекции получаемых векторных представлений и должны влиять на то, какие признаки считаются важными. Такой подход может привлечь внимание модели к первичным признакам объекта, а не вторичным, таким как фон. Цель статьи — выявить, какая триплетная функция потерь позволяет получить модель, которая имеет более высокую точность и корректно выделяет признаки класса. Корректными признаками класса считаются первичные признаки принадлежащих ему объектов. Для достижения поставленной цели необходимо провести экспериментальный анализ различных модификаций триплетной функции потерь. Чтобы выяснить, какие признаки модели используют для предсказания, применяется метод интерпретации результатов на основе Grad-CAM.

Методы и принципы исследования

Для выявления областей изображения, где были обнаружены признаки искомого класса, используются различные методы. Для трансформеров зрения — это методы, основанные на анализе результатов работы механизма внимания [4]. Для сверточных нейронных сетей — это методы, анализирующие зависимость значений градиентов сверточных слоев от предсказания модели, такие как Grad-CAM и его разновидности [5]. Также существуют способы интерпретации, не зависящие от архитектуры — это отдельные обучаемые компоненты, встраиваемые в сеть, например T-TAME [6]. Их недостаток состоит в необходимости подбора параметров для конкретной нейросети в процессе обучения, что ограничивает возможности их применения. В данной статье рассматриваются сверточные нейронные сети, поэтому используется вариация Grad-CAM.

Обычно интерпретация результатов классификации предполагает оценку влияния обнаруженных признаков на уверенность модели в принадлежности объекта классу. Чем больше значение, тем больше уверенность. Поэтому, чтобы понять, что повлияло на отнесение объекта к выбранному классу, достаточно разобраться, какие признаки во входных данных увеличивают значение уверенности. Такой подход возможен для моделей, обученных с помощью кросс-энтропии. Однако при использовании метрического обучения задача усложняется. Модель предсказывает не уверенность, а векторное представление (англ. embedding) для каждого объекта, соответственно необходимо адаптировать методы интерпретации результатов. Для Grad-CAM доработки предполагают вычисление косинусного сходства между векторным представлением текущего изображения объекта и некоторым эталонным или усредненным векторным представлением объектов исследуемого класса [7], [8]. Чем выше метрика косинусного сходства между этими векторными представлениями, тем больше уверенность, что объект принадлежит классу. Соответственно для интерпретации результата достаточно обнаружить, какие признаки входных данных увеличивают косинусное сходство.

В данной статье для визуализации областей изображения, на которых были обнаружены свойственные классу признаки, используются идеи из вышеуказанных подходов. *Алгоритм*:

1) для каждого класса подсчитывается эталонное векторное представление как арифметическое среднее представлений всех экземпляров этого класса в обучающем датасете;

2) для тестового изображения вычисляется косинусное сходство между его векторным представлением и эталонным векторным представлением его истинного класса;

3) для интерпретации полученного значения используется Grad-CAM, получающий в качестве входных данных выходы последнего сверточного слоя модели и значение косинусного сходства;

4) результаты Grad-CAM используются для затемнения областей, которые не увеличивают косинусное сходство между векторами тестового изображения и эталона.

Для экспериментов использовалась простая сверточная сеть с размером входа 128x128 пикселей, состоящая из 5 сверточных блоков (64, 64, 128, 256, 512 фильтров соответственно) и двух полносвязных слоев (256 и 128 нейронов соответственно). К последнему слою применена L2-нормализация для ограничения длины получаемых векторных представлений. Для предотвращения переобучения использовалась регуляризация Тихонова. Использование относительно неглубокой архитектуры поощряет модель использовать фон и прочие нерелевантные признаки для распознавания класса, так как авторы статьи [9] показали, что первые слои сверточных нейронных сетей в основном выявляют именно признаки фона. Классификация полученных векторных представлений производилась с помощью k ближайших соседей.

Модель была обучена с помощью оригинальной триплетной функции потерь, фокальной [10] и функции с гибкой границей [3]. Фокальная функция потерь выбрана как промежуточная, так как функция потерь с гибкой границей так же использует фокальный подсчет расстояний. Такой подход позволяет изолированно определить влияние использования дополнительной информации о классах и изменения в подсчете расстояний. Во всех случаях использовалась стратегия подбора полу-сложных триплетов.

Для обучения использован датасет SKU CLASSIFICATION [11], так как он содержит дополнительные данные о классах, качественные разнообразные изображения, и для его визуального анализа достаточно обыденного знания.

*Оценка результатов п*роизводилась по изображениям, на которых хотя бы одна из трех обученных моделей ошиблась. Такие изображения были приняты сложными. К сложным изображениям был применен вышеописанный алгоритм, использующий Grad-CAM. Его целью было получить карту областей изображения, на которых были обнаружены признаки, приближающие векторное представление к эталонному для истинного класса. Вне зависимости от того, к какому классу изображение было отнесено в итоге. В результате для всех изображений с ошибками были получены их версии с затемненными областями для каждой из моделей.

 © Авторы статьи / Authors of the article

Далее для каждого изображения была проведена визуальная оценка областей, которые не были затемнены, а значит, с точки зрения модели, содержат признаки, важные для предсказания истинного класса. Для каждой пары «сложное изображение — модель» было отмечено:

1) соответствует ли истинному предсказанный моделью класс;

2) соответствует ли незатемненная область изображения реальному местоположению объекта целевого класса.

В случае, если для пары выполнялись оба условия, принималось, что модель корректно выявила признаки объекта целевого класса на изображении. На рисунке 1 приведены слева направо:

1) оригинальное изображение,

2) пример корректно выделенных признаков и верно предсказанного класса,

3) пример корректно выделенных признаков и неверно предсказанного класса,

4) пример некорректно выделенных признаков, но верно предсказанного класса.

Рисунок 1 - Примеры корректного и некорректного выделения признаков класса
DOI: https://doi.org/10.60797/IRJ.2026.167.89.1

На рисунке 2 приведены оригинальное изображение и три примера корректного выделения признаков класса.

Рисунок 2 - Примеры корректно выделенных признаков класса
DOI: https://doi.org/10.60797/IRJ.2026.167.89.2

На рисунке 3 приведены оригинальное изображение, пример корректного выделения признаков и двух некорректных.

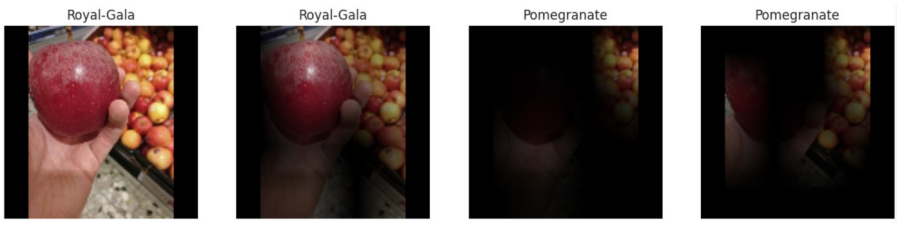

Рисунок 3 - Примеры корректно и некорректно выделенных признаков
DOI: https://doi.org/10.60797/IRJ.2026.167.89.3

На рисунке 4 приведен пример, когда все модели некорректно выделили признаки класса.

 © Авторы статьи / Authors of the article

Рисунок 4 - Примеры некорректно выделенных признаков
DOI: https://doi.org/10.60797/IRJ.2026.167.89.4

Основные результаты

Общее количество сложных изображений — 185.

Таблица 1 - Результаты обучения моделей с помощью триплетных функций потерь
DOI: https://doi.org/10.60797/IRJ.2026.167.89.5

Функция потерь, с помощью которой обучена модель	macro F1 итоговой модели	Общее количество ошибок модели	Количество сложных изображений, на которых корректно распознаны класс и признаки объекта
Триплетная функция потерь	0,86	118	50
Фокальная триплетная функция потерь	0,87	110	63
Триплетная функция потерь с гибкой границей	**0,89**	**98**	**65**

Полученные в таблице 1 результаты показывают, что модель, обученная с помощью триплетной функции потерь с гибкой границей имеет наименьшее количество ошибок и наиболее высокую точность. Разница в общем количестве ошибок между моделями, обученными с помощью модифицированной и оригинальной триплетной функциями потерь составила 17%. При этом количество сложных изображений, на которых были корректно распознаны класс и признаки объекта, выросло на 30% относительно оригинальной триплетной функции потерь и на 3% относительно фокальной триплетной функции потерь. Доля корректно распознанных сложных изображений (как с точки зрения класса, так и с точки зрения его признаков) выросла с 27% для оригинальной триплетной функции потерь до 35% для функции потерь с гибкой границей.

Обсуждение

Полученные результаты показывают, что учет дополнительных меток класса в триплетной функции потерь с гибкой границей существенно влияет на увеличение итоговой точности и сокращение количества ошибок. При этом на выявление признаков класса оказывает наибольшоее влияние именно фокальный подсчет расстояний, введенный в фокальной триплетной функции потерь. Введение дополнительных меток в функции потерь с гибкой границей усиливает этот эффект, но незначительно. Наиболее высокую точность классификации и корректное выделение признаков класса удалось получить при обучении с помощью триплетной функции потерь с гибкой границей. Исследование опирается на Grad-CAM, поэтому приводимые выводы могут быть дополнены при использовании других методов интерпретации предсказаний сверточных нейронных сетей.

Заключение

В результате проведенного эксперимента по обучению моделей с разными модификациями триплетной функции потерь было выяснено, что на корректное выделение признаков оказывает существенное влияние фокальный подсчет расстояний. При внедрении фокального подсчета расстояний количество изображений, на которых корректно распознан класс и его признаки, увеличивается на 26% относительно оригинальной триплетной функции потерь. При этом добавление дополнительной информации о классах в триплетную функцию потерь с гибкой границей увеличивает точность итоговой модели и снижает общее количество ошибок (на 17% относительно оригинальной триплетной функции потерь), но не оказывает существенного влияния на выделение признаков с помощью Grad-CAM. Количество изображений, на которых корректно распознан класс его признаки, увеличилось на 3% в сравнении с моделью, обученной посредством фокальной триплетной функцией потерь.

Дальнейшие исследования могут быть направлены на доработку триплетной функции потерь с гибкой границей с целью усиления влияния дополнительных меток на точность итоговой модели и интерпретируемость получаемых векторных представлений. Также на развитие самих методов интерпретации результатов метрических моделей, к примеру, использование преобразования изображения для выявления признаков, на которые опирается модель [12].

 © Авторы статьи / Authors of the article

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Moayeri M. A comprehensive study of image classification model sensitivity to foregrounds, backgrounds, and visual attributes / M. Moayeri, P. Pope, Y. Balaji et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; — New Orleans: IEEE, 2022. — P. 19087–19097.

2. Xiao K. Noise or signal: The role of image backgrounds in object recognition / K. Xiao, L. Engstrom, A. Ilyas et al. // *arXiv preprint arXiv:2006.09994*. — 2020. — URL: https://arxiv.org/abs/2006.09994. (дата обращения: 04.03.26) doi: 10.48550/arXiv.2006.09994

3. Петрова Я.С. Методика обучения классификаторов изображений с использованием дополнительных меток / Я.С. Петрова // Моделирование, оптимизация и информационные технологии. — 2025. — 13 (2). — С. 1–13. — DOI: 10.26102/2310-6018/2025.49.2.041

4. Ayyar M.P. More to Attention: Statistical Filtering Enhances Explanations in Vision Transformers / M.P. Ayyar, J. Benois-Pineau, A. Zemmari // *arXiv preprint arXiv:2510.06070*. — 2025. — URL: https://arxiv.org/abs/2510.06070. (дата обращения: 04.03.26) doi: 10.48550/arXiv.2510.06070

5. Chattopadhay A. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks / A. Chattopadhay, A. Sarkar, P. Howlader et al. // 2018 IEEE winter conference on applications of computer vision (WACV); — New Orleans: IEEE, 2018. — P. 839–847.

6. Ntrougkas M.V. T-TAME: trainable attention mechanism for explaining convolutional networks and vision transformers / M.V. Ntrougkas, N. Gkalelis, V. Mezaris. // IEEE Access; — 12. — New Orleans: IEEE, 2024. — P. 76880–76900.

7. Chen L. Adapting grad-cam for embedding networks / L. Chen, J. Chen, H. Hajimirsadeghi et al. // Proceedings of the IEEE/CVF winter conference on applications of computer vision; — New Orleans: IEEE, 2020. — P. 2794–2803.

8. Zhu S. Visual explanation for deep metric learning / S. Zhu, T. Yang, C. Chen. // IEEE Transactions on Image Processing; — 30. — New Orleans: IEEE, 2021. — P. 7593–7607.

9. Loke J. Human Visual Cortex and Deep Convolutional Neural Network Care Deeply about Object Background / J. Loke, N. Seijdel, L. Snoek et al. // Journal of Cognitive Neuroscience. — 2024. — 36(3). — P. 551–566.

10. Zhang S. Person Re-Identification With Triplet Focal Loss / S. Zhang, Q. Zhang, X. Wei et al. // IEEE Access. — 2018. — 6. — P. 78092–78099.

11. siva SKUCLASSIFICATION Dataset / siva // Roboflow Universe. — 2024. — URL: https://universe.roboflow.com/siva-4or6j/skuclassification. (дата обращения: 04.03.26)

12. Erukude S.T. Identifying bias in deep neural networks using image transforms / S.T. Erukude, A. Joshi, L. Shamir // Computers. — 2024. — 13(12). — P. 341.

Список литературы на английском языке / References in English

1. Moayeri M. A comprehensive study of image classification model sensitivity to foregrounds, backgrounds, and visual attributes / M. Moayeri, P. Pope, Y. Balaji et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; — New Orleans: IEEE, 2022. — P. 19087–19097.

2. Xiao K. Noise or signal: The role of image backgrounds in object recognition / K. Xiao, L. Engstrom, A. Ilyas et al. // *arXiv preprint arXiv:2006.09994*. — 2020. — URL: https://arxiv.org/abs/2006.09994. (accessed: 04.03.26) doi: 10.48550/arXiv.2006.09994

3. Petrova Ya.S. Metodika obucheniya klassifikatorov izobrazhenij s ispol'zovaniem dopolnitel'ny'x metok [Method of training image classifiers using additional labels] / Ya.S. Petrova // Modeling, Optimization and Information Technology. — 2025. — 13 (2). — P. 1–13. — DOI: 10.26102/2310-6018/2025.49.2.041 [in Russian]

4. Ayyar M.P. More to Attention: Statistical Filtering Enhances Explanations in Vision Transformers / M.P. Ayyar, J. Benois-Pineau, A. Zemmari // *arXiv preprint arXiv:2510.06070*. — 2025. — URL: https://arxiv.org/abs/2510.06070. (accessed: 04.03.26) doi: 10.48550/arXiv.2510.06070

5. Chattopadhay A. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks / A. Chattopadhay, A. Sarkar, P. Howlader et al. // 2018 IEEE winter conference on applications of computer vision (WACV); — New Orleans: IEEE, 2018. — P. 839–847.

6. Ntrougkas M.V. T-TAME: trainable attention mechanism for explaining convolutional networks and vision transformers / M.V. Ntrougkas, N. Gkalelis, V. Mezaris. // IEEE Access; — 12. — New Orleans: IEEE, 2024. — P. 76880–76900.

7. Chen L. Adapting grad-cam for embedding networks / L. Chen, J. Chen, H. Hajimirsadeghi et al. // Proceedings of the IEEE/CVF winter conference on applications of computer vision; — New Orleans: IEEE, 2020. — P. 2794–2803.

8. Zhu S. Visual explanation for deep metric learning / S. Zhu, T. Yang, C. Chen. // IEEE Transactions on Image Processing; — 30. — New Orleans: IEEE, 2021. — P. 7593–7607.

 © Авторы статьи / Authors of the article

9. Loke J. Human Visual Cortex and Deep Convolutional Neural Network Care Deeply about Object Background / J. Loke, N. Seijdel, L. Snoek et al. // Journal of Cognitive Neuroscience. — 2024. — 36(3). — P. 551–566.

10. Zhang S. Person Re-Identification With Triplet Focal Loss / S. Zhang, Q. Zhang, X. Wei et al. // IEEE Access. — 2018. — 6. — P. 78092–78099.

11. siva SKUCLASSIFICATION Dataset / siva // Roboflow Universe. — 2024. — URL: https://universe.roboflow.com/siva-4or6j/skuclassification. (accessed: 04.03.26)

12. Erukude S.T. Identifying bias in deep neural networks using image transforms / S.T. Erukude, A. Joshi, L. Shamir // Computers. — 2024. — 13(12). — P. 341.