

ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ/INFORMATICS AND INFORMATION PROCESSES

DOI: <https://doi.org/10.60797/IRJ.2026.172.12> EDN: WZOCMAПРИМЕНЕНИЕ SELF-SUPERVISED ПРЕДОБУЧЕНИЯ SIMCLR ДЛЯ СНИЖЕНИЯ ЗАВИСИМОСТИ
КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ ОТ ОБЪЕМА РУЧНОЙ РАЗМЕТКИ

Научная статья

Мельников Н.С.^{1,*}¹ ORCID : 0009-0000-6138-2637;¹ Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, Екатеринбург, Российская Федерация

* Корреспондирующий автор (nikitamelnikov502[at]gmail.com)

Предложена: 26.07.2026; Принята: 14.10.2026; Опубликовано: 16.10.2026

Аннотация

Исследуется возможность уменьшения объёма ручной разметки изображений за счёт self-supervised предобучения визуального backbone методом SimCLR. Предобучение выполнено на неразмеченной части датасета STL-10, после чего параметры backbone перенесены в классификационную модель YOLOv8n-cls. Классификатор дообучался на 5, 10, 30 и 50% размеченных данных и сравнивался с идентичной по архитектуре моделью, обученной с нуля. Для каждого бюджета разметки проведено по три запуска с начальными состояниями 42, 43 и 44; качество оценивалось по Top-1 Accuracy, Macro-F1 и Balanced Accuracy. Во всех 12 попарных сравнениях SimCLR + fine-tuning превзошёл обучение с нуля. Средний прирост Top-1 Accuracy составил 24,91; 25,98; 22,93 и 17,52 процентного пункта для 5, 10, 30 и 50% разметки соответственно. Модель с SimCLR, дообученная на 5% размеченных данных, достигла Accuracy 58,65%, превысив результат scratch-модели на 50% данных (56,87%). Это показывает, что в пределах достигнутого уровня качества предварительное обучение позволило сократить используемый бюджет разметки в десять раз. Эффект сохранялся на всех исследованных долях, однако уменьшался по мере роста объёма размеченных данных.

Ключевые слова: self-supervised learning, SimCLR, контрастивное обучение, классификация изображений, ограниченная разметка, transfer learning, YOLOv8, STL-10.

THE APPLICATION OF SELF-SUPERVISED SIMCLR PRE-TRAINING TO REDUCE THE DEPENDENCE OF
IMAGE CLASSIFICATION ON THE VOLUME OF MANUAL LABELLING

Research article

Melnikov N.S.^{1,*}¹ ORCID : 0009-0000-6138-2637;¹ Ural Federal University named after the first President of Russia B.N. Yeltsin, Ekaterinburg, Russian Federation

* Corresponding author (nikitamelnikov502[at]gmail.com)

Suggested: 26.07.2026; Accepted: 14.10.2026; Published: 16.10.2026

Abstract

The possibility of reducing the amount of manual image labeling through self-supervised pre-training of a visual backbone using the SimCLR method is studied. Pre-training was carried out on an unlabelled portion of the STL-10 dataset, after which the backbone parameters were transferred to a YOLOv8n-cls classification model. The classifier was fine-tuned on 5, 10, 30 and 50% of the labelled data and compared with an architecturally identical model trained from scratch. For each labelling budget, three runs were conducted with initial states of 42, 43 and 44; quality was assessed using Top-1 Accuracy, Macro-F1 and Balanced Accuracy. In all 12 head-to-head comparisons, SimCLR + fine-tuning outperformed training from scratch. The average increase in Top-1 Accuracy was 24.91; 25.98; 22.93 and 17.52 percentage points for 5%, 10%, 30% and 50% of the labelled data, respectively. The model with SimCLR, fine-tuned on 5% of the labelled data, achieved an accuracy of 58.65%, exceeding the result of the scratch model on 50% of the data (56.87%). This demonstrates that, within the achieved quality range, pre-training enabled a tenfold reduction in the annotation budget required. The effect persisted across all proportions studied, although it diminished as the volume of annotated data increased.

Keywords: self-supervised learning, SimCLR, contrastive learning, image classification, limited annotation, transfer learning, YOLOv8, STL-10.

Введение

Современные модели компьютерного зрения способны достигать высокого качества классификации, однако их обучение обычно требует значительного объёма размеченных данных. Подготовка таких данных включает отбор изображений, назначение классов, проверку качества аннотаций и разрешение спорных случаев. Даже в задаче классификации эта работа создаёт заметные временные и финансовые затраты, а для специализированных предметных областей дополнительно требует участия экспертов. Self-supervised learning позволяет использовать изображения без ручных меток для предварительного обучения представлений. Модель получает обучающий сигнал не из классов, назначенных человеком, а из структуры самих данных. В контрастивных методах две аугментированные версии одного изображения рассматриваются как положительная пара, а представления различных изображений разделяются в

пространстве признаков. Одним из базовых и широко используемых подходов является SimCLR, предложенный Т. Ченом и соавторами [1]. Его ключевыми элементами являются сильные аугментации, encoder, проекционная голова и контрастивная функция потерь NT-Xent. Результаты SimCLR, MoCo и последующих методов показали, что self-supervised представления могут успешно переноситься на downstream-задачи [1], [2], [3], [4]. Практический интерес представляет не только абсолютное качество модели, но и её label efficiency: какой объём размеченных данных требуется для достижения заданного уровня качества. Именно эта постановка особенно важна для организаций, располагающих большим количеством необработанных изображений, но ограниченным бюджетом ручной разметки. Цель работы — определить, снижает ли предобучение SimCLR зависимость классификатора изображений от объёма ручной разметки. Для этого сравниваются две схемы: обучение YOLOv8n-cls с нуля и перенос SimCLR-предобученного backbone с последующим finetuning. Эксперимент проводится на нескольких бюджетах разметки и повторяется для трёх начальных состояний генераторов случайных чисел.

Основные задачи исследования:

- 1) реализовать предобучение backbone методом SimCLR на неразмеченных изображениях;
- 2) корректно перенести веса в классификационную архитектуру и переинициализировать выходную голову;
- 3) сравнить SimCLR + fine-tuning со scratch при одинаковых условиях;
- 4) оценить изменение преимущества SimCLR при росте доли размеченных данных;
- 5) определить, позволяет ли подход достичь качества baseline при меньшем бюджете разметки.

Методы и принципы исследования

2.1. Данные и экспериментальные доли

В качестве набора данных выбран STL-10 [5], предназначенный для исследований в области unsupervised и self-supervised обучения. Датасет содержит изображения десяти классов и отдельную крупную неразмеченную часть. Неразмеченные изображения использовались только для self-supervised предобучения, а размеченная часть — для обучения и проверки классификаторов. Для supervised-этапа были сформированы подмножества, соответствующие 5, 10, 30 и 50% доступных размеченных данных. Эти значения выбраны для анализа режима сильного дефицита меток, промежуточного режима и ситуации, когда доступна половина обучающей выборки. Разбиение данных было одинаковым для сравниваемых методов внутри каждого seed.

2.2. Предобучение SimCLR

На этапе SimCLR для каждого исходного изображения формировались две независимые аугментированные версии. В пайплайн входили RandomResizedCrop, ColorJitter, GaussianBlur и другие преобразования, изменяющие внешний вид изображения при сохранении его семантики. Аугментированные изображения проходили через общий backbone и проекционную голову. Для пары представлений z_i и z_j минимизировалась нормализованная температурно-масштабированная кросс-энтропийная потеря:

$$\mathcal{L}_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k \neq i} \exp(\text{sim}(z_i, z_k)/\tau)},$$

где $\text{sim}(\cdot)$ — косинусная близость, а τ — температурный параметр. Такая оптимизация сближает представления двух видов одного изображения и разделяет представления различных изображений. Предобучение выполнялось 100 эпох с seed 42. Итоговое значение функции потерь на обучающей выборке составило 2,3634, а продолжительность этапа — 6041 с, то есть около 1 ч 41 мин.

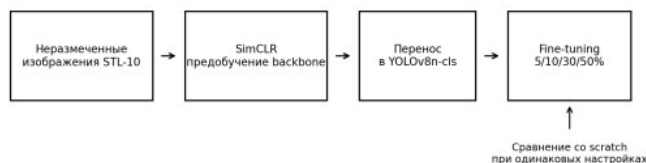


Рисунок 1 - Схема экспериментального пайплайна

DOI: <https://doi.org/10.60797/IRJ.2026.172.12.1>

2.3. Перенос backbone и supervised-обучение

Базовой моделью являлась YOLOv8n-cls из библиотеки Ultralytics [6]. Число параметров модели составляло 1 110 618. После SimCLR-предобучения проекционная голова удалялась, веса backbone переносились в классификатор, а классификационная голова переинициализировалась. Корректность переноса проверялась автоматически для каждого запуска. Все 12 отчётов о переносе получили статус passed; доля совпавших параметров backbone составила 1,0, классификационная голова была переинициализирована, а неполная загрузка с strict=False не использовалась. Таким образом, сравнение отражает различие именно в инициализации backbone, а не в архитектуре или числе параметров. Для каждого бюджета разметки обучались два варианта:

- scratch — YOLOv8n-cls со случайной инициализацией;
- SimCLR + fine-tuning — та же модель с перенесённым self-supervised backbone. Эксперименты проведены для seed 42, 43 и 44. Внутри каждой пары сохранялись одинаковые split, размер входного изображения, batch size, число эпох и остальные параметры supervised-обучения. Всего завершено 24 запуска: 12 scratch и 12 SimCLR + fine-tuning.

2.4. Метрики и агрегация

Основной метрикой являлась Top-1 Accurasy. Дополнительно рассчитывались Macro-F1 и Balanced Accurasy. Macro-F1 усредняет F1 по классам и позволяет проверить, не достигается ли прирост только за счёт отдельных категорий. Balanced Accurasy рассчитывает среднюю полноту по классам и полезна при потенциальном дисбалансе распределения. Для каждой комбинации метода и бюджета разметки вычислялись среднее значение и выборочное стандартное отклонение по трём seed. Дополнительно использовались попарные разности между SimCLR и scratch при одинаковых seed и доле разметки. Из-за малого числа повторов результаты трактуются прежде всего как оценка устойчивости направления и масштаба эффекта, а не как исчерпывающий статистический тест.

Таблица 1 - Средние метрики по трём seed

DOI: <https://doi.org/10.60797/IRJ.2026.172.12.2>

Разметка, %	Scratch Acc.	SimCLR Acc.	Δ Acc.
5	33,74 $\pm$ 0,91	58,65 $\pm$ 0,81	24,91
10	39,05 $\pm$ 0,45	65,03 $\pm$ 1,22	25,98
30	48,58 $\pm$ 1,29	71,51 $\pm$ 1,54	22,93
50	56,87 $\pm$ 0,76	74,39 $\pm$ 2,32	17,52

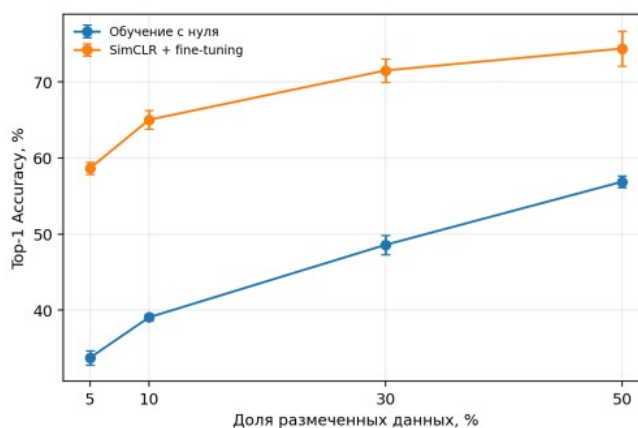


Рисунок 2 - Зависимость Top-1 Accurasy от доли размеченных данных

DOI: <https://doi.org/10.60797/IRJ.2026.172.12.3>

Примечание: показаны среднее и стандартное отклонение по трём seed

Основные результаты

3.1. Качество классификации

Средние результаты представлены в табл. 1. На всех четырёх бюджетах разметки SimCLR + fine-tuning существенно превосходит обучение с нуля. При 5% разметки средняя Accurasy возрастает с 33,74 до 58,65%, при 10% — с 39,05 до 65,03%, при 30% — с 48,58 до 71,51%, а при 50% — с 56,87 до 74,39%. Значения Macro-F1 согласуются с Accurasy. Для scratch они составили 33,25; 38,83; 48,30 и 56,68%, а для SimCLR — 58,27; 64,89; 71,38 и 74,36%. Прирост Macro-F1 находился в диапазоне от 17,68 до 26,06 п.п. Balanced Accurasy практически совпала с обычной Accurasy, что ожидаемо для сбалансированной десятиклассовой постановки и дополнительно подтверждает отсутствие выигрыша только на отдельных классах. На рис. 2 видно, что кривые обоих методов возрастают при увеличении разметки, однако SimCLR сохраняет значительный отрыв на всём исследованном диапазоне. Наибольший средний абсолютный выигрыш наблюдается при 10% разметки: 25,98 п.п. При 50% он уменьшается до 17,52 п.п., но остаётся практически значимым.

3.2. Устойчивость попарного преимущества

Во всех 12 сравнениях при одинаковых seed результат SimCLR оказался выше scratch. Минимальный попарный прирост Accurasy составил 15,33 п.п., максимальный — 27,31 п.п. При 5% разметки прирост по отдельным seed находился в диапазоне 23,68–25,66 п.п.; при 10% — 24,22–27,31 п.п.; при 30% — 22,29–23,38 п.п.; при 50% — 15,33–20,90 п.п. Стандартное отклонение Accurasy для scratch не превышало 1,29 п.п., а для SimCLR — 2,32 п.п. Наибольшая вариативность SimCLR наблюдалась при 50% данных: результаты отдельных запусков составили 73,85; 76,94 и 72,39%. Даже наименее успешный SimCLR-запуск в этом режиме превосходил лучший scratch-запуск более чем на 14 п.п.

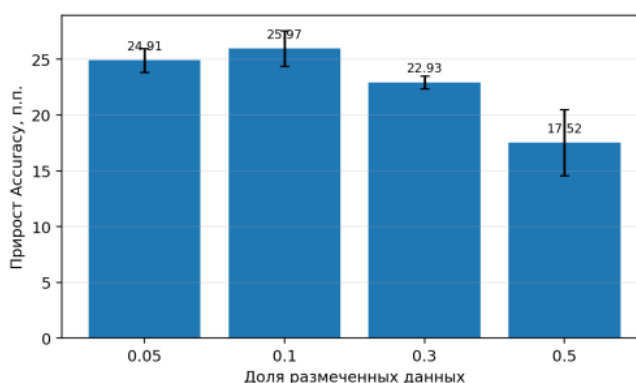


Рисунок 3 - Средний прирост Accuracy от SimCLR относительно scratch

DOI: <https://doi.org/10.60797/IRJ.2026.172.12.4>

3.3. Снижение требуемого бюджета разметки

Ключевое наблюдение состоит в сравнении разных бюджетов. При использовании только 5% размеченных данных SimCLR + fine-tuning достиг Accuracy 58,65%. Scratch-модель, обученная на 50% данных, показала 56,87%. Разница составила 1,78 п.п. в пользу SimCLR, несмотря на десятикратное различие в доле размеченной выборки. Следовательно, для достижения качества около 57% в рамках проведённого эксперимента self-supervised предобучение позволило перейти от 50 к 5% разметки. Это соответствует десятикратному уменьшению используемого бюджета, или сокращению на 90% относительно baseline бюджета. Данный вывод относится именно к указанному уровню качества и исследованным долям; он не означает универсального десятикратного сокращения для любого целевого значения Accuracy.

3.4. Вычислительные затраты

Среднее wall-clock время одного supervised-запуска составляло около 6,1 мин как для scratch, так и для SimCLR + fine-tuning. Отдельные различия времени объясняются выбором лучшей эпохи и вариациями выполнения, но систематического роста стоимости fine-tuning относительно scratch не наблюдалось. Основная дополнительная стоимость связана с однократным SimCLR-предобучением, занявшим около 101 мин. После получения backbone он может применяться повторно для нескольких бюджетов разметки, seed или близких downstream-задач. Поэтому на практике вычислительная стоимость предобучения может распределяться между серией последующих экспериментов, тогда как экономия ручной разметки сохраняется для каждого из них.

Обсуждение

Полученные результаты подтверждают исходную гипотезу: self-supervised предобучение уменьшает зависимость классификатора от количества ручных меток. Наиболее заметный эффект наблюдается в low-label режиме.

При 5 и 10% данных scratch-модель ещё не формирует достаточно устойчивые представления, тогда как SimCLRbackbone уже содержит признаки, извлечённые из неразмеченных изображений. Supervised-этапу остаётся адаптировать их к десятиклассовой задаче и настроить классификационную голову. Уменьшение абсолютного выигрыша с 25,98 п.п. при 10% до 17,52 п.п. при 50% соответствует ожидаемой границе применимости метода. По мере роста размеченной выборки supervised-модель получает больше прямого сигнала и частично компенсирует отсутствие предобучения. При этом эффект не исчезает: даже на максимальном исследованном бюджете SimCLR сохраняет значительный отрыв. Важно отличать абсолютный прирост качества от относительного. Например, увеличение Accuracy с 33,74 до 58,65% составляет 24,91 процентного пункта, но относительное улучшение по отношению к scratch равно примерно 73,8%. Для научной интерпретации в работе используются прежде всего процентные пункты, поскольку они непосредственно отражают расстояние между значениями одной метрики. Согласованность Accuracy и Macro-F1 показывает, что улучшение не ограничено несколькими классами. Полное совпадение направления эффекта в 12 попарных сравнениях также снижает вероятность того, что итог объясняется одним удачным seed. Вместе с автоматической проверкой полного переноса backbone это делает эксперимент технически воспроизводимым и внутренне согласованным. С практической точки зрения SimCLR особенно оправдан, когда накоплен большой массив неразмеченных изображений, а ручная классификация дорога или медленна. Подход позволяет перенести часть затрат с аннотации на вычисления. Такой обмен выгоден, если предобученный backbone используется многократно либо если стоимость экспертной разметки значительно превышает стоимость вычислительных ресурсов.

Ограничения исследования

Работа имеет несколько ограничений. Во-первых, использованы три seed. Этого достаточно для первичной проверки устойчивости, однако для более строгой статистической оценки желательно увеличить число повторов. Во-вторых, эксперимент проведён на одном датасете и одной компактной архитектуре, поэтому переносимость выводов на другие домены требует дополнительной проверки. В-третьих, в supervised-сравнение включены доли до 50%. Отсутствие scratch-результатов на 75 и 100% не позволяет утверждать, что 5% SimCLR эквивалентны полному объёму разметки. Корректный вывод ограничивается сравнением 5% SimCLR и 50% scratch при качестве около 57%. В-четвёртых, baseline представлен обучением с нуля. Для более полного исследования необходимо добавить стандартное



supervised-предобучение, например ImageNet initialization, а также другие self-supervised методы. Это позволит отделить общий эффект предварительного обучения от специфического вклада SimCLR. Наконец, работа оценивает качество классификации, но не переводит экономию разметки в денежные показатели. Для прикладного внедрения следует измерять фактическое время аннотаторов, стоимость проверки меток и стоимость вычислений.

Заключение

В работе исследовано применение SimCLR для снижения зависимости классификации изображений от объёма ручной разметки. Предобученный на неразмеченной части STL-10 backbone был перенесён в YOLOv8n-cls и дообучен на 5, 10, 30 и 50% размеченных данных. Сравнение выполнено с моделью той же архитектуры, обученной с нуля, при трёх одинаковых seed. SimCLR + fine-tuning превзошёл scratch во всех 12 попарных сравнениях. Средний прирост Accurasy составил от 17,52 до 25,98 п.п., а Macro-F1 — от 17,68 до 26,06 п.п. Наиболее важный результат получен при межбюджетном сравнении: 5% разметки с SimCLR дали 58,65% Accurasy, тогда как 50% разметки при обучении с нуля — 56,87%. Таким образом, в пределах достигнутого уровня качества self-supervised предобучение позволило уменьшить используемый объём размеченных данных в десять раз. Результаты подтверждают целесообразность SimCLR в условиях дефицита меток и показывают, что его преимущество постепенно уменьшается, но сохраняется при увеличении supervised-бюджета.

Благодарности

Исследование выполнено в рамках производственной практики в ООО «ЛамАрт» по образовательной программе УрФУ 09.03.03 «Прикладная информатика».

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.

Acknowledgement

The research was carried out as part of a work placement at 'LamArt' Ltd under the Ural Federal University's 09.03.03 educational programme in 'Applied Computer Science'.

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 на английском языке / References in English

1. Chen T. A Simple Framework for Contrastive Learning of Visual Representations / T. Chen, G. Hinton, M. Norouzi et al. // Proceedings of the 37th International Conference on Machine Learning. — PMLR, 2020. — P. 1597–1607.
2. He K. Momentum Contrast for Unsupervised Visual Representation Learning / K. He, H. Fan, Y. Wu et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; — Los Alamitos: IEEE, 2020. — P. 9729–9738. doi: 10.1109/CVPR42600.2020.00975
3. Coates A. An Analysis of Single-Layer Networks in Unsupervised Feature Learning / A. Coates, A. Ng, H. Lee. // Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics; — Fort Lauderdale: PMLR, 2011. — P. 215–223.
4. Chen T. Big Self-Supervised Models are Strong Semi-Supervised Learners / T. Chen, S. Kornblith, K. Swersky et al. // Advances in Neural Information Processing Systems; — Red Hook: Curran Associates, 2020. — P. 22243–22255.
5. Wang T. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere / T. Wang, P. Isola // Proceedings of the 37th International Conference on Machine Learning; — PMLR, 2020. — P. 9929–9939.
6. Henaff O. Data-Efficient Image Recognition with Contrastive Predictive Coding / O. Henaff // Proceedings of the 37th International Conference on Machine Learning; — PMLR, 2020. — P. 4182–4192.
7. Zbontar J. Barlow Twins: Self-Supervised Learning via Redundancy Reduction / J. Zbontar, L. Jing, I. Misra et al. // Proceedings of the 38th International Conference on Machine Learning; — PMLR, 2021. — P. 12310–12320.
8. Wu Z. Unsupervised Feature Learning via Non-Parametric Instance Discrimination / Z. Wu, Y. Xiong, S.X. Yu et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; — Salt Lake City: IEEE, 2018. — P. 3733–3742. — DOI: 10.1109/CVPR.2018.00393
9. Caron M. Deep Clustering for Unsupervised Learning of Visual Features / M. Caron, P. Bojanowski, A. Joulin et al. // Computer Vision – ECCV 2018; — Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-030-01264-9_9
10. Noroozi M. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles / M. Noroozi, P. Favaro. // Computer Vision – ECCV 2016; — Cham: Springer International Publishing, 2016. — P. 69–84. doi: 10.1007/978-3-319-46466-4_5