

УДК 004.932

Повышение точности реидентификации людей в системах видеонаблюдения за счет выбора гиперпараметров при обучении сверточных нейронных сетей и повторного ранжирования нейросетевых признаков

С. А. Игнатьева, Р. П. Богуш

Полоцкий государственный университет имени Евфросинии Полоцкой,
Новополоцк, Беларусь

С. В. Абламейко

Белорусский государственный университет,
Минск

Введение

Результативность реидентификации людей на основе сверточных нейронных сетей (СНС) определяется не только архитектурой СНС, но и рядом других факторов, включая метод обучения, установленные гиперпараметры, состав тренировочной выборки, а также алгоритм сравнения нейросетевых признаков [1]. Таким образом, подходы для повышения точности реидентификации людей можно разделить на две группы: применяемые на этапе обучения СНС и на этапе постобработки извлеченных признаков. В настоящей работе показана возможность повышения точности реидентификации людей за счет выбора гиперпараметров при тренировке СНС и повторного ранжирования полученных нейросетевых признаков.

Выбор гиперпараметров при обучении сверточных нейронных сетей

В качестве гиперпараметров используются установленные перед началом тренировки СНС значения количества эпох (итераций обновления весов), скорости обучения, размера пакета. Данные значения в значительной мере определяют результативность работы СНС и время, затраченное на ее сходимость. Под сходимостью понимают момент в обучении, когда значения функции потерь перестают уменьшаться и колеблются в окрестностях точки локального минимума, при этом точность принимает удовлетворительное значение.

Выбор гиперпараметров обычно является компромиссом между временем, затраченным на тренировку СНС, и точностью обученной модели. Наилучшее сочетание значений скорости обучения, размера пакета и коли-

чества эпох зависит от большого числа случайных факторов, решаемой задачи и обучающих данных. Поэтому чаще всего значения гиперпараметров определяются опытным путем.

Для нахождения наиболее эффективных гиперпараметров СНС при реидентификации проведен ряд экспериментов с использованием алгоритма [2], СНС ResNet-50 и DenseNet-121, набора данных Market-1501, при которых обучение выполнялось для пакетов разного размера b с различной скоростью η . Использовались пакеты размером 2, 4, 8, 16, 32 и 64 изображения. Скорость тренировки изменялась от 0,001 до 0,01 с шагом 0,001, от 0,01 до 0,1 с шагом 0,01 и далее с шагом 0,1. После 40-й эпохи скорость обучения уменьшалась в 10 раз. Обучение осуществлялось на протяжении 80 эпох на персональном компьютере с основными характеристиками Intel Core i5 3.11 GHz, 16 Gb RAM, Nvidia GeForce RTX-3060 6 Gb. Среднее время, затраченное на тренировку СНС ResNet-50 и DenseNet-121 на наборе данных Market-1501 при использовании пакетов разных размеров, показано на рис. 1.

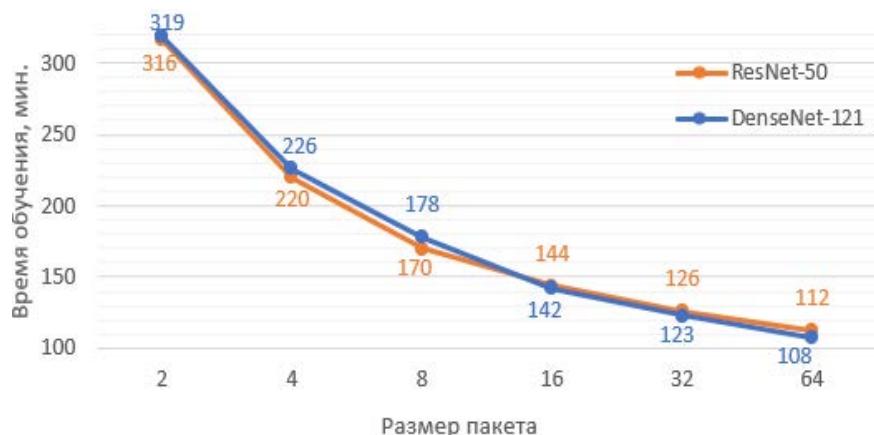


Рис. 1. Зависимость времени обучения СНС ResNet-50 и DenseNet-121 на наборе данных Market-1501 от размера пакета изображений

Как видно из графика на рис. 1, по мере увеличения пакета влияние его размера на время, затраченное на обучение, уменьшается. Например, при увеличении размера пакета с двух до четырех изображений среднее время, затраченное на обучение, уменьшается для СНС ResNet-50 и DenseNet-121 на 96 и 93 мин соответственно, что составляет примерно 29–30 %. При увеличении пакета с 32 до 64 изображений выигрыш времени составляет 11–12 %. Скорость обучения при фиксированном количестве эпох оказывает влияние не на время тренировки, а на скорость сходимости модели.

Из анализа полученных результатов для функции потерь и ошибки top1 следует, что при низкой скорости обучения при увеличении размера пакета количество эпох тренировки СНС необходимо увеличивать для

обеспечения сходимости модели. Это приведет к увеличению временных затрат. В то же время чем больше пакет изображений, тем большую скорость обучения можно установить без риска взрывных градиентов. При использовании небольших пакетов ($b=2$) сеть практически не обучается из-за того, что двух изображений в пакете не хватает для получения достаточной обобщающей способности. Если анализировать только критерии, контролируемые при обучении, то можно сделать предположение, что следует увеличивать размер пакета и скорость обучения для уменьшения значений функции потерь и точности работы сети до определенного порога, после которого возникает риск возникновения неконтролируемого увеличения градиентов. Однако результаты тестирования на изображениях, которые не использовались при обучении, показывают, что при меньшем размере пакета точность в метриках Rank1, mAP и mINP имеет наибольшие показатели Rank1 при сочетании гиперпараметров ($b; \eta$) = (8; 0,03), (8; 0,05) и (16; 0,07); для метрики mAP наилучшие значения гиперпараметров (8; 0,04), (8; 0,05) и (16; 0,09); для метрики mINP (8; 0,04) и (16; 0,1). Наибольшая точность реидентификации на наборе данных Market-1501 достигается при скорости обучения $\eta = 0,09$ или $\eta = 0,1$, однако на наборе данных DukeMTMC-ReID при такой скорости обучения происходит неконтролируемый рост ошибок. Чтобы избежать взрывных градиентов на других наборах данных, принимаются значения гиперпараметров $\eta = 0,07$ и $b = 16$.

На рис. 2 представлены графики функций потерь и ошибки top1 при обучении СНС с гиперпараметрами $b = 8$, $\eta = 0,05$ (рис. 2, *а*) и $b = 16$, $\eta = 0,04$ (рис. 2, *б*).

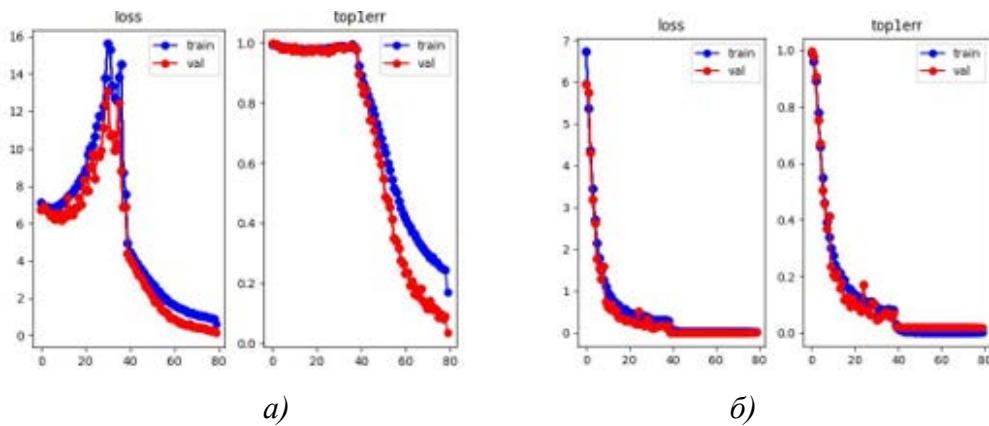


Рис. 2. Графики изменения функции потерь и ошибки top1 при обучении ResNet-50 на наборе данных Market-1501 с разными значениями гиперпараметров:
а) $b = 8$ и $\eta = 0,05$; *б*) $b = 16$ и $\eta = 0,04$

Анализ рис. 2, *а* показывает, что после 10-й эпохи обучения значения функции потерь начали сильно увеличиваться, а снижение скорости обучения позволило остановить этот процесс. На рис. 2, *б* видно, что уменьшение скорости позволило снизить значение функции потерь. Необходи-

мое количество эпох обучения определялось экспериментально аналогично [3]. Модель обучалась на протяжении 110 эпох с применением метода аугментации, рассмотренного в работе [4], и технологии двухэтапного обучения, представленной в работе [5]. Для каждой пятой эпохи определялась точность работы алгоритма реидентификации в метриках Rank1, mAP и mINP. Наибольшая точность реидентификации для СНС ResNet-50 и DenseNet-121, обученных на наборе данных Market-1501, достигается на 80-й эпохе.

Повторное ранжирование нейросетевых признаков

Для поиска соответствия изображений людей в галерее для заданного запроса применяется вычисление расстояний между нейросетевым вектором признаков изображения запроса и нейросетевыми векторами признаков изображений в галерее. После определения сходства векторов и первой сортировки по схожести между изображением-запросом и наиболее схожими изображениями из галереи для повышения точности реидентификации перспективным является повторное ранжирование, которое позволяет уточнить результат. Эффективность такого подхода в значительной мере зависит от качества исходного ранжирования, и, если ближайшими соседями являются неверные предсказания, повторная сортировка признаков может привести к нежелательному результату [6]. Например, на рис. 3, *a* представлен пример реидентификации, на котором вверху списка ближайших соседей не все предсказания верные, поэтому усреднение отличительных признаков для изображений из ранжированного списка будет давать некорректный результат. В связи с этим в работе [7] используются взаимно ближайшие соседи (*k*-reciprocal nearest neighbors). Два изображения считаются *k*-взаимно ближайшими соседями, если они оба занимают место среди первых *k* результатов для другого изображения в качестве запроса (рис. 3, *b*).

Первоначальное ранжирование выполняется с использованием расстояния Махalanобиса. Формируются дополнительные векторы признаков для *k*-взаимно ближайших соседей изображения-запроса и изображений галереи, предсказанных СНС в числе *k* первых по метрике Жаккара. Для объединения двух ранжированных списков вычисляется расстояние между признаками запроса и изображений галереи. Результаты исследования подтверждают эффективность применения повторного ранжирования. Повторное ранжирование признаков использовалось для реидентификации с применением архитектур СНС ResNet-50 и DenseNet-121, трех наборов данных Market-1501, DukeMTMC-ReID и сформированного объединенного набора данных, состоящего из 52 035 изображений для 657 человек. Для реидентификации использовался алгоритм [2]. Анализ результатов показал,

что применение повторного ранжирования [7] позволяет повысить точность реидентификации на разных наборах данных для СНС ResNet-50 по метрике Rank1 на 1–9 %, по метрике mAP на 11–25 %, по метрике mINP на 55–112 %; для СНС DenseNet-121 по метрике Rank1 на 1–5 %, по метрике mAP на 10–25 %, по метрике mINP на 43–107 %.

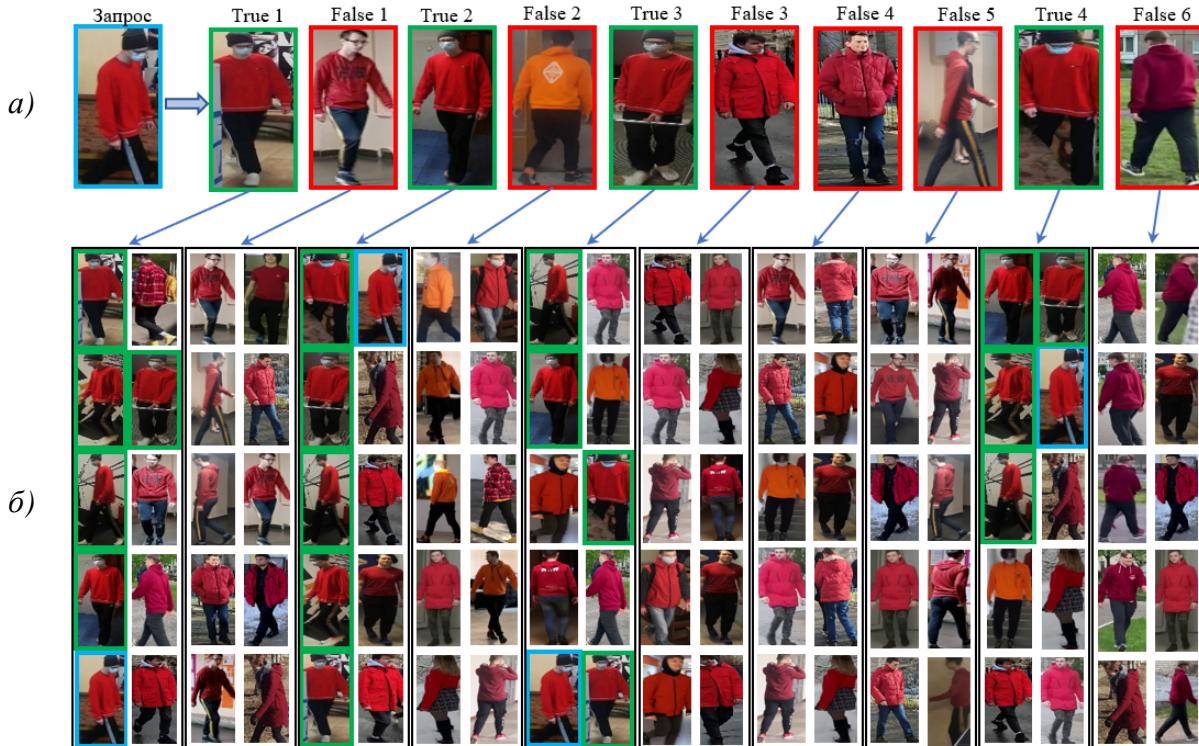


Рис. 3. Пример реидентификации человека:
 а) изображение-запрос и 10 ближайших соседей; б) 10 ближайших соседей для каждого из предсказаний для запроса. Голубая рамка – изображение-запрос, зеленая – верные предсказания, красная – ошибочные

Заключение

Для определения наиболее эффективных значений гиперпараметров при обучении СНС ResNet-50 и DenseNet-121 проведены экспериментальные исследования, которые показали, что для реидентификации лучшие показатели точности достигаются при использовании размера пакета $b = 16$. При скорости обучения $\eta = 0,09$ или $\eta = 0,1$ на наборе данных Market-1501 достигается наилучшее значение точности, однако при использовании другого набора данных возникает риск взрывных градиентов. Так, при обучении на наборе данных DukeMTMC-ReID неконтролируемый рост градиента ошибок возникает уже на скорости $\eta = 0,08$. Таким образом, в качестве гиперпараметров, применяемых для обучения СНС, принимаются $b = 16$ и $\eta = 0,07$.

Применение повторного ранжирования признаков позволяет увеличить точность реидентификации, что подтверждается результатами исследований на наиболее используемых наборах данных, таких как Market-1501, DukeMTMC-ReID, PolReID, и CHC ResNet-50 и DenseNet-121 до 9 % по метрике Rank1, на 10–25 % по метрике mAP и на 43–112 % по метрике mINP.

Список использованных источников

1. Повторная идентификация людей в системах видеонаблюдения с использованием глубокого обучения: анализ существующих методов / Х. Чен, С. А. Игнатьева, Р. П. Богуш, С. В. Абламейко // Автоматика и телемеханика. – 2023. – № 5. – С. 61–112.
2. GitHub // Pytorch ReID. – URL: https://github.com/layumi/Person_reID_baseline_pytorch (date of access: 10.07.2024).
3. Игнатьева, С. А. Аугментация данных для регуляризации в нейросетевых алгоритмах повторной идентификации людей по видеоданным / С. А. Игнатьева, Р. П. Богуш, С. В. Абламейко // Материалы Междунар. конгр. по информатике «Информационные системы и технологии» (CSIST'2022) : в 3 ч., 27–28 окт. 2022 г., Минск. – Минск : БГУ, 2022. – Ч. 2. – С. 132–137.
4. Игнатьева, С. А. Формирование обучающей выборки для сверточных нейронных сетей при реидентификации людей по видеоданным / С. А. Игнатьева, Р. П. Богуш // Доклады БГУИР. – 2023. – № 3. – С. 87–95.
5. Data augmentation and fine tuning of convolution neural network during training for person re-identification in video surveillance systems / S. Ye, R. Bohush, S. Ihnatsyeva, S. V. Ablameyko // Optical Memory and Neural Network. – 2023. – Vol. 4. – P. 233–246.
6. Bohush, R. Person re-identification accuracy improvement by training a CNN with the new large joint dataset and re-rank / R. Bohush, S. Ihnatsyeva, S. Ablameyko // Machine Graphics and Vision. – 2022. – Vol. 31(1/4). – P. 93–109.
7. Re-ranking Person Re-identification with k -Reciprocal Encoding / Z. Zhong, L. Zheng, D. Cao, S. Li // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR-2017), Honolulu, 21– 26 July 2017. – Honolulu, 2017. – P. 3652–3661.