

Булгаков Денис Геннадьевич

*студент 4 курса бакалавриата, Национальный исследовательский ядерный
университет «МИФИ», Россия, г. Москва*

РАСПОЗНАВАНИЕ РУКОПИСНЫХ ПОДПИСЕЙ НА ОСНОВЕ МЕТРИЧЕСКОГО ОБУЧЕНИЯ И СИНТЕТИЧЕСКОЙ АУГМЕНТАЦИИ ОБУЧАЮЩЕЙ ВЫБОРКИ

Аннотация. В статье рассматривается задача оффлайн-распознавания рукописных подписей в writer-independent режиме, при котором решение о подлинности проверяемого образца принимается по ограниченному числу эталонов пользователя, не участвовавшего в обучении модели. Актуальность темы обусловлена сохраняющейся ролью рукописной подписи как средства подтверждения личности при заключении договоров, банковских операциях и нотариальном удостоверении документов, а также ограниченным объёмом публично доступных наборов данных. Цель исследования — разработка нейросетевой модели распознавания на основе метрического обучения и прототипической классификации, а также оценка возможности расширения обучающей выборки средствами генеративного моделирования. В работе используются методы метрического обучения с триплетной функцией потерь, прототипической классификации в пространстве эмбеддингов ResNet-50 и вариационного автоэнкодирования с составной функцией потерь и управляемым локальным сэмплированием. Экспериментальная проверка выполнена на датасете CEDAR с разбиением выборок строго по авторам. Показано, что прототипический классификатор достигает $EER = 0,215$ при пяти эталонных образцах, а синтетическая аугментация обучающей выборки снижает этот показатель до 0,183.

Ключевые слова: рукописная подпись, метрическое обучение, прототипическая классификация, вариационный автоэнкодер, синтетическая аугментация, ResNet-50, CEDAR.

Annotation. This article addresses the problem of offline handwritten signature verification in a writer-independent setting, where the authenticity of a query sample is decided from a limited number of reference signatures of a user who did not participate in training. The relevance of the topic stems from the continuing role of the handwritten signature as a means of identity confirmation in contracts, banking operations and notarial certification of documents, as well as from the limited size of publicly available datasets. The purpose of the study is to develop a recognition model based on metric learning and prototypical classification, and to evaluate whether the training set can be extended by means of generative modelling. The study employs metric learning with a triplet loss, prototypical classification in a ResNet-50 embedding space, and variational autoencoding with a composite loss function and controlled local sampling. Experiments were carried out on the CEDAR dataset with a writer-disjoint split. It is shown that the prototypical classifier reaches $EER = 0.215$ with five reference samples, and that synthetic augmentation of the training set reduces this figure to 0.183.

Keywords: handwritten signature, metric learning, prototypical classification, variational autoencoder, synthetic augmentation, ResNet-50, CEDAR.

Рукописная подпись остаётся одним из наиболее распространённых способов подтверждения личности при заключении договоров, проведении банковских операций и нотариальном удостоверении документов [4]. Массовая цифровизация документооборота повышает нагрузку на экспертов-почерковедов и создаёт потребность в автоматических системах распознавания подписей по статическим изображениям. Практическая ценность таких систем определяется их способностью работать в writer-independent режиме — принимать решение о подлинности подписи нового, не встречавшегося при обучении пользователя, на основании лишь нескольких эталонных образцов [9]. Дополнительным ограничением выступает малый объём публично доступных датасетов подписей, что делает актуальной задачу

синтеза дополнительных обучающих образцов средствами генеративного моделирования.

Задача формулируется как бинарная классификация в режиме ограниченного числа примеров. Исходными данными служат множество из N эталонных образцов подписи пользователя, где N принимает значения от 1 до 10, и проверяемое изображение, о принадлежности которого тому же пользователю необходимо вынести решение. Исторически задача решалась конструированием набора геометрических и текстурных дескрипторов с последующим обучением классического классификатора, например метода опорных векторов [1]; глубокие свёрточные и сиамские архитектуры [2, 3] превзошли этот подход по качеству разделения классов. В отличие от writer-dependent постановки, требующей обучения отдельного классификатора на подписях и имитациях каждого пользователя, в настоящей работе строится единое метрическое пространство, в котором геометрическое расстояние между точками отражает биометрическую несхожесть подписей; такая модель применима к новому пользователю без дообучения и без сбора его имитаций.

Предобработка приводит изображение произвольного разрешения и полярности к единому виду. Выполняются коррекция полярности по средней интенсивности, морфологическая очистка от линовки и рамок операцией открытия, обрезка по минимальному ограничивающему прямоугольнику содержательной области, билинейное масштабирование к разрешению 224×224 и бинаризация по методу Оцу. На этапе обучения добавляется стохастическая аугментация — поворот на угол $\alpha \sim U(-5^\circ, 5^\circ)$, аффинное преобразование с масштабом $s \sim U(0,95; 1,05)$, перспективное искажение и случайное стирание, — отключаемая при валидации и тестировании ради воспроизводимости метрик.

Функция-кодировщик реализована как композиция предобученного бэкбона и обучаемой проекционной головы: $\phi(I) = h_\theta(f_{bb}(I))$, где f_{bb} — сеть ResNet-50 [5] с замороженными весами ImageNet, формирующая 2048-мерный признаковый вектор, а h_θ — последовательность полносвязного

преобразования $2048 \rightarrow 512$, пакетной нормализации, ReLU, Dropout с вероятностью зануления 0,3 и полносвязного преобразования $512 \rightarrow 256$ с финальной L2-нормализацией. Заморозка бэкбона сокращает число оптимизируемых параметров до примерно 1,2 млн и предотвращает переобучение на ограниченном датасете. Нормализация помещает все эмбединги на единичную гиперсферу, где справедливо тождество $\|e_1 - e_2\|^2 = 2(1 - e_1 \cdot e_2)$, то есть минимизация евклидова расстояния эквивалентна максимизации косинусного сходства.

Обучение проводится в два этапа. На первом параметры проекционной головы оптимизируются по триплетной функции потерь [10]:

$$\ell(e^{\text{anc}}, e^{\text{pos}}, e^{\text{neg}}) = \max(0, \|e^{\text{anc}} - e^{\text{pos}}\|_2 - \|e^{\text{anc}} - e^{\text{neg}}\|_2 + m), \quad (1)$$

где $m = 0,3$ — параметр зазора. Случайный выбор триплетов неэффективен: большинство троек уже удовлетворяет условию зазора и даёт нулевой градиент. Поэтому применяется онлайн-майнинг трудных примеров — внутри мини-батча для каждого якоря отбираются наиболее удалённый подлинный образец того же пользователя и наиболее близкий образец чужого. Оптимизация выполняется методом Adam с начальной скоростью обучения 10^{-3} и ранней остановкой по валидационной выборке.

На втором этапе поверх зафиксированного метрического пространства обучается прототипический классификатор [11]. По множеству эталонов строится нормализованный центроид — среднее их эмбедингов, приведённое к единичной длине. Одного расстояния до прототипа недостаточно: разные подписи могут отстоять от него одинаково, различаясь по локальным компонентам отклонения. Поэтому формируется признаковый вектор размерности 515, объединяющий косинусное сходство, евклидово и манхэттенское расстояния с поэлементными разностью и произведением Адамара вектора проверяемого образца и прототипа. На него накладывается двухслойная полносвязная сеть $515 \rightarrow 128 \rightarrow 1$ с сигмодой на выходе, обучаемая минимизацией бинарной кросс-энтропии; итоговое решение

принимается по пороговому правилу. Качество оценивается метриками FAR, FRR, EER и AUC.

Для расширения обучающей выборки разработан вариационный автоэнкодер [8] с латентным пространством размерности $L = 128$. Энкодер состоит из свёрточных блоков с ядром 4×4 и шагом 2, пакетной нормализацией и LeakyReLU, предсказывающих через две независимые головы параметры факторизованного нормального апостериорного распределения. Декодер использует гибридную стратегию апсэмплинга — интерполяцию ближайшего соседа на низких разрешениях и билинейную на высоких, что исключает шахматные артефакты транспонированных свёрток, — с Instance Normalization и функцией гиперболического тангенса на выходе. Прямая подстановка гауссовского правдоподобия ведёт к равномерной среднеквадратичной потере, неприемлемой для изображений подписей, где штрихи занимают малую долю площади. Реконструкционный член заменён составным суррогатом:

$$L = w_{l1} \cdot L_{l1} + w_{ssim} \cdot L_{ssim} + w_{perc} \cdot L_{perc} + w_{edge} \cdot L_{edge} + \\ + w_{ch} \cdot L_{ch} + w_{bin} \cdot L_{bin} + w_{topo} \cdot L_{topo} + \beta \cdot L_{KL}, \quad (2)$$

компоненты которого контролируют пиксельную точность со взвешиванием штрихов в 10 раз сильнее фона, структурное сходство [12], перцептивную близость [13], чёткость границ по картам Собеля и Лапласа, геометрическое расположение штрихов, бинарность выхода через энтропию Шеннона, топологическую связность и регулярность латентного пространства. KL-слагаемое дополнено механизмом Free Bits, гарантирующим минимальный информационный поток через каждое измерение, а его вес линейно нарастает в течение первых 30 эпох [6].

Управляемая генерация опирается на метод репараметризации и выполняется в три шага: эталонное изображение кодируется в параметры апостериорного распределения μ и σ , формируется возмущённый латентный вектор

$$z^* = \mu + \lambda \cdot \sigma \odot \varepsilon, \quad \varepsilon \sim N(0, I_L), \quad (3)$$

и полученный вектор декодируется. Параметр λ линейно регулирует радиус отклонения вокруг центра апостериорного распределения; рабочий диапазон $\lambda \in (0,5; 1,5]$ определён геометрически: при $\lambda \leq 0,5$ вариации слишком малы для обогащения выборки, при $\lambda > 1,5$ возмущение выводит вектор за область высокого апостериорного правдоподобия и деградирует визуальное качество.

Программная реализация выполнена на Python 3.10 с использованием PyTorch 2.x, torchvision, PyTorch Metric Learning (стратегия BatchHardMiner), scikit-image и библиотеки lpips. Система построена из девяти подсистем, связанных через базу данных обученных моделей и подсистему конфигурации, что позволяет развивать обучение и инференс независимо друг от друга. Поверх модулей реализован веб-интерфейс: шлюз на FastAPI и клиентское приложение на React, организованное по методологии Feature-Sliced Design.

Экспериментальное исследование проведено на датасете CEDAR [7]: 55 авторов, по 24 подлинных и 24 поддельных образца на каждого, всего 2640 изображений при разрешении 300 dpi. Разбиение выполнено строго по авторам: 38 авторов в обучающей выборке, 8 — в валидационной и 9 — в тестовой. Генеративная модель обучалась на отдельном корпусе из 30 157 изображений 960 авторов, что исключает пересечение с тестовыми данными. Протокол оценки предусматривает случайный выбор N эталонов для каждого тестового автора с повторением процедуры 20 раз; фиксируются средние значения метрик и стандартные отклонения.

В базовых условиях ($N = 5$, все типы отрицательных примеров) прототипический классификатор достигает $EER = 0,215 \pm 0,024$ и $AUC = 0,793 \pm 0,028$ при доле верных решений 0,785. Значение AUC заметно превышает уровень случайного угадывания, но не достигает порога 0,9, что соответствует приемлемой, но не высокой разделяющей способности метрического пространства. Близкие значения FAR (0,233) и FRR (0,198) при пороге 0,5 указывают на отсутствие смещения в сторону одного из типов ошибок. На обучающей выборке тот же классификатор даёт $EER = 0,097$; разрыв в 0,118

типичен для writer-independent задачи при ограниченном числе обучающих авторов и определяет главное направление совершенствования модели.

При тестировании исключительно на квалифицированных подделках EER возрастает до 0,246, а AUC снижается до 0,758. Средняя апостериорная вероятность подлинности для имитаций составляет 0,410, то есть остаётся ниже порога принятия решения, однако доля подделок с вероятностью выше 0,70 достигает 11,4 % — это группа имитаций, которые классификатор уверенно принимает за подлинные подписи. Наиболее проблематичны подделки, точно воспроизводящие глобальную структуру при расхождении только в деталях штрихов.

Зависимость качества распознавания от числа эталонных образцов приведена в таблице 1. Она монотонна и обнаруживает выраженное насыщение: при $N = 1$ значение $EER = 0,368$ неприемлемо для практического применения, наибольший прирост даёт переход к $N = 2$, диапазон от 3 до 5 эталонов обеспечивает хорошее соотношение качества и объёма данных, а начиная с $N = 5$ прирост AUC замедляется до 1,6 % и ниже.

Таблица 1 – Зависимость качества распознавания от числа эталонных образцов

N	EER	AUC	Accuracy
1	$0,368 \pm 0,040$	$0,698 \pm 0,040$	$0,632 \pm 0,040$
2	$0,304 \pm 0,035$	$0,740 \pm 0,036$	$0,696 \pm 0,035$
3	$0,266 \pm 0,030$	$0,763 \pm 0,032$	$0,734 \pm 0,030$
5	$0,215 \pm 0,024$	$0,793 \pm 0,028$	$0,785 \pm 0,024$
7	$0,194 \pm 0,021$	$0,806 \pm 0,025$	$0,806 \pm 0,021$
10	$0,179 \pm 0,020$	$0,816 \pm 0,023$	$0,821 \pm 0,020$

Аблационное исследование генеративной модели подтвердило значимость каждой компоненты составной функции потерь: полная конфигурация превосходит базовую по SSIM на 4,8 п. п. и по LPIPS на 3,7 п. п. Наибольший вклад вносит потеря на границах штрихов, без которой LPIPS возрастает на 26,8 %; исключение топологической компоненты почти не

меняет агрегированные метрики, однако при качественном анализе проявляются разрывы штрихов в местах пересечений — ошибки, критичные для биометрической корректности, но слабо уловимые пиксельными мерами. Размерность $L = 128$ принята как оптимальная: при $L = 32$ и $L = 64$ ёмкость пространства исчерпана, а при $L = 256$ активны лишь 47 % измерений, что снижает предсказуемость локального сэмплирования. Финальная модель, обученная 150 эпох, достигает $SSIM = 0,7752$ и $LPIPS = 0,1259$ при полной утилизации всех 128 измерений латентного пространства; метод главных компонент показывает, что 91 компонента объясняет 80 % суммарной дисперсии, а 118 компонент — 95 %.

Синтетическая аугментация применялась в двух режимах: генерация трёх вариаций с $\lambda \in [0,6; 1,0]$ для каждого из 912 подлинных образцов обучающей выборки, что даёт 2736 дополнительных образцов, и сэмплирование 1500 образцов из приора $N(0, I_{128})$ в роли дополнительных отрицательных примеров. Добавление синтетических подделок оказалось эффективнее добавления подлинных вариантов (−9,9 % против −6,3 % по EER): расширение разнообразия отрицательных примеров улучшает граничную поверхность классификатора сильнее, чем уточнение уже хорошо представленного положительного класса. Комбинированная стратегия даёт наилучший результат по всем сценариям, что отражено в таблице 2.

Таблица 2 – Сводное сравнение качества распознавания: базовая конфигурация и комбинированная аугментация

Сценарий	Метрика	Baseline	+Both	Δ , %
Базовое исследование, $N = 5$	EER	0,215	0,183	−14,9
Базовое исследование, $N = 5$	AUC	0,793	0,812	+2,4
Квалифицированные подделки, $N = 5$	EER	0,246	0,229	−6,9
Квалифицированные подделки, $N = 5$	AUC	0,758	0,768	+1,3
Зависимость от числа эталонов, $N = 3$	EER	0,266	0,234	−12,0
Зависимость от числа эталонов, $N = 7$	EER	0,194	0,164	−15,5

Аугментация стабильно улучшает показатели во всех сценариях, однако величина эффекта существенно варьируется. Наибольший прирост достигается там, где прототип уже стабилен, наименьший — по квалифицированным подделкам. Причина различия принципиальна: образцы, сэмплированные из приора, биометрически правдоподобны по структуре, но не принадлежат конкретному автору, то есть эквивалентны неквалифицированным подделкам — они расширяют класс «случайных» имитаций, но не помогают распознавать целенаправленные.

Проведённое исследование позволяет утверждать, что связка метрического обучения и прототипической классификации обеспечивает работоспособное решение задачи распознавания рукописных подписей в writer-independent режиме при ограниченном числе эталонов. Разработанная модель применима к новому пользователю без дообучения, добавление эталонов не требует изменения весов, а достигнутые показатели соответствуют требованиям систем умеренной критичности. Практическим оптимумом является диапазон от 5 до 7 эталонных образцов, поскольку при меньших значениях качество падает резко, а при больших прирост становится незначительным.

Применение генеративной модели для расширения обучающей выборки подтвердило свою эффективность: комбинированная стратегия аугментации снижает EER с 0,215 до 0,183 и повышает AUC с 0,793 до 0,812. Вариационный автоэнкодер оказался предпочтительнее альтернативных генеративных парадигм за счёт регулярности латентного пространства, которая делает возможным управляемое локальное сэмплирование вблизи эталона, а не только безусловную генерацию. Составная функция потерь при этом необходима: равномерная пиксельная мера игнорирует штрихи в пользу точного воспроизведения фона, а отдельные аспекты биометрической корректности — прежде всего связность штрихов — вообще не улавливаются агрегированными метриками качества изображений.

Вместе с тем исследование выявило ограничения, без преодоления которых дальнейшее повышение качества затруднительно. Даже с аугментацией значение $EER = 0,183$ при $N = 5$ означает примерно одну ошибку на каждые пять-шесть решений — уровень, достаточный для задач с умеренными требованиями к надёжности, но недостаточный для применений повышенной критичности без дополнительных мер. Квалифицированные имитации остаются главной нерешённой задачей: разрыв между базовым сценарием и сценарием с квалифицированными подделками после аугментации не сокращается, а возрастает с 0,031 до 0,046. Главным ограничивающим фактором выступает не качество синтеза, а узость обучающей выборки: 38 реальных авторов задают конечное разнообразие стилей, которое синтетические образцы воспроизводят, но не расширяют принципиально.

Перспективными направлениями дальнейшей работы, вытекающими из полученных результатов, являются увеличение числа реальных авторов в обучающей выборке за счёт объединения нескольких публичных датасетов и разработка условного генератора имитаций, способного порождать целенаправленные подделки конкретного автора. Именно эти две меры адресуют выявленные ограничения напрямую, тогда как ни увеличение числа эталонов, ни аугментация неквалифицированными образцами по отдельности их не устраняют.

Список литературы

1. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. М.: Наука, 1974. 416 с.
2. Dey S., Dutta A., Toledo J. I., Ghosh S. K., Lladós J., Pal U. SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification // arXiv preprint arXiv:1707.02131. 2017. 7 p.
3. Hafemann L. G., Sabourin R., Oliveira L. S. Learning Features for Offline Handwritten Signature Verification using Deep Convolutional Neural Networks // Pattern Recognition. 2017. Vol. 70. P. 163–176.

4. Hafemann L. G., Sabourin R., Oliveira L. S. Offline Handwritten Signature Verification – Literature Review // Proceedings of the International Conference on Image Processing Theory, Tools and Applications (IPTA). 2017. P. 1–8.
5. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778.
6. Higgins I., Matthey L., Pal A., Burgess C., Glorot X., Botvinick M., Mohamed S., Lerchner A. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework // Proceedings of the 5th International Conference on Learning Representations (ICLR). 2017. 22 p.
7. Kalera M. K., Srihari S., Xu A. Offline Signature Verification and Identification Using Distance Statistics // International Journal of Pattern Recognition and Artificial Intelligence. 2004. Vol. 18, No. 7. P. 1339–1360.
8. Kingma D. P., Welling M. Auto-Encoding Variational Bayes // Proceedings of the 2nd International Conference on Learning Representations (ICLR). 2014. 14 p.
9. Rantzsch H., Yang H., Meinel C. Signature Embedding: Writer Independent Offline Signature Verification with Deep Metric Learning // Proceedings of the International Conference on Biometrics (ICB). 2016. P. 616–623.
10. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 815–823.
11. Snell J., Swersky K., Zemel R. S. Prototypical Networks for Few-shot Learning // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 4077–4087.
12. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Transactions on Image Processing. 2004. Vol. 13, No. 4. P. 600–612.
13. Zhang R., Isola P., Efros A. A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric // Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 586–595.

References

1. Vapnik V. N., Chervonenkis A. Ya. Teoriya raspoznavaniya obrazov [Pattern recognition theory]. Moscow, Nauka Publ., 1974. 416 p.
2. Dey S., Dutta A., Toledo J. I., Ghosh S. K., Lladós J., Pal U. SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification // arXiv preprint arXiv:1707.02131. 2017. 7 p.
3. Hafemann L. G., Sabourin R., Oliveira L. S. Learning Features for Offline Handwritten Signature Verification using Deep Convolutional Neural Networks // Pattern Recognition. 2017. Vol. 70. P. 163–176.
4. Hafemann L. G., Sabourin R., Oliveira L. S. Offline Handwritten Signature Verification – Literature Review // Proceedings of the International Conference on Image Processing Theory, Tools and Applications (IPTA). 2017. P. 1–8.
5. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778.
6. Higgins I., Matthey L., Pal A., Burgess C., Glorot X., Botvinick M., Mohamed S., Lerchner A. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework // Proceedings of the 5th International Conference on Learning Representations (ICLR). 2017. 22 p.
7. Kalera M. K., Srihari S., Xu A. Offline Signature Verification and Identification Using Distance Statistics // International Journal of Pattern Recognition and Artificial Intelligence. 2004. Vol. 18, No. 7. P. 1339–1360.
8. Kingma D. P., Welling M. Auto-Encoding Variational Bayes // Proceedings of the 2nd International Conference on Learning Representations (ICLR). 2014. 14 p.
9. Rantzsch H., Yang H., Meinel C. Signature Embedding: Writer Independent Offline Signature Verification with Deep Metric Learning // Proceedings of the International Conference on Biometrics (ICB). 2016. P. 616–623.

10. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 815–823.
11. Snell J., Swersky K., Zemel R. S. Prototypical Networks for Few-shot Learning // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 4077–4087.
12. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Transactions on Image Processing. 2004. Vol. 13, No. 4. P. 600–612.
13. Zhang R., Isola P., Efros A. A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 586–595.