

DOI: 10.15514/ISPRAS-2026-38(2)-13



Совмещение динамических характеристик почерка и модели ConvNeXtV2 для распознавания рукописных символов на русскоязычных и англоязычных наборах данных

Д.В. Яценко, ORCID: 0000-0003-1014-0376 <iatcenko@sfedu.ru>

*Институт высоких технологий и нэзотехники ЮФУ,
Россия, 344090, г.Ростов-на-Дону, ул. Мильчакова, д. 10.*

Аннотация. Представленная работа развивает результаты исследования автора по использованию динамических характеристик движения пера для повышения качества распознавания рукописного текста. В настоящей статье мы объединяем ранее предложенную динамическую компоненту (угловые и спектральные признаки траекторий) с флагманской моделью глубокой нейронной сети распознавания изображений ConvNeXtV2 tiny, и проводим исследование на трех наборах данных: EMNIST (by_class), UJI Pen Characters 2 и собственном наборе Russian Handwritings Tracked. Показано, что после умеренной настройки аугментаций визуальная ветвь достигает на наборе EMNIST результата уровня SOTA среди сравниваемых моделей, тогда как динамическая ветвь и их ансамбль обеспечивают прирост устойчивости на наборах с большей вариативностью почерка (UJI, Russian Handwritings Tracked). Итоги подтверждают комплементарность визуальных и кинематических признаков и указывают на перспективность метода для сценариев «сложного почерка».

Ключевые слова: распознавание рукописного текста; онлайн-почерк; кинематические признаки; траектория пера; спектральные признаки; модель ConvNeXtV2; набор данных EMNIST; набор данных UJI2; набор данных Russian Handwritings Tracked; ансамблирование.

Для цитирования: Яценко Д.В. Совмещение динамических характеристик почерка и модели ConvNeXtV2 для распознавания рукописных символов на русскоязычных и англоязычных наборах данных. Труды ИСП РАН, том 38, вып. 2, 2026 г., стр. 195–206. DOI: 10.15514/ISPRAS–2026–38(2)–13.

Combining Handwriting Dynamics with ConvNeXtV2 Convolutional Backbone for Handwritten Character Recognition on Russian- and English-language Datasets

D.V. Iatsenko, ORCID: 0000-0003-1014-0376 <iatsenko@sfedu.ru>

*Institute of High Technologies and Piezotechnics, SFedU,
Russia, 344090, Rostov-on-Don, Milchakova st., bldg. 10.*

Abstract. This work extends the author’s previous research on leveraging dynamic pen-motion characteristics to improve handwritten text recognition. We combine the earlier proposed dynamic component—angular and spectral trajectory features—with a modern visual backbone, ConvNeXtV2_tiny, and evaluate on three datasets: EMNIST (by_class), UJI Pen Characters 2, and our own Russian Handwritings Tracked. We show that, after moderate augmentation tuning, the visual branch achieves state-of-the-art performance on EMNIST among the models compared, while the dynamic branch and their ensemble improve robustness on datasets with greater handwriting variability (UJI, Russian Handwritings Tracked). The results confirm the complementarity of visual and kinematic features and highlight the promise of the method for “difficult handwriting” scenarios.

Keywords: handwriting recognition; online handwriting; kinematic features; pen trajectory; spectral features; ConvNeXtV2; EMNIST; UJI2; Russian Handwritings Tracked; ensembling.

For citation: Iatsenko D.V. Combining handwriting dynamics with the modern ConvNeXtV2 convolutional backbone for handwritten character recognition on Russian- and English-language Datasets. *Trudy ISP RAN/Proc. ISP RAS*, vol. 38, issue 2, 2026, pp. 195-206 (in Russian). DOI: 10.15514/ISPRAS-2026-38(2)-13.

1. Введение

Распознавание рукописных символов традиционно решается методами оптического распознавания изображений символов. Однако реальные рукописные данные зачастую имеют малое число наблюдений на класс, значительную вариативность почерка и неоднозначные классы, вследствие чего качество чисто визуальных методов резко падает. В предыдущей работе автора была предложена методика извлечения динамических признаков движения пера (угол направления, спектральное представление траектории движения пера), инвариантных к положению, масштабу и наклону, и показано, что их ансамблирование с классическими визуальными моделями заметно повышает точность классификации в трекированных наборах данных.

Используемые подходы. В работе рассматриваются две комплементарные ветви: (i) оптическая – современная сверточная модель ConvNeXtV2_tiny [1-3] с умеренными аугментациями; (ii) динамическая – инвариантные к положению/масштабу/наклону угловые и спектральные признаки траектории движения пера, классифицируемые простыми одномерными сверточными нейронными сетями (1D-CNN) или рекуррентной нейронной сетью с долгой краткосрочной памятью (LSTM). Предсказания ветвей объединяются взвешенным усреднением вероятностей (temperature-softmax, label smoothing). Такой дизайн акцентирует объединение устойчивых кинематических признаков с сильной визуальной моделью и далее он сравнивается на наборах данных EMNIST, UJI и Russian Handwritings Tracked, подробности в разделах 4.1-4.3 и табл. 1.

Цель настоящей работы – расширить и углубить указанные результаты, дополнив визуальную ветвь моделью ConvNeXtV2_tiny и оценив влияние архитектуры Сверточных Нейронных Сетей (Convolutional Neural Network, далее CNN) при фиксированных ограничениях по данным.

2. Связь с предыдущими публикациями

2.1 Предшествующие результаты

В ранней статье [4] были представлены следующие ключевые положения: (i) разработка архитектур, учитывающих как визуальные, так и динамические признаки; (ii) метод построения признаков на основе угла траектории и его спектра; (iii) значимый выигрыш при ансамблировании ветвей «картинка + динамика» на опубликованных наборах данных (базах данных) UJI и RUS с точностью (метрика ассурасу, как отношение правильно предсказанных результатов к общему количеству) до 0,75 и 0,62 соответственно. Там же приведены базовые значения точности для классической модели нейронной сети, известной как LeNet-5, а также для других простых сверточных нейронных сетей и многослойных полносвязных сетей на преобразованных изображениях и на динамических признаках.

2.2 Публикация о наборе данных Russian Handwritings Tracked

В отдельной публикации [5] подробно описан набор Russian Handwritings Tracked: 12 писателей, 76 классов (строчные/прописные буквы и цифры), а также слова из панграммы (словосочетания, содержащего все буквы алфавита) «съешь ещё этих мягких французских булок, да выпей же чаю». Данные содержат координаты траекторий пера и утилиты для перевода траекторий в изображения формата 28×28 . Показано, что при использовании базовых моделей нейронных сетей LeNet-5 и ResNet-18, качество на этом наборе данных заметно ниже, чем на популярном наборе данных EMNIST. Это объясняется большим количеством классов, меньшим числом образцов и высокой межклассовой схожестью.

3. Наборы данных

Чтобы объективно оценить влияние динамических признаков и визуальной ветви, мы подбираем наборы данных, различающиеся по модальности, алфавиту и размеру выборки: EMNIST (by_class), UJI Pen Characters 2 и Russian Handwritings Tracked. Это позволяет сопоставлять результаты в условиях как фиксированных изображений, так и треков пера.

3.1 EMNIST (by_class)

Набор данных EMNIST – это расширение набора данных MNIST до 62 классов (10 цифр + буквы латиницы), порядка $8 \cdot 10^5$ образцов (рис. 1). В экспериментах применена корректировка ориентации (поворот на -90 градусов и горизонтальное отражение), соответствующая официальным рекомендациям.



Рис. 1 Визуализация набора данных EMNIST (by_class).
Fig. 1. Visualization of the EMNIST (by_class) dataset.

3.2 UJI Pen Characters

UJI Pen Characters является онлайн-набором данных с траекториями испанского алфавита и цифр [6-7] (62 класса), порядка $1.16 \cdot 10^4$ образцов. Используются как исходные треки, так и их конвертация в растровое представление 32×32 (рис. 2).

3.3 Russian Handwritings Tracked (RUS)

Собранный и опубликованный ранее автором набор рукописных символов Russian Handwritings Tracked [8]: содержит 2812 образцов, 76 классов, 12 писателей, доступен для загрузки в виде векторных треков, так и в виде изображений 28×28 (рис. 3), также доступны записи целых слов для задач атрибуции авторства.

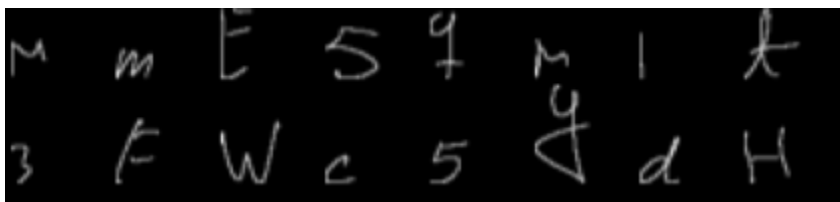


Рис. 2: Визуализация набора данных UJI.
Fig. 2. Visualization of the UJI Dataset.



Рис. 3: Визуализация набора данных RUS.
Fig. 3. Visualization of the RUS Dataset.

4. Методы

4.1 Визуальная ветвь: ConvNeXtV2_tiny

В качестве визуальной модели применена ConvNeXtV2_tiny. Чтобы оставаться в сопоставимом вычислительном бюджете, но в то же время использовать минимальный размер, релевантный архитектуре модели изображения, входное разрешение установлено в 96×96 (масштабирование исходных изображений). Оптимизация: алгоритм AdamW, функция потерь кроссэнтропия, план обучения с тёплым стартом (warmup) и косинусным спадом, пачка данных для одновременной подачи на вход модели (батч данных) размером 512 образцов. В текущей работе мы используем модель нейронной сети ConvNeXt V2 с классификационным полносвязным слоем.

4.1.1 Блок ConvNeXtV2 (вариант Tiny)

Почему выбран вариант ConvNeXt V2-Tiny? Эта архитектура даёт хороший баланс точности и скорости на ограниченном бюджете, оставаясь полностью свёрточной (без self-attention, ресурсоемкой операции вычисления глобального внимания), что упрощает развёртывание и ускоряет прямой проход (inference) на CPU/GPU. Ключевое отличие версии V2 сети ConvNeXt заключается в Global Response Normalization (GRN) модуле нормализации сигнала, позволяющем стабилизировать и усиливать признаки, инвариантные к масштабу внутри MLP-подсети блока (рис. 4); она особенно полезна при пониженном входном разрешении. Линейные проекции 1×1 и глубокие свёртки делают блоки лёгкими по параметрам и вычислительной сложности (FLOPs – числу операций с плавающей точкой), а вариант Tiny с каналами [96, 192, 384, 768] и глубинами [3, 3, 9, 3] служит компактной, но сильной «базовой» конфигурацией для дальнейшей адаптации. В представленной работе применяется стандартный начальный свёрточный блок (stem) с шагом 4 (patch size 4) [9].

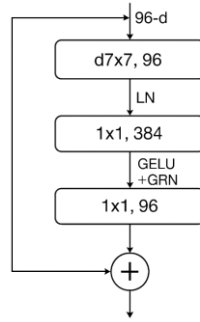


Рис. 4: Архитектура блока ConvNeXt.
Fig. 4. Architecture of the ConvNeXt block.

Далее формально опишем вычисления в одном резидуальном блоке ConvNeXt V2 [1] с учётом Global Response Normalization (GRN) [2]. Обозначим вход блока $X \in \mathbb{R}^{H \times W \times C}$ (формат channels-last).

- 1) Глубинная свёртка (depthwise).

$$U = DWConv_{7 \times 7}(X) \quad (1)$$

- 2) Нормализация и расширяющая 1×1 -проекция (MLP на свёртках).

$$\tilde{U} = LM(U), \quad (2)$$

$$V = \phi(\tilde{U}W_1 + b_1), \quad W_1 \in \mathbb{R}^{C \times rC}, \quad r = 4 \quad (3)$$

где ϕ – нелинейность (типично GELU), а умножение реализуется 1×1 -свёрткой.

- 3) Global Response Normalization (GRN) [3]. Для каждого канала $i = 1, \dots, C$ вычислим L_2 -агрегацию по пространству:

$$g_i = \left(\sum_{h=1}^H \sum_{w=1}^W V_{h,w,i}^2 \right)^{1/2} \quad (4)$$

Нормируем на среднее по каналам:

$$n_i = \frac{g_i}{\frac{1}{C} \sum_{j=1}^C g_j + \varepsilon} \quad (5)$$

и калибруем отклики:

$$\hat{V}_i = V_i \cdot n_i \quad (6)$$

С учётом обучаемых аффинных параметров γ, β и резидуального соединения итоговый выход подсети после GRN записывается как

$$Y = \gamma \hat{V} + \beta + V \quad (7)$$

- 4) Сжимающая 1×1 -проекция и резидуальное сложение.

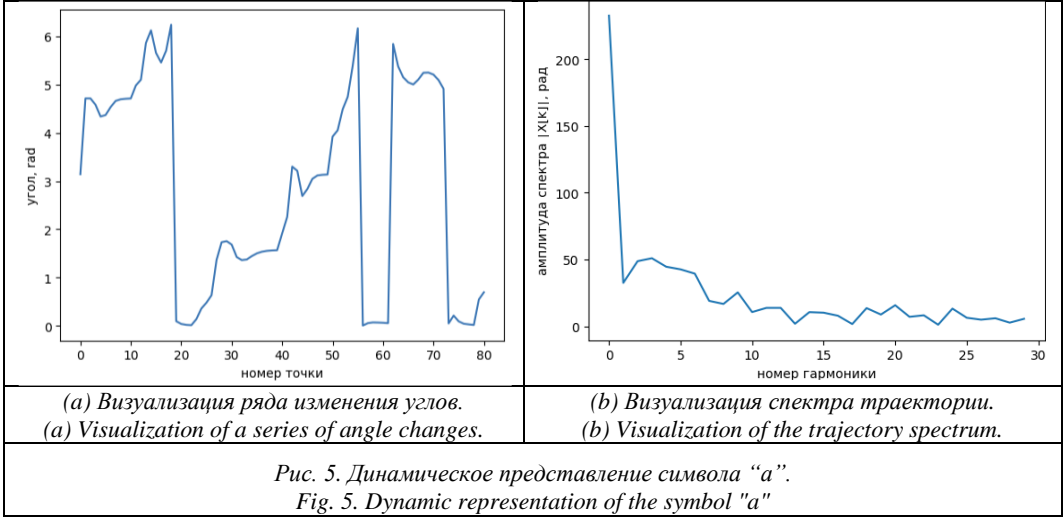
$$Z = Y W_2 + b_2, \quad W_2 \in \mathbb{R}^{rC \times C} \quad (8)$$

$$\text{Out} = X + Z \quad (9)$$

4.2 Динамическая ветвь

Для UJI и RUS извлекается дискретный ряд углов направлений траектории $\alpha_i = \arctan \frac{\Delta y}{\Delta x}$, и параллельно используется его спектральное представление (rFFT) с усечением до первых 30 компонент. Таким образом фактически получали два табличных набора данных: один с признаками угловых изменений и второй со спектрами движений. К наборам данных применялись различные классификаторы – полносвязная (FNN)/сверточная 1D (CNN)/рекуррентная (LSTM). На рисунке 5.а представлен график изменения углов по траектории

для русского символа “а” а на рисунке 5.б представлена визуализация спектра траектории для этой же буквы.



4.3 Ансамблирование моделей

Для финального предсказания используется ансамбль из M независимых моделей (разные инициализации/аугментации) визуальной и динамической ветви. Для ансамблирования применялось взвешенное усреднение численной оценки/вероятностей динамической ветви с визуальной ветвью.

Численные оценки (логиты) модели $z \in R^K$ преобразуются в вероятности через нелинейную функцию Softmax с температурой τ :

$$\hat{p}_i = \frac{\exp(z_i/\tau)}{\sum_{j=1}^K \exp(z_j/\tau)} \quad (10)$$

Обучение ведётся с функцией потерь кросс-энтропии с label smoothing параметра ϵ :

$$\mathcal{L}^{(\epsilon)} = -(1 - \epsilon) \log \hat{p}_y - \sum_{i \neq y} \frac{\epsilon}{K-1} \log \hat{p}_i \quad (11)$$

Ансамбль усредняет апостериорные распределения:

$$\bar{p} = \frac{1}{M} \sum_{m=1}^M \widehat{p}^{(m)}, \quad \hat{y} = \arg \max_i \bar{p}_i \quad (12)$$

Выводы об ансамблировании. Компромисс смещения-дисперсии (Bias-Variance Tradeoff [10, 11]) гласит: обобщающая ошибка модели складывается из систематического смещения, разброса предсказаний и неизбежного шума; уменьшая одно, мы обычно увеличиваем другое. Практическая цель – найти баланс, где суммарная ошибка минимальна (в том числе с помощью приёмов вроде ансамблирования, которое снижает дисперсию за счёт усреднения).

Пусть $f_j(x)$ – предсказание j -й модели, $y(x)$ – истинная (детерминированная) цель, а ошибка j -й модели $\xi_j(x) = f_j(x) - y(x)$. Предположим:

$$E_x[\xi_j(x)] = 0, E_x[\xi_j(x)^2] = E_1 \text{ (одинаково для всех } j), E_x[\xi_i(x)\xi_j(x)] = 0 \text{ при } i \neq j$$

Тогда для усреднённого ансамбля $\bar{f}(x) = \frac{1}{N} \sum_{j=1}^N f_j(x)$ имеем:

$$E_N = E_x \left(\bar{f}(x) - y(x) \right)^2 = E_x \left(\frac{1}{N} \sum_{j=1}^N \xi_j(x) \right)^2 =$$

$$\frac{1}{N^2} E_x \left(\sum_{j=1}^N \xi_j(x)^2 + \sum_{i \neq j} \xi_i(x) \xi_j(x) \right) = \frac{1}{N^2} (N E_1 + 0) = \frac{E_1}{N} \quad (13)$$

И хотя в представленной работе не исследовалась коррелированность ошибок между моделями, тем не менее наши эксперименты показывают, что улучшение метрик ансамбля по сравнению с отдельными моделями, эмпирически снижает ошибку, что согласуется с ожиданиями смещения-дисперсии и объясняется уменьшением дисперсии/повышением разнообразия между моделями.

4.4 Протокол обучения и метрики

Для EMNIST разделение: обучающая/валидационная части – случайное разделение 95/5 обучающей подвыборки; а тест – официальный тестовый набор. Отчётные метрики: top-1/top-3/top-5 accuracy. Для UJI/RUS оценка проводится по 3-кратному повторению со случайными сплитами и усреднением.

5. Результаты

Автором были проведены серии экспериментов по работе с ConvNeXtV2_tiny на различных наборах данных, а также агрегация (ансамблирование) ее с динамической ветвью моделей. Исходные коды к этой и предыдущим работам, конфигурации обучения и скрипты опубликованы в репозитории [12].

5.1 Сопоставление с базовыми моделями

В табл. 1 сведены базовые значения из предыдущих работ авторов и новые результаты визуальной ветви на архитектуре ConvNeXtV2_tiny. Для полноты включены результаты динамической ветви и ансамблей. В таблице значения из предыдущих работ приведены курсивом; новые результаты – прямым шрифтом.

Табл. 1. Сводные результаты (top-1, %) на трёх наборах данных.
Table 1. Summary results (top-1, %) on three datasets.

Набор данных	Архитектура (ветвь)	Вход	Вход Точность
<i>EMNIST by_class</i>	<i>LeNet-5 (изобр.)</i>	<i>28 × 28</i>	<i>86%</i>
<i>UJI</i>	<i>LeNet-5 (изобр.)</i>	<i>32 × 32</i>	<i>60%</i>
<i>RUS</i>	<i>LeNet-5 (изобр.)</i>	<i>28 × 28</i>	<i>34%</i>
<i>RUS</i>	<i>Динамика (CNN по углам)</i>	–	<i>51%</i>
<i>RUS</i>	<i>Ансамбль (изобр.+динамика)</i>	–	<i>63%</i>
EMNIST by_class	ConvNeXtV2_tiny (изобр.)	96 × 96	79%
UJI	ConvNeXtV2_tiny (изобр.)	96 × 96	81%
RUS	ConvNeXtV2_tiny (изобр.)	96 × 96	68%
RUS	ConvNeXtV2_tiny (изобр.+динамика)	–	73%

5.2 Наблюдения

Визуальная ветвь на базе ConvNeXtV2_tiny показала на наборе данных EMNIST результат немного хуже, чем у сравниваемой модели (см. Табл. 1). Это вызвано тем, что модель избыточно сильная и не так хорошо обобщает слишком простые данные, как это делает на пару порядков более простая модель LeNet-5. Но вот на сложном наборе данных RUS ситуация меняется на противоположную: ConvNeXtV2_tiny обеспечивает вдвое более высокую точность по сравнению с LeNet-5, 68% против 34%.

В то же время цель работы состоит не в замене визуальной ветви, а в повышении точности за счёт кинетических признаков (траектория пера, скорость, ускорение, спектральные характеристики). На наборах данных с ограниченным числом примеров и/или с сильной вариативностью написания (UJL, RUS) динамическая ветвь остаётся более устойчивой к искажениям и даёт прирост при ансамблировании с изображениями. В совокупности ансамбль «изображения + динамика» демонстрирует наиболее стабильные результаты на всех наборах данных.

Особенно важно то, что самая простая модель LeNet-5 [13], в паре с динамической ветвью, на наборе данных Russian Handwritings Tracked [8] подошла вплотную к лучшей опубликованной (state of the art, SOTA) оптической модели: 63% против 68%. Это подчёркивает комплементарность кинематических признаков к визуальным.

Итог: Добавление динамических признаков обеспечивает прирост качества, дополнительную устойчивость и переносимость между наборами данных.

5.3 Возможности применения

Также исследовалось (Iatsenko, Dissertation, in prep.) применение данных о спектре функции изменения угла движения пера в процессе написания слов для атрибуции авторства по словам/фразам. В разделе extra набора данных Russian Handwritings Tracked есть образцы слов из фразы-панграммы “съешь еще этих мягких французских булок да выпей чаю”. Для распознавания применяется предобученная на спектральных данных модель, у которой удален слой классификатора. То есть модель работает в режиме извлечения признаков (feature extractor) – выходное значение сети описывает исследуемый объект данных как вектор значений в признаковом пространстве. Используя метрику косинусного сходства (14), можно оценить степень схожести двух образцов.

$$similarity = \frac{x_1 \cdot x_2}{\max(|x_1|_2, |x_2|_2, \epsilon)} \quad (14)$$

Если для анализа используется не одно слово, а фраза, в среднем состоящая из 5 слов, то мера косинусного сходства между фразами написанными одним или разными авторами различается очень заметно (рис. 6).

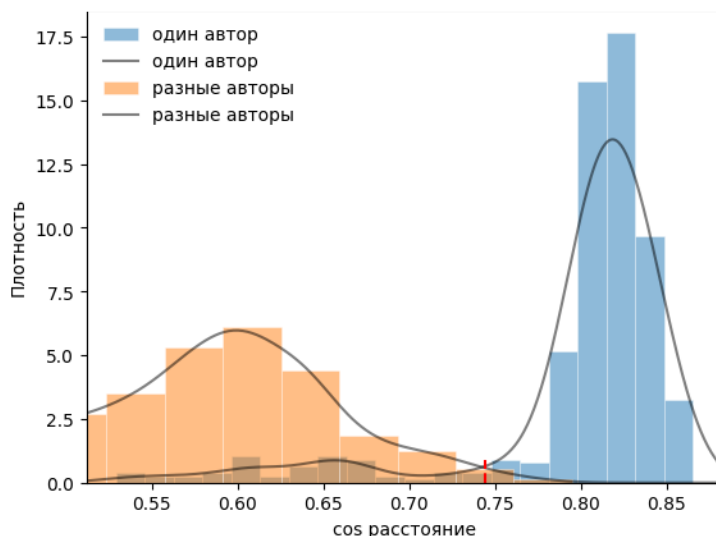


Рис. 6. Сходство/различие фраз.
Fig. 6. Similarity/difference of phrases.

Исходя из полученного распределения было выбрано пороговое значение 0.744. В результате эксперимента была получена общая сбалансированная точность определения сходства образцов фраз одного автора в 93%.

6. Анализ и обсуждение

Цель и общий вывод. Цель представленной работы состояла не столько в замене ветви алгоритма для оптического распознавания (визуальной части алгоритма) на ConvNeXtV2_tiny, сколько в повышении точности и устойчивости за счёт кинетических признаков (траектория, скорость, ускорение, спектральные характеристики) и их объединения с визуальной ветвью. ConvNeXtV2_tiny ожидаемо достигла на наборе данных EMNIST наиболее высокого результата среди сравниваемых решений (см. Табл. 1). При этом добавление динамической информации остаётся важным фактором стабильности на наборах с высокой вариативностью почерка и ограниченным числом примеров.

Влияние аугментаций визуальной ветви. В ходе серии экспериментов на наборе EMNIST было выявлено заметное увеличение точности с уменьшением интенсивности геометрических трансформаций: для символов, сходных с символами из набора MNIST и его производных - слишком сильные искажения «съедают» мелкие штрихи и приводят к деградации. Умеренные аугментации позволяют ConvNeXt V2 раскрыть потенциал без излишней вариативности данных. Практический вывод: для мелких бинарных/полутоновых знаков агрессивность аугментаций критична и требует подбора под домен.

Комплементарность модальностей и ансамблирование. При работе с наборами данных UJI и RUS динамическая ветвь демонстрирует большую устойчивость к стиливым колебаниям и шуму записи, чем чисто визуальная. Совместное использование признаков изображения и движения даёт выигрыш за счёт понижения дисперсии ошибок и меньшей их коррелированности между ветвями (см. раздел о компромиссе смещения–дисперсии). Отсюда – более стабильные результаты ансамбля «изображения + динамика» на рассмотренных наборах данных.

Поведение на «сложном почерке». Кинематические признаки особенно полезны, когда геометрия нестабильна: спешное письмо, тремор, индивидуальные моторные особенности (например, у медработников). В таких сценариях траекторные и ритмические паттерны сохраняют информативность и помогают визуальной ветви разрешать неоднозначные контуры. Это объясняет, почему даже простая нейронная сеть LeNet-5 в паре с динамической ветвью значительно сокращает разрыв с более мощными оптическими моделями на наборе RUS.

Переносимость и ограничения. Сочетание модальностей повышает переносимость между наборами данных и условиями записи; однако для офлайн-документов без треков пера эффект ограничен (необходима реконструкция штриха или сбор онлайн-записей). Также отмечается чувствительность итогов к выбору схемы аугментаций и нормализаций в визуальной ветви; эти гиперпараметры следует подбирать отдельно под каждый домен.

7. Практическая значимость

Объединение современной визуальной модели с динамическими признаками позволяет:

- повышать устойчивость к плохому почерку и редким буквенным формам;
- переносить стиль письма и выполнять атрибуцию авторства (по словам и фразам) ;
- встраивать решение в планшетные интерфейсы «онлайн»-распознавания.

8. Заключение

Мы изучили подход, сочетающий современную визуальную архитектуру (ConvNeXtV2_tiny)

и признаки кинематики письма. После коррекции стратегии аугментаций визуальная ветвь достигла уровня, сопоставимого с передовыми алгоритмами на наборе данных EMNIST, тогда как динамическая ветвь и их ансамбль обеспечили выигрыш устойчивости на наборах с большей вариативностью почерка. Тем самым подтверждается ключевая идея работы: анализ движения пера несёт комплементарную к изображению информацию и улучшает распознавание.

Практическая релевантность. Мы ожидаем, что учёт кинематики даст особенно заметные преимущества в сценариях «трудного почерка» – у медицинских работников, пользователей с тремором/микротремором, при спешном письме и др. Геометрия знаков при этом нестабильна, тогда как паттерны движения (скоростные профили, ритмика, спектральные характеристики) сохраняют индивидуальные и классовые регулярности, помогая корректно интерпретировать неоднозначные контуры.

Ограничения и будущая работа.

- (1) Для офлайн-документов без траекторий эффект ограничен; требуется восстановление штриха или сбор онлайн-записей.
- (2) Нужны целевые наборы данных со «сложным почерком» и клиническими группами; перспективным направлением является проведение отдельных экспериментов и стресс-тестов с искусственными дрожаниями/рывками.
- (3) Перспективно самосупервизорное (self-supervised) обучение на больших корпусах онлайн-почерка и более тесная интеграция динамики в визуальную ветвь (модульная фузия, адаптивное взвешивание).

Список литературы / References

- [1]. Woo S., Debnath S., Hu R., Chen X., Liu Z., Kweon I. S., Xie S. Convnext v2: Codesigning and scaling convnets with masked autoencoders, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 16133-16142. Available at: <https://arxiv.org/pdf/2301.00808>, accessed 10.01.2026.
- [2]. Woo S. et al., Convnext v2: Co-designing and scaling convnets with masked autoencoders,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023. Global Response Normalization (GRN)
- [3]. Saenko I., Lauta O., Iatsenko D. The use of dynamic characteristics in handwriting recognition tasks, in 2023 International Ural Conference on Electrical Power Engineering (UralCon), 2023, pp. 609-614.
- [4]. Яценко Д. В., Граецкая О. В., Смирнов К. В. Набор данных рукописных символов русского языка. Вестник компьютерных и информационных технологий, Feb. 2025, с. 46-54.
- [5]. Prat F., Castro M., Llorens D., Marzal A., Vilar J. UJI Pen Characters (Version 2), 2008. Dataset is available at: <https://doi.org/10.24432/C5FG8S>, accessed 10.01.2026.
- [6]. Llorens D., Prat F., Marzal A., Vilar J. M., Castro M. J., Amengual J. C., Barrachina S., Castellanos A., España S., Gómez J. A., Gorbe J., Gordo A., Palazón V., Peris G., Ramos-Garijo R., Zamora F. The UJPenchars database: a pen-based database of isolated handwritten characters, in Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), (Marrakech, Morocco), European Language Resources Association (ELRA), 2008.
- [7]. Iatsenko D., Smirnov K. Russian handwritings tracked. Mendeley Data, V2, 2023. Dataset.
- [8]. Hugging Face, Convnext v2 – transformers documentation, 2025. Available at: https://huggingface.co/docs/transformers/model_doc/convnextv2, accessed 04.10.2025.
- [9]. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2 ed., 2009.
- [10]. Geman S., Bienenstock E., Doursat R. Neural networks and the bias/variance dilemma, *Neural Computation*, vol. 4, no. 1, pp. 1-58, 1992.
- [11]. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

- [12]. Yacenko D. Article17: репозиторий GitHub (релиз v1.0) [Электронный ресурс]. URL: github.com/d-yacenko/Article17.git (дата обращения 10.01.2026)
- [13]. Cohen G., Afshar S., Tapson J., van Schaik A. EMNIST: an extension of MNIST to handwritten letters, 2017. Available at: <https://arxiv.org/pdf/1702.05373>, accessed 10.01.2026.

Информация об авторах / Information about authors

Дмитрий Владимирович ЯЦЕНКО – старший преподаватель кафедры прикладной информатики и инноватики Института Высоких Технологий и Пьезотехники ЮФУ с 2020 года. Сфера научных интересов: машинное обучение, глубокие модели, криптографические методы защиты информации.

Dmitry Vladimirovich IATSENKO – senior lecturer in the Department of Applied Informatics and Innovation at the Institute of High Technologies and Piezotechnics at SFedU since 2020. His research interests include machine learning, deep learning models, and cryptographic methods for information security.

