

МЕТОД АВТОМАТИЧЕСКОГО ПОИСКА ОПТИМАЛЬНОГО МАСШТАБА ПРИ ПРИМЕНЕНИИ ОБУЧЕННЫХ МОДЕЛЕЙ АНАЛИЗА ГИСТОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЙ*

© 2023 г. М. А. Пенкин^{a,*} (ORCID: 0000000280279333),

А. В. Хвостиков^{a,**} (ORCID: 0000000242177141), А. С. Крылов^{a,***} (ORCID: 0000000199104501)

^aЛаборатория математических методов обработки изображений, Факультет вычислительной математики и кибернетики, Московский государственный университет имени М.В. Ломоносова, Россия, 119991, Москва, Ленинские горы, д. 1, стр. 52

*E-mail: penkin97@gmail.com

**E-mail: khvostikov@cs.msu.ru

***E-mail: kryl@cs.msu.ru

Поступила в редакцию 09.01.2023 г.

После доработки 15.01.2023 г.

Принята к публикации 20.01.2023 г.

Подготовка входных данных для нейронной сети является ключевым шагом для достижения высокой точности ее предсказаний. Известно, что сверточные нейронные модели обладают низкой инвариантностью к изменению масштаба входных данных. Так, обработка многомасштабных полнослайдовых гистологических изображений сверточными сетями естественным образом поднимает вопрос выбора оптимального масштаба обработки. В данной работе эта задача решается путем итеративного анализа расстояний, выдаваемых сверточным классификатором, до разделяющей гиперплоскости при различных входных масштабах. Предлагаемый метод проверен на предобученной на данных PATH-DT-MSU глубокой архитектуре DenseNet121, решающей задачу по-патчевой классификации полнослайдовых гистологических изображений.

DOI: 10.31857/S0132347423030032, EDN: DEIZNX

1. ВВЕДЕНИЕ

Значительные технологические достижения привели к появлению инновационных методов цифровой визуализации в гистопатологии — разделе медицины, изучающем процессы и состояния в тканях организма.

Поиск решений известных проблем световой микроскопии таких как: затраты на подготовку и хранение препаратов, потеря качества препаратов со временем, сложность организации дистанционной работы, — привел к стремительной эволюции взаимодействия с гистологическими препаратами, и за последние 100 лет был осуществлен переход от световой микроскопии к цифровой [1].

На данный момент одним из передовых решений является сбор и анализ данных в виде полнослайдовых изображений (*англ.* Whole slide images (WSI)) [2]. Пример полнослайдового изображения стенки желудка из набора данных PATH-DT-MSU¹ приведен на рис. 1. Снимок сделан на оп-

тическом увеличении (масштабе) $\times 40$ с разрешением 65809×99600 пикселей.

Технология WSI продолжает набирать популярность среди патологоанатомов в диагностических, образовательных [3] и исследовательских целях. В большинстве случаев, полнослайдовые изображения представляют собой многомасштабные пирамиды из нескольких уровней, каждый из которых хранит изображение одного и того же образца ткани в различном разрешении. Нижний уровень содержит RGB-изображение образца в максимальном оптическом разрешении. Очередной уровень пирамиды хранит последовательно уменьшенную копию нулевого уровня, вплоть до небольшого обзорного снимка. Каждый уровень, включая нулевой, представляет собой композицию, составленную из небольших фрагментов квадратной формы, видимые границы между которыми отсутствуют. Такая структура представления обеспечивает высокую скорость доступа к любому участку полнослайдового

* Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 22-41-02002.

¹ <https://imaging.cs.msu.ru/en/research/histology/path-dt-msu>

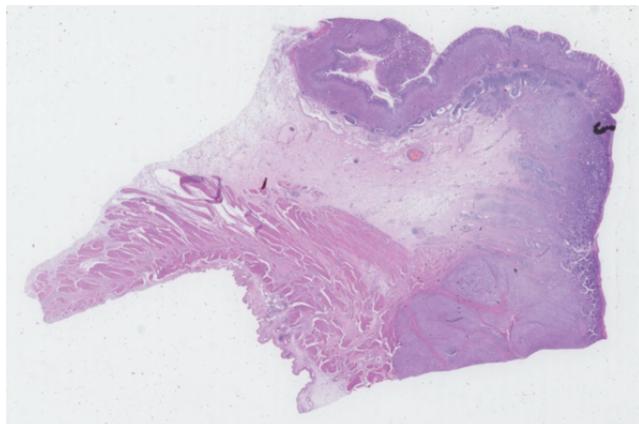


Рис. 1. Полнослайдовое гистологическое изображение стенки желудка из набора данных PATH-DT-MSU.

изображения на желаемом оптическом увеличении (масштабе).

Как любая технология, WSI имеет и недостатки [4, 5]. Основным недостатком WSI по сравнению со световой микроскопией является критическая необходимость в компьютерной автоматизированной обработке, без которой практическое использование полнослайдовых изображений в ручном режиме может оказаться неэффективным, дорогостоящим и требующим неприемлемо больших временных затрат. В зависимости от размера образца ткани размер изображений может достигать 100000×100000 пикселей и более.

В настоящее время в области компьютерной обработки изображений большое распространение приобрели методы глубокого обучения. В первую очередь — сверточные нейронные сети. Ввиду специфики медицинских изображений (отсутствие большого разнообразия выборки) обучение глубоких нейронных моделей крайне редко проводится с нуля, чаще всего используются предварительно

обученные архитектуры с дальнейшей тонкой настройкой под конкретную медицинскую задачу. Примерами успешного применения сверточных сетей в медицинской обработке изображений являются работы [6–8]. Сверточные нейронные сети отображают изображения в признаковые представления. В обученных признаковых пространствах большой размерности ведется необходимая обработка, например, классификация или сегментация.

По своей природе, сверточные модели инвариантны к преобразованию сдвига, однако, напротив, обладают низкой инвариантностью к преобразованию масштаба [9]. Так, задача выбора масштаба обработки (см. рис. 2), оптимального для выбранной обученной нейронной сети, становится особо важной при автоматизации анализа многомасштабных полнослайдовых изображений, а также при разработке комплексных систем работы с гистологическими изображениями [10].

В данной работе предлагается алгоритм автоматического поиска величины оптимального масштаба при обработке полнослайдовых гистологических изображений сверточными нейронными сетями. Предлагаемый алгоритм протестирован на задаче классификации полнослайдовых гистологических изображений и основан на анализе расстояний до разделяющих гиперплоскостей при различных входных масштабах.

2. АКТУАЛЬНОСТЬ

Оценить важность обработки полнослайдовых гистологических изображений на оптимальном для выбранной нейронной модели масштабе можно на примере предобученного классификатора опухолей желудка на базе DenseNet121 [11].

Алгоритм обучен [12] классификации 5 типов тканей, имеющихся в наборе данных PATH-DT-MSU (WSS1, WSS2):

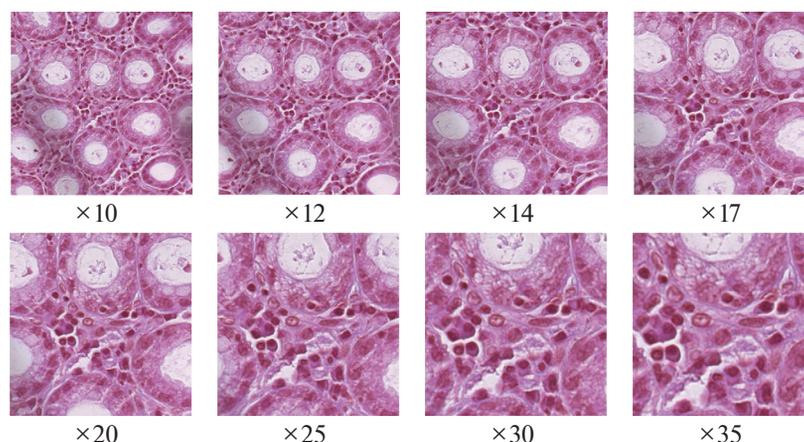


Рис. 2. Участок полнослайдового гистологического изображения, рассмотренный в различных масштабах.

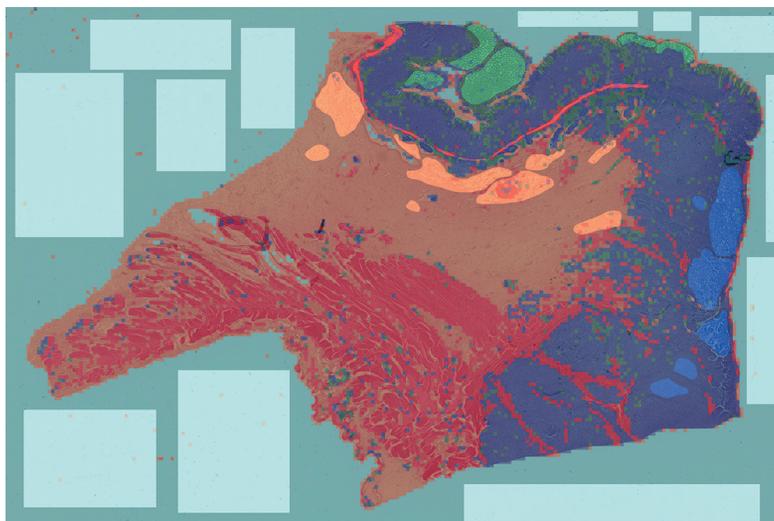


Рис. 3. Визуализация по-патчевой классификации полнослайдового гистологического изображения сверточной сетью DenseNet121. Области, размеченные специалистами, выделены увеличенной яркостью.

1. участки подслизистой основы, собственный мышечный слой желудка и субсерозные отделы (АТ);
2. фон изображения (BG);
3. неизменные участки собственной пластинки слизистой оболочки (LP);
4. неизменные участки мышечной пластинки слизистой оболочки (ММ);
5. участки аденокарциномы желудка (TUM).

Обучающая и валидационная выборки были сформированы из RGB фрагментов (224×224), вырезанных из WSI на масштабе $\times 20$.

Высокое разрешение полнослайдовых изображений затрудняет применение методов глубокого обучения *напрямую*, как при классификации или сегментации обычных изображений архитектурами ResNet18 [13], VGG19 [14], InceptionNet [15], DenseNet121 [11], UNet [6]. Поэтому рассматриваемый классификатор опухолей желудка осуществляет *по-патчевую* классификацию (см. рис. 3):

- полнослайдовое изображение разделяется на небольшие фрагменты – патчи;
- сверточная нейронная сеть присваивает каждому фрагменту метку соответствующего класса;
- полученный набор классифицированных фрагментов склеивается в единое изображение.

Таким образом получается *грубая* карта сегментации, точность которой можно увеличить, уменьшая шаг окна (*англ.* stride) и усредняя предсказания.

В табл. 1 приведены значения точности классификации полнослайдового гистологического изображения на оптимальном увеличении (масштаб, на котором велось обучение) $\times 20$, на меньшем увеличении – $\times 10$ и на большем увеличении

– $\times 40$. При выборе масштаба $\times 10$, отличного от оптимального для выбранной сверточной модели, наблюдается значительное снижение точности детекции класса LP и, как следствие, падение средней точности классификации. А при выборе масштаба $\times 40$, также отличного от оптимального, падение точности классификации еще более существенно и наблюдается на классах LP и MM.

На практике, в отличие от рассмотренного примера, информация об оптимальном масштабе для выбранной модели может быть недоступна или же может быть утеряна информация об оптическом увеличении, на котором было получено анализируемое WSI изображение. Подобные риски усиливают практическую актуальность поставленной задачи.

Проблема выбора оптимального масштаба обработки встречается не только при попытках применить предобученную нейронную модель к полнослайдовым медицинским изображениям, но и при работе с обычными изображениями. К примеру, авторы [16] убедились в низкой инвариантности сверточных классификаторов ResNet50, ResNet18 [13], VGG19 [14], AlexNet [17] к изменению масштаба входных изображений из набора ImageNet [18].

Таблица 1. Точность классификации полнослайдового гистологического изображения в зависимости от входного масштаба (μ – средняя точность по пяти классам)

	АТ	BG	LP	ММ	TUM	μ
$\times 10$	87.80	99.84	61.74	17.47	98.19	73.01
$\times 20$	84.42	99.58	82.41	15.74	98.04	76.04
$\times 40$	78.85	99.46	15.01	7.35	94.52	59.04

Сверточный нейронный классификатор можно разделить на два блока:

- блок нелинейного отображения изображений в многомерное признаковое пространство;
- блок разделяющих гиперплоскостей, осуществляющих классификацию признаков представлений на заданное число классов.

Предлагаемый в статье [16] алгоритм выбора масштаба сосредоточен на анализе процессов внутри блока нелинейного отображения. Алгоритм опирается на принцип сверточного фильтра [19] – одного из основных компонентов сверточных нейронных сетей. Этот принцип утверждает, что высокая корреляция фильтра с очередным паттерном сигнализирует об успехе распознавания. Таким образом, в работе [16] формулируется гипотеза, что изображения в оптимальном масштабе провоцируют экстремум активаций внутренних слоев. Выдвинутая гипотеза эмпирически доказывается и проверяется на задаче классификации как обычных изображений, так и медицинских гистологических изображений.

Данная работа, в отличие от исследования [16], обращает свое внимание на анализ процессов в блоке разделяющих гиперплоскостей при различных масштабах входных изображений.

3. ПРЕДЛАГАЕМЫЙ МЕТОД

В настоящей работе решение о выборе оптимального масштаба обработки полнослайдовых гистологических изображений принимается путем анализа расстояний до разделяющих гиперплоскостей при различных входных масштабах (см. Алгоритм 1).

Алгоритм 1

Вход:

- I_{WSI} – WSI изображение;
 - $\mathcal{A}[\Phi, W]$ – сверточный классификатор гистологических RGB-изображений, имеющих размер $\varepsilon \times \varepsilon \times 3$.
 - Φ – нелинейное отображение изображений размером $\varepsilon \times \varepsilon \times 3$ в признаковое пространство размерности M .
 - W – матрица $\mathbb{R}^{M \times C}$, определяющая C разделяющих гиперплоскостей классификатора \mathcal{A} .
- 1: $\mathcal{S} \leftarrow \{s_1, \dots, s_N\}, |\mathcal{S}| = N$.
 - 2: $\mathcal{P} \leftarrow \{x_k\}_{k=1}^P: x_k \in I_{WSI}, k \in \overline{1..P}$.
 - 3: **for** $k \leftarrow \overline{1..P}$ **do**
 - 4: **for** $i \leftarrow \overline{1..N}$ **do**
 - 5: $x_k^\varepsilon(s_i) \leftarrow U(x_k(s_i); \varepsilon)$.

$$6: \quad f_i^{(k)} \leftarrow \Phi(x_k^\varepsilon(s_i)): f_i^{(k)} \in \mathbb{R}^M.$$

7: **end for**

$$8: \quad X^{(k)} \leftarrow \{f_i^{(k)}\}_{i=1}^N: X^{(k)} \in \mathbb{R}^{N \times M}.$$

$$9: \quad D^{(k)} \leftarrow \frac{X^{(k)} \cdot W}{\|W\|}: D^{(k)} \in \mathbb{R}^{N \times C}.$$

$$10: \quad \hat{s}^{(k)} \leftarrow \operatorname{argmax}(\operatorname{ReduceMax}(D^{(k)}, \text{axis} = -1)).$$

$$11: \quad \hat{\mathcal{S}} \leftarrow \hat{\mathcal{S}} \cup \hat{s}^{(k)}$$

12: **end for**

Выход:

– $\hat{\mathcal{S}} = \{\hat{s}^{(k)}\}_{k=1}^P$ – множество предлагаемых масштабов обработки изображения I_{WSI} по P -патчам.

В качестве входа предлагаемый алгоритм получает: полнослайдовое изображение I_{WSI} , обученный классификатор \mathcal{A} RGB-фрагментов гистологических изображений и число патчей P , многомасштабный анализ которых формирует результирующее множество предлагаемых масштабов $\hat{\mathcal{S}} = \{\hat{s}^{(k)}\}_{k=1}^P$. Поиск оптимальной величины масштаба обработки ведется среди масштабов $\mathcal{S} = \{s_1, \dots, s_N\}$, где s_1 – наименьшее оптическое увеличение, а s_N – наибольшее.

На полнослайдовом изображении I_{WSI} случайным образом выбирается P точек $\{x_k\}_{k=1}^P$.

Вокруг очередной точки x_k на каждом рассматриваемом масштабе s_i вырезается квадратный фрагмент $x_k^\varepsilon(s_i)$ формы $\varepsilon \times \varepsilon \times 3$. Признаковые представления $\{f_i^{(k)} \in \mathbb{R}^M\}_{i=1}^N$ патча x_k^ε в N масштабах образуют матрицу *объектов-признаков* $X^{(k)} \in \mathbb{R}^{N \times M}$.

Обработка фрагмента x_k^ε , представленного в N масштабах, завершается вычислением матрицы расстояний $D^{(k)}$ до каждой из C разделяющих гиперплоскостей. Масштаб $\hat{s}^{(k)}$, отвечающий наибольшему расстоянию, сохраняется в множество предлагаемых масштабов $\hat{\mathcal{S}}$, и алгоритм переходит к следующей точке множества \mathcal{P} .

Окончательное решение об оптимальном масштабе обработки полнослайдового изображения I_{WSI} принимается голосованием по множеству предлагаемых масштабов $\hat{\mathcal{S}}$ обработки фрагментов $\{x_k^\varepsilon\}_{k=1}^P$.

Предлагаемый метод мотивирован геометрическим смыслом линейного классификатора $\langle \omega, x \rangle$, инкапсулированного в большинство сверточных нейронных классификаторов. Геометрически линейный классификатор соответствует гиперплоскости с вектором нормали ω . Величина

скалярного произведения $\langle \omega, x \rangle$ пропорциональна расстоянию от гиперплоскости до точки x , а его знак показывает, с какой стороны от гиперплоскости находится данная точка. Таким образом, линейный классификатор разделяет пространство на две части, и при этом одно полупространство он относит к положительному классу, а другое — к отрицательному. Нетрудно показать [20], что расстояние от точки x до разделяющей гиперплоскости ω вычисляется как $\rho(x) = \frac{|\langle \omega, x \rangle|}{\|\omega\|}$.

В сверточные нейронные классификаторы встроены C разделяющих гиперплоскостей, так что вектор ω превращается в матрицу W , столбцы которой задают нормали этих гиперплоскостей, а скалярное произведения — в матричное умножение.

4. РЕЗУЛЬТАТЫ

Представленный алгоритм поиска оптимального масштаба реализован на языке программирования Python 3 с использованием библиотеки глубокого обучения TensorFlow 2 и библиотеки OpenSlide для работы с полнослайдовыми изображениями.

Метод протестирован на задаче классификации полнослайдовых гистологических изображений стенок желудка из набора данных PATH-DT-MSU (WSS1, WSS2) обученной сверточной нейронной сетью DenseNet121 на масштабе $\times 20$. Параметры алгоритма (см. Алгоритм 1) в таком случае принимают следующие значения: $\mathcal{S} = \{\times 10, \times 12, \times 14, \times 17, \times 20, \times 25, \times 30, \times 35, \times 43\}$, $\mathcal{A} := \text{DenseNet121}$, $M = 1024$, $C = 5$, $\epsilon = 224$. Множество \mathcal{S} составлено так, что $\frac{s_{i+1}}{s_i} \approx 1.2, i = 1 \dots N - 1$.

Визуализации применения алгоритма к пяти различным многомасштабным патчам, относящимся к пяти различным классам (AT, BG, LP, MM, TUM) и вырезанных из полнослайдового изображения (рис. 1), представлены на рис. 4.

Для всех пяти фрагментов наибольший отступ (англ. margin) порождается классификатором именно от гиперплоскости, отвечающей референсному классу. Причем наблюдается заметная разница между величинами отступов от референсной гиперплоскости и от всех остальных.

Низкая устойчивость распознавания класса LP к изменению входного масштаба, отмеченная в секции 2 (см. табл. 1), находит свое отражение и

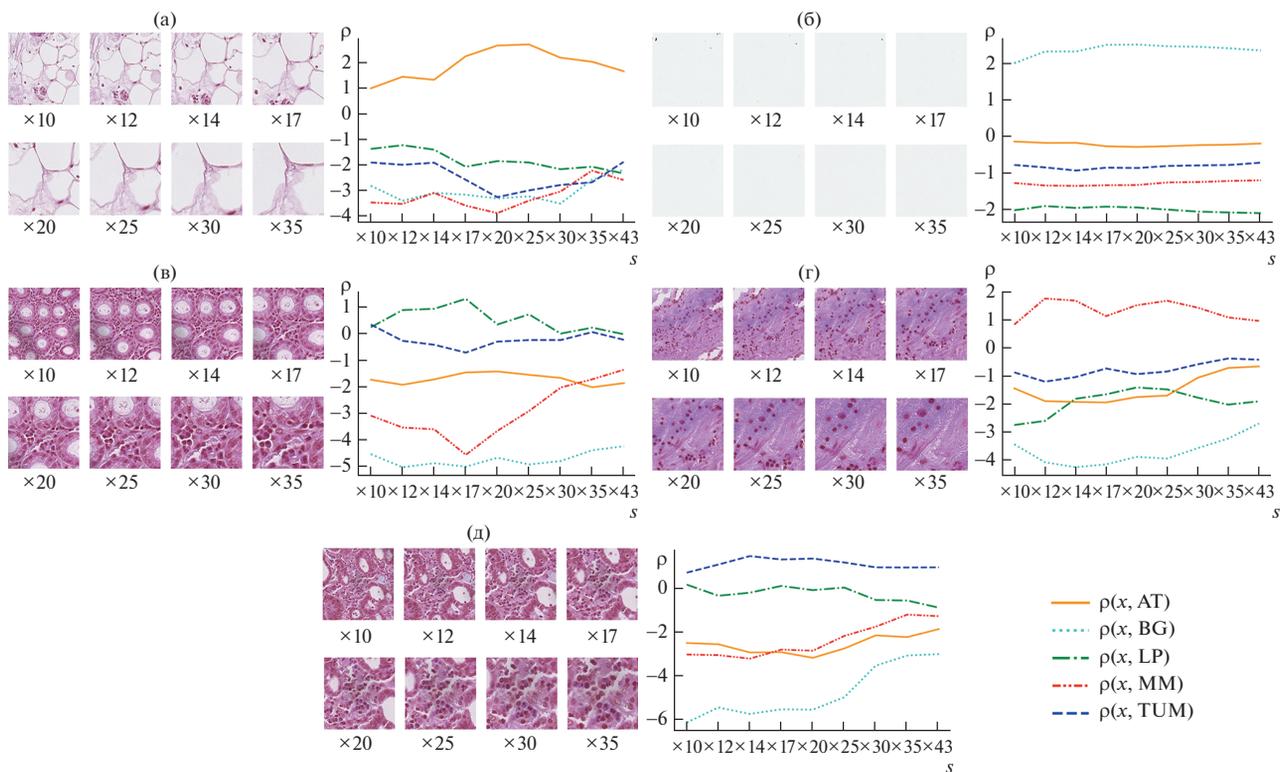


Рис. 4. Визуализация работы алгоритма на различных многомасштабных фрагментах полнослайдового изображения. На графиках показаны зависимости расстояний (ρ) до разделяющих гиперплоскостей в зависимости от масштаба (s) обрабатываемого фрагмента. (а) — фрагмент класса AT; (б) — фрагмент класса BG; (в) — фрагмент класса LP; (г) — фрагмент класса MM; (д) — фрагмент класса TUM.

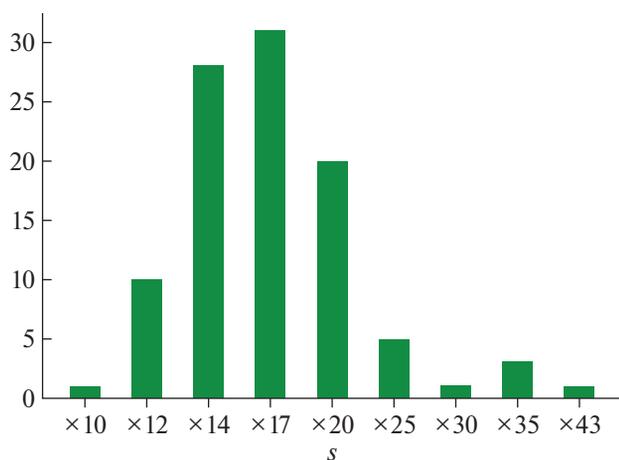


Рис. 5. Гистограмма по множеству предложенных алгоритмом масштабов $\hat{\mathcal{S}}$ для 100 случайных LP-патчей.

на рис. 4 (в): достигая экстремума на масштабе $\times 17$ (близком к истинному оптимальному масштабу $\times 20$, на котором велось обучение), кривая расстояний до референсной гиперплоскости существенно сближается с кривой расстояний до гиперплоскости TUM, при обработке патча на большем или меньшем масштабе в сравнении с $\times 17$.

В зависимости от специфики текстур патча и ширины множества поиска \mathcal{S} (под шириной здесь понимается $|s_N - s_1|$) варьируется характер зависимостей расстояний до гиперплоскостей от масштаба:

- сильная выпуклость (рис. 4 (а, в)) — алгоритм способен сделать однозначный вывод об оптимальном масштабе с высокой уверенностью, ширина множества \mathcal{S} выбрана верно;
- монотонность (рис. 4 (г, д)) — алгоритм способен сделать некоторый вывод об оптимальном масштабе, однако следует увеличить ширину множества поиска \mathcal{S} ;
- почти константность (рис. 4 (б)) — анализируемые масштабы равнозначны.

На рис. 5 представлена гистограмма по множеству предложенных алгоритмом масштабов $\hat{\mathcal{S}}$ для 100 случайных патчей, размеченных специалистами, как LP-участки WSI-изображения. Класс LP выбран ввиду особо низкой устойчивости модели DenseNet121 к распознаванию LP-фрагментов в различных масштабах (см. табл. 1), а значит высокой актуальности предлагаемого алгоритма для анализа многомасштабных LP-фрагментов WSI-изображения. Топ-3 предлагаемых масштаба: $\times 17$, $\times 14$ и $\times 20$, соответственно. В доверительный интервал истинного оптимального масштаба ($\times 17 \leftarrow \times 20 \rightarrow \times 25$) попало 56% всех прогнозов из $\hat{\mathcal{S}}$. В более широкий доверительный интервал истинного оптимально-

го масштаба ($\times 14 \leftarrow \times 20 \rightarrow \times 30$) попало 85% всех прогнозов из $\hat{\mathcal{S}}$.

5. ЗАКЛЮЧЕНИЕ

Данная работа обращает внимание на низкую инвариантность сверточных нейронных сетей к изменению масштаба входных данных и предлагает автоматический метод выбора оптимального масштаба обработки, что особенно актуально при анализе многомасштабных полнослайдовых гистологических изображений.

Условно разделив сверточную нейронную сеть на два блока: блок нелинейного отображения и блок разделяющих гиперплоскостей, — данная работа представляет метод выбора масштаба на основе анализа расстояний до разделяющих гиперплоскостей.

Предлагаемый алгоритм протестирован на задаче классификации полнослайдовых изображений стенки желудка из набора данных PATH-DT-MSU (WSS1, WSS2) обученной сверточной нейронной сетью DenseNet121.

Работа в данном направлении продолжается, и одним из возможных вариантов развития алгоритма является его адаптация под иерархическую композицию разделяющих гиперплоскостей.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научно-го проекта № 22-41-02002.

СПИСОК ЛИТЕРАТУРЫ

1. Park S., Pantanowitz L., Parwani A.V. Digital imaging in pathology // Clinics in laboratory medicine. 2012. М. 32. № 4. С. 557–584.
2. Pantanowitz L., Valenstein P.N., Evans A.J., Kaplan K.J., Pfeifer J.D., Wilbur D.C., Collins L.C., Colgan T.J. Review of the current state of whole slide imaging in pathology // Journal of pathology informatics. 2012. V. 2. № 1. P. 36.
3. Saco A., Bombi J.A., Garcia A., RamGrez J., Ordi J. Current status of whole-slide imaging in education // Pathobiology. 2016. V. 83. № 2–3. P. 79–88.
4. Farahani N., Parwani A.V., Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives // Pathol Lab Med Int. 2015. V. 7. № 23–33. P. 4321.
5. Rojo M.G., Garcia G.B., Mateos C.P., Garcia J.G., Vicente M.C. Critical comparison of 31 commercially available digital slide systems in pathology // International journal of surgical pathology. 2006. V. 14. № 4. P. 285–305.
6. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // In International Conference on Medical image com-

- puting and computer-assisted intervention. 2015. P. 234–241.
7. *Khvostikov A., Krylov A.S., Mikhailov I., Malkov P.* CNN Assisted Hybrid Algorithm for Medical Images Segmentation // In Proceedings of the 2020 5th International Conference on Biomedical Signal and Image Processing. 2020. P. 14–19.
 8. *Getmanskaya A.A., Sokolov N.A., Turlapov V.E.* Multi-class U-Net Segmentation of Brain Electron Microscopy Data Using Original and Semi-Synthetic Training Datasets // Programming and Computer Software. 2022. V. 48. № 3. P. 164–171.
 9. *Gong Y., Wang L., Guo R., Lazebnik S.* Multi-scale orderless pooling of deep convolutional activation features // In European conference on computer vision. 2014. P. 392–407.
 10. *Khvostikov A.V., Krylov A.S., Mikhailov I.A., Malkov P.G.* Visualization of Whole Slide Histological Images with Automatic Tissue Type Recognition // Pattern Recognition and Image Analysis. 2022. V. 32. № 3. P. 483–488.
 11. *Huang G., Liu Z., Van Der Maaten L., Weinberger K.Q.* Densely connected convolutional networks // In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. P. 4700–4708.
 12. *Kingma D.P., Ba J.* Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014.
 13. *He K., Zhang X., Ren S., Sun J.* Deep residual learning for image recognition // Proceedings of the CVPR IEEE Conference. 2016. P. 770–778.
 14. *Simonyan K., Zisserman A.* Very deep convolutional networks for large-scale image recognition // arXiv preprint arXiv:1409.1556. 2014.
 15. *Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A.* Going deeper with convolutions // In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015. P. 1–9.
 16. *Penkin M.A., Khvostikov A.V., Krylov A.S.* Optimal Input Scale Transformation Search for Deep Classification Neural Networks // In Graphicon-Conference on Computer Graphics and Vision. 2022. V. 32. P. 668–677.
 17. *Krizhevsky A., Sutskever I., Hinton G.E.* Imagenet classification with deep convolutional neural networks // Communications of the ACM. 2017. V. 60. № 6. P. 84–90.
 18. *Deng J., Dong W., Socher R., Li L.J., Li K., Fei-Fei L.* Imagenet: A large-scale hierarchical image database // In 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009. P. 248–255.
 19. *Bolme D.S., Beveridge J.R., Draper B.A., Lui Y.M.* Visual object tracking using adaptive correlation filters // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010. P. 2544–2550.
 20. *Mohri M., Rostamizadeh A., Talwalkar A.* Foundations of machine learning. MIT Press, 2018. 475 p.