Применение технологий глубокого обучения для детекции пиков яркости на биспиральных изображениях флуоресцирующих точечных излучателей

А.В. Наумов [1], А.А. Роженцов [2], А.А. Баев [2], К.О. Иванов [2]
[1] Институт спектроскопии Российской академии наук, 108840, Россия, г. Москва, г. Троицк, ул. Физическая, д. 5;
[2] Поволжский государственный технологический университет, 424000, Россия, г. Йошкар-Ола, пл. Ленина, д. 3

Аннотация

В статье рассмотрены подходы к формированию обучающей выборки, особенности представления исходных данных и архитектур нейронных сетей при решении задач локализации дифракционных пиков на биспиральных изображениях флуоресцирующих точечных излучателей (молекул, квантовых точек и других) в трехмерной локализационной микроскопии сверхвысокого пространственного разрешения с использованием оптики вращающихся световых полей. Показано преимущество вычисления субпиксельных координат пиков яркости по тепловым картам с помощью операции soft-argmax и использование предобученных энкодеров при ограниченном объёме данных. В частности, средняя абсолютная ошибка (MAE) локализации для модели с предобученным энкодером ConvNeXt-Tiny и операцией soft-argmax составила 0,6 пикселя, что примерно в 5 раз меньше, чем при прямой регрессии координат с тем же энкодером (3,08 пикселя), и в 10 раз меньше, чем у сверточной нейронной сети со случайно инициализированными весами (6,03 пикселя). Для энкодеров EfficientNet-B0 и ResNet-18 с операцией soft-argmax MAE составила 0,92 и 2,66 пикселя соответственно.

Ключевые слова: биспиральные изображения, глубокое обучение, искусственный интеллект, компьютерное зрение, машинное обучение, наноскопия, сверточные нейронные сети, цифровая микроскопия.

Цитирование: **Наумов, А.В.** Применение технологий глубокого обучения для детекции пиков яркости на биспиральных изображениях флуоресцирующих точечных излучателей / А.В. Наумов, А.А. Роженцов, А.А. Баев, К.О. Иванов // Компьютерная оптика. – 2026. – Т. 50, № 5. – 1919. – doi: 10.18287/COJ1919.

Citation: Naumov AV, Rozhentsov AA, Baev AA, Ivanov KO. Application of Deep Learning Technologies for Brightness Peak Detection in Bispiral Images of Fluorescent Point Emitters. Computer Optics 2026; 50(5): 1919. doi: 10.18287/COJ1919.

Введение

Применение методов обработки данных с использованием глубокого обучения в различных областях знаний позволяет значительно расширить возможности исследователей. В области трехмерной локализационной микроскопии сверхвысокого пространственного разрешения (трехмерной наноскопии) актуальной является задача обработки биспиральных изображений [1, 2] флуоресцирующих точечных излучателей. К таким излучателям относятся одиночные молекулы, квантовые точки, центры окраски в кристаллах, размеры которых существенно меньше длины волны флуоресценции. В данном подходе используется преобразование пучков света во вращающиеся световые поля с помощью специально рассчитанных фазовых масок. Угол поворота в биспиральном изображении, состоящем из двух дифракционно ограниченных пятен (лепестков), характеризует значение продольной координаты одиночного излучателя, а положение центра тяжести изображения позволяет определить продольные координаты излучателя с субдифракционной точностью. На основе обработки таких изображений можно получать трехмерные реконструкции микроскопических объектов с нанометровым разрешением.

В результате проведения эксперимента формируется серия изображений, в каждом из которых наблюдается некоторое количество двулепестковых изображений от одиночных точечных излучателей, флуоресцирующих случайным образом. Сложность обработки подобных изображений заключается в том, что они обладают существенным разбросом возможных углов поворота в двулепестковых изображениях. Форма лепестков также может существенно меняться при изменении продольной координаты, изображения от различных точечных излучателей могут пересекаться, а часть биспиральных изображений может выходить за пределы кадра (рис. 1). Таким образом, для решения задачи анализа биспиральных изображений одновременно должны решаться задачи обнаружения, разрешения и оценки параметров двулепестковых изображений. Методы, базирующиеся на формировании и обработке биспиральных изображений одиночных точечных излучателей, представлены в работах [1 – 4], однако им присущи такие ограничения, как низкая точность локализации при высоком уровне шума, ограниченная способность обрабатывать перекрывающиеся отклики и высокая вычислительная сложность.

В подобных случаях именно применение нейросетевых методов обработки может дать положительный результат. Основными моделями глубокого обучения для обработки и анализа изображений в настоящее время

являются сверточные нейронные сети (CNN) [5, 6]. За счет обучения на больших объемах данных и сложности архитектуры они демонстрируют повышенную устойчивость к артефактам и способность к разделению перекрывающихся объектов на изображении [7, 8].

Задачей данной работы является сравнительный анализ методов глубокого обучения на основе сверточных нейронных сетей для субпиксельной локализации центров пиков яркости в биспиральных наноскопических изображениях флуоресцирующих нанокристаллов. При этом будут рассмотрены подходы на основе прямой регрессии координат [9], вычислении взвешенных сумм значений тепловых карт с помощью функции soft-argmax [10] и комбинированных методов с использованием нескольких независимых голов нейронной сети [11]. В качестве энкодеров применяются сверточные нейронные сети со случайно инициализированными весами, а также предобученные модели ResNet [12], EfficientNet [13] и ConvNeXt [14] из библиотеки torchvision.

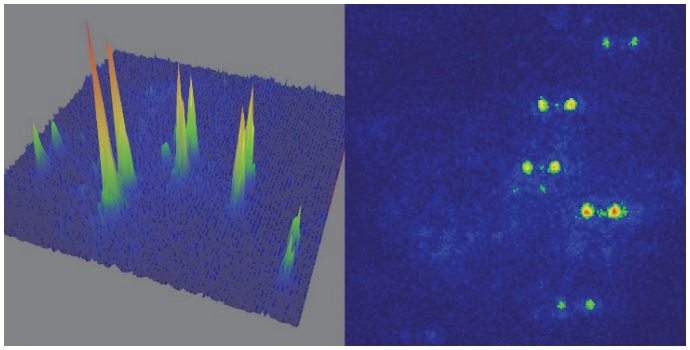


Рис. 1. Примеры биспиральных изображений

Формирование обучающей выборки

Формирование достаточного объема обучающей выборки на реальных экспериментальных данных при разработке нейронной сети для обработки наноскопических изображений представляет собой сложную задачу ввиду случайного характера «включения» точечных излучателей, большого объема получаемых видеоданных, требующих разметки, значительного количества различных возможных комбинаций взаимного расположения отметок, которые впоследствии могут возникнуть при обработке реальных данных и к восприятию которых должна быть готова нейронная сеть. Все указанные факторы могут быть учтены при формировании обучающей выборки путем моделирования реальных изображений.

Модель биспирального изображения точечных отметок в n-м кадре может быть задана с помощью суммы двух гауссовских пиков [15 – 18]:

$$\mathrm{d}\left(A,\mathbf{X},\overline{\mathbf{X}},\mathbf{D},\phi(z)\right)=\mathrm{f}\left(A,\mathbf{X},\overline{\mathbf{X}}+\mathbf{P},\mathbf{D},\phi(z)-\tfrac{\pi}{2}\right)+\mathrm{f}\left(A,\mathbf{X},\overline{\mathbf{X}}-\mathbf{P},\mathbf{D},\phi(z)\phi(z)-\tfrac{\pi}{2}\right),$$

где $\mathrm{f}\left(A,\mathbf{X},\overline{\mathbf{X}},\mathbf{D},\phi(z)\right)=\frac{A}{2\pi\sqrt{\boldsymbol{D}}}\exp\left(-\frac{1}{2}\cdot\left(\mathbf{R}\cdot\left(\mathbf{X}-\overline{\mathbf{X}}\right)\right)^{\mathrm{T}}\cdot\mathbf{D}^{-1}\cdot\left(\mathbf{R}\cdot\left(\mathbf{X}-\overline{\mathbf{X}}\right)\right)\right)$.

В свою очередь, A – амплитуда отклика отметки в очередном кадре, $\mathbf{D}=\begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{bmatrix}$ – матрица, задающая ширину отметки по осям x и y, $\mathbf{X}$ – вектор координат точки, в которой вычисляется значение функции, $\overline{\mathbf{X}}$ – вектор координат отметки, для которой формируется модель биспирального изображения, $\phi(z)$ – угол поворота биспирального изображения, однозначно связанный с координатой z точечного излучателя, $\mathbf{R}=\begin{bmatrix}\cos(\varphi(z)) & \sin(\varphi(z))\\ -\sin(\varphi(z)) & \cos(\varphi(z))\end{bmatrix}$ – матрица вращения, $\mathbf{P}=\mathbf{R}\cdot\begin{bmatrix}R\\0\end{bmatrix}$ – вектор сдвига отметок в биспиральном изображении относительно координат исходного точечного излучателя, R – радиус биспирального изображения (рис. 2).

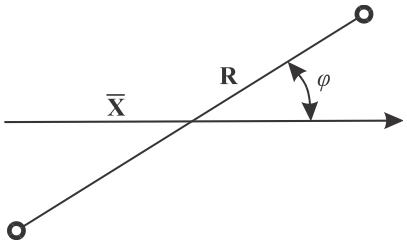


Рис. 2. Параметры биспирального изображения

В связи с тем, что при формировании биспирального изображения возможно дополнительное «размывание» изображения отметки вдоль линии окружности с радиусом R, в матрице $\mathbf{D}$ может быть задана соотвествующая

функциональная зависимость $\sigma_x^2 = F\big(\varphi(z)\big)$. На рис. 3 представлены примеры отдельных отметок биспирального изображения.

Для моделирования тестовых изображений был сформирован набор трехмерных объектов в виде облаков точек с заданными в трехмерном пространстве координатами x, y, z. На рис. 4 представлен пример изображения такого объекта.

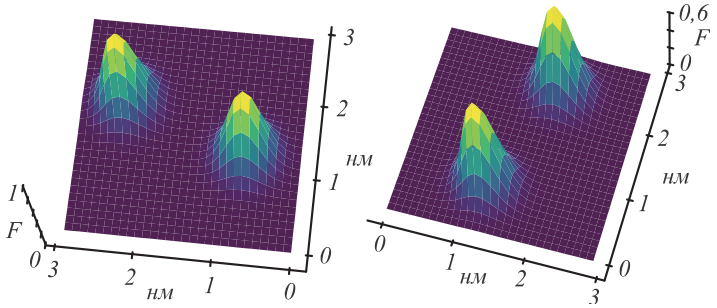


Рис. 3. Примеры моделирования отдельных отметок биспирального изображения

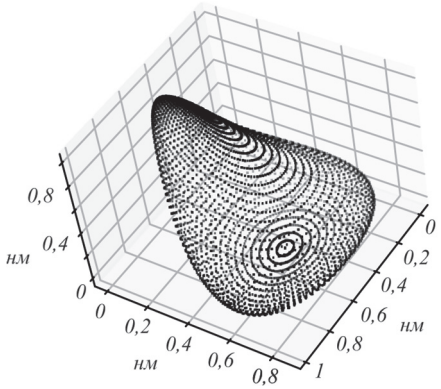


Рис. 4. Пример трехмерного изображения, заданного облаком точек

При формировании каждого кадра случайным образом с учетом заданной вероятности появления каждой отметки осуществляется выборка точек из исходного облака, и с учетом их параметров формируется биспиральное изображение (рис. 5).

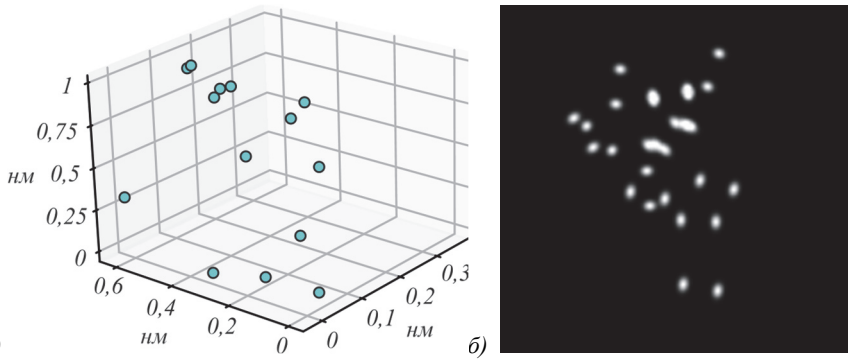


Рис. 5. Отображение выборки точек в трехмерном пространстве (а) и сформированное модельное биспиральное изображение (б)

При необходимости может осуществляться зашумление изображений тестовой выборки с учетом статистических характеристик шумов в реальных изображениях.

Исследуемые архитектуры нейронных сетей

Сверточная нейронная сеть преобразует двумерное изображение в трехмерный тензор карт признаков. При этом каждая 2D-карта, полученная путем применения фильтра свертки, отражает наличие и выраженность визуальных паттернов, что эффективно для задач классификации, обнаружения объектов, семантической сегментации и сегментации экземпляров [6].

Одним из наиболее очевидных подходов к получению координат центров откликов на наноскопическом изображении является добавление нескольких полносвязных слоев к выходу сверточной нейронной сети, чтобы полученная модель решала задачу регрессии для оценки координат. При этом последний полносвязный слой

должен содержать $2 \cdot N$ нейронов, где N – максимальное число откликов на изображении, а в выходном тензоре будут чередоваться горизонтальные и вертикальные координаты центров откликов. Если текущее количество откликов на изображении M меньше максимального значения N, то оставшимся элементам выходного тензора присваиваются значения, равные – 1. Кроме того, для лишних $N - M$ точек можно не вычислять функцию потерь за счет использования маски, которая обнуляет ошибку для неиспользуемых позиций [9, 19, 20].

Помимо прямой регрессии координат, для оценки координат центров откликов на изображении может использоваться операция soft-argmax. В этом случае на выходе энкодера в виде сверточной нейронной сети необходимо получить N карт признаков (N по-прежнему равно максимальному числу откликов на изображении) и к каждой карте признаков применить эту операцию. В результате для каждой карты признаков будет получена оценка координат центра одного отклика. Работа операции soft-argmax заключается в последовательном применении операции softmax к k-й карте признаков и вычислении взвешенных оценок координат $(\hat{y}_k, \hat{x}_k)$ центра отклика на текущей карте:

$$\hat{y}_k = \sum_{i=0}^{H-1}\sum_{j=0}^{W-1} i \cdot soft_k(i,j),\ \hat{x}_k = \sum_{i=0}^{H-1}\sum_{j=0}^{W-1} j \cdot soft_k(i,j),$$

где $soft_k$ – результат применения операции softmax к k-й карте признаков, H и W – высота и ширина карты признаков изображения соответственно. Далее полученные координаты масштабируются к размеру исходного изображения путем умножения на коэффициент s, показывающий, во сколько раз карта признаков меньше размера исходного изображения [21, 22].

Другой подход к определению координат центров пиков яркости основан на использовании многоцелевой нейронной сети, включающей CNN-энкодер, слой soft-argmax и специализированные головы для вычисления субпиксельного смещения координат и оценки вероятности наличия пиков яркости на изображении [21 – 24].

Поскольку во всех рассмотренных подходах к оценке координат центров пиков яркости на наноскопических изображениях используются карты признаков, полученные с помощью сверточных нейронных сетей, то для получения этих карт признаков также могут использоваться различные предобученные модели, такие как ResNet, EfficientNet и ConvNeXt [12 – 14]. При этом вычисление центров пиков яркости по полученным картам признаков также может выполняться с помощью прямой регрессии координат, операции soft-argmax или с помощью многоголовой нейронной сети. Также отметим, что для повышения точности предсказанных координат центров пиков яркости с использованием всех рассмотренных подходов возможно применение постобработки, например, гауссовой интерполяции [25].

На начальном этапе исследований для оценки способности сверточных нейронных сетей извлекать признаки из данных были обучены модели со случайно инициализированными весами, без использования предобученных архитектур. Модели включали блоки прямой регрессии координат (рис. 6*а*), операцию soft-argmax (рис. 6*б*) и отдельную голову для вычисления субпиксельного смещения координат пиков яркости (рис. 6*в*). Энкодеры всех моделей состояли из четырех сверточных слоев с функциями активации ReLU [6] и двух слоев субдискретизации (pooling). Эта конфигурация была признана оптимальной на основе эмпирической оценки альтернативных архитектур с различным количеством уровней. Также установлено предпочтительное использование адаптивного усредняющего пуллинга (AdaptiveAvrPool) [6] вместо слоя выравнивания (Flatten) [6] на выходе сверточной нейронной сети. Это позволило не только уменьшить количество параметров полносвязных слоев, которые отвечают за непосредственное определение координат центров откликов, но и повысить точность определения координат пиков яркости на изображении. Модель с адаптивным усредняющим пуллингом в одинаковых условиях справлялась с задачей лучше, чем модель со слоем выравнивания.

Каждая нейронная сеть выдает координаты центров откликов в формате $\mathbf{c} = [[y_0, x_0], [y_1, x_1], \ldots, [y_{N-1}, x_{N-1}]]$ и вектор вероятностей присутствия соответствующих откликов на изображении $\mathbf{p} = [p_0, p_1, \ldots, p_{N-1}]$. Наличие дополнительной головы, вычисляющей вектор вероятностей **p,** обоснованно неопределенным количеством откликов. Порог вероятности, равный 0,5, выбран как стандартное значение для бинарной классификации, поскольку он представляет собой точку равновесия между положительным и отрицательным классами, позволяя эффективно разделять значимые отклики от шума. Низкие значения $p_i < 0{,}5$ указывают, что соответствующие координаты можно игнорировать.

Для обучения нейронной сети применялась комбинированная функция потерь, представляющая собой сумму потерь для каждой из голов модели. Голова, ответственная за определение координат центров откликов на изображении, оптимизировалась с использованием среднеквадратичной ошибки (MSE) [6]. Для головы, вычисляющей вектор вероятностей p, использовалась бинарная кросс-энтропия (BCE) [6]. В модели, представленной на рис. 6*в*, добавлена дополнительная голова для оценки вектора смещения координат центров откликов, для которой также применялась среднеквадратичная ошибка (MSE) в качестве функции потерь.

Также в работе были исследованы модели, где в качестве энкодеров выступали предобученные на базе данных ImageNet [27] нейронные сети ResNet18, EfficientNet-B0 и ConvNeXt-Tiny (рис. 7). При этом головы, отвечающие за вычисление координат, вектора оценки вероятностей наличия всплеска и вектора субпиксельного смещения остались без изменения, как для моделей со случайно инициализированными весами.

Для энкодера на базе ResNet18 (рис. 7*а*) в исходной модели был изменен первый слой, так чтобы он принимал изображение с одним, а не тремя каналами яркости. При этом размер ядра и шаг свертки остались без изменения. Кроме того, из модели были удалены последние два слоя: адаптивный глобальный усредняющий пулинг и полносвязный слой для классификации 1000 классов.

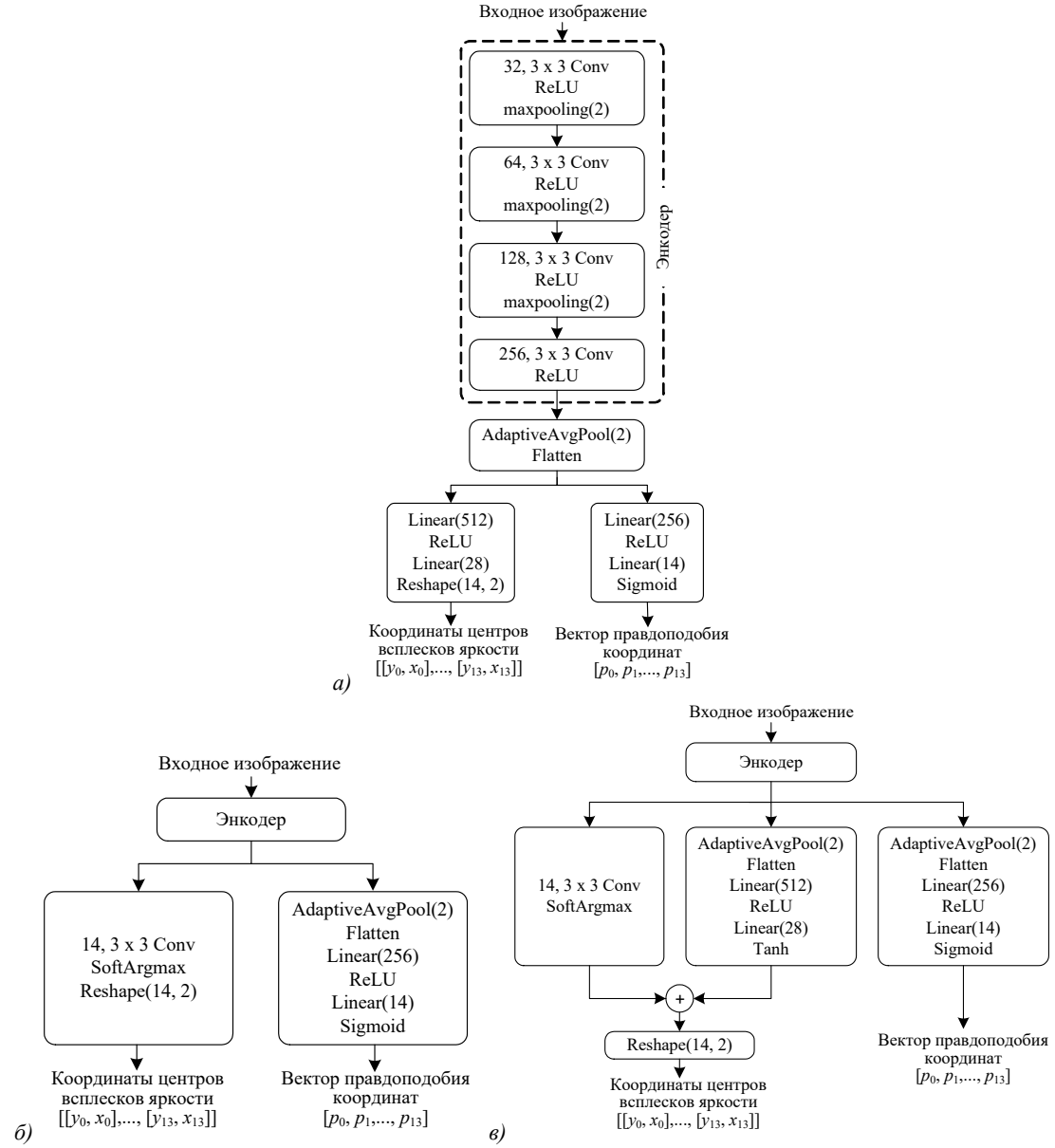


Рис. 6. Исследуемые архитектуры нейронных сетей со случайными весами: (а) с прямой регрессией координат, (б) с операцией soft-argmax для вычисления координат пиков яркости по тепловым картам, (в) с операцией soft-argmax и головой для вычисления вектора оценки смещения координат

В энкодере на базе нейронной сети EfficientNet-B0 (рис. 7*б*) у исходной модели также был изменен первый сверточный слой, чтобы нейронная сеть работала с одноканальными изображениями. При этом начальные веса этого слоя были получены как усредненные по каналам значения стандартных весов оригинального слоя. Далее карта признаков с выхода модифицированной EfficientNet-B0 подвергалась серии обратных сверток (ConvTranspose) для постепенного увеличения размера и уменьшения количества каналов, чтобы в итоге к полученной карте признаков можно было применить операцию soft-argmax для нахождения координат точек. Аналогичные изменения стандартной модели и добавления новой цепочки слоев были сделаны для энкодера на базе нейронной сети ConvNeXt-Tiny (рис. 7*в*).

Эксперименты и обучение моделей

В соответствии с описанной методикой формирования обучающей выборки для обучения различных архитектур сверточных нейронных сетей было сформировано 35000 биспиральных наноскопических

изображений флуоресцирующих нанокристаллов. В данные добавлялся аддитивный белый гауссов шум в отношении сигнал/шум 0,1. Размер всех изображений был равен 100 × 100 пикселей, а для представления яркости каждого пикселя использовалось 2 байта в формате целого числа без знака (uint16). Максимальное количество откликов на всех изображениях не превышало 14 ($N \leq 14$). Предобработка данных включала минимаксную нормализацию пикселей до интервала $[-1, 1]$. Для этого по всей обучающей выборке предварительно были вычислены глобальные минимальное и максимальное значения яркости.

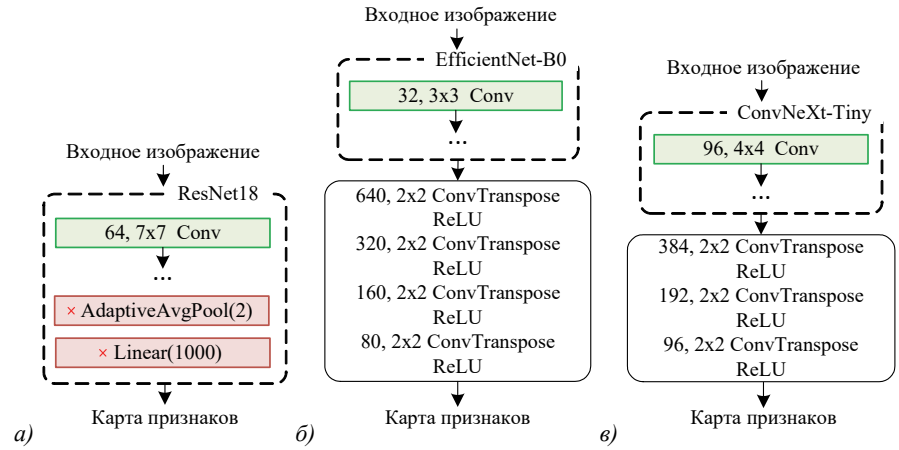


Рис. 7. Структурные схемы энкодеров на базе предобученных моделей нейронных сетей: (а) на базе ResNet18, (б) на базе EfficientNet-B0, (в) на базе ConvNeXt-Tiny

Процесс обучения моделей осуществлялся в течение 150 эпох с использованием оптимизатора Adam и прерывался при наступлении переобучения. Для контроля переобучения исходный датасет был случайным образом разделен в соотношении 70/15/15 на обучающую, проверочную и тестовую выборки. Помимо MSE, используемой для оптимизации параметров моделей, на каждой эпохе также вычислялась средняя абсолютная ошибка, обеспечивающая интуитивно понятную оценку качества работы нейронной сети. Наилучшую точность среди моделей со случайно инициализированными весами продемонстрировала архитектура нейронной сети, в которой использовалась операция soft-argmax для определения субпиксельных координат пиков яркости (рис. 6*б*), график обучения которой представлен на рис. 8*а*. Из графика (рис. 8*а*) видно, что переобучение наступает в районе 23 эпохи, при этом средняя абсолютная ошибка модели составляет 6 пикселей.

Наилучшую точность среди нейронных сетей, где в качестве энкодеров выступали предобученные модели, показала модель на основе энкодера ConvNeXt-Tiny с операцией soft-argmax для субпиксельной регрессии координат (рис. 7*б*). Ее кривая обучения представлена на рис. 8*б*, а пример локализации пиков яркости – на рис. 9.

Значения MAE, полученные на тестовых данных для всех исследуемых моделей до наступления переобучения, представлены в табл. 1.

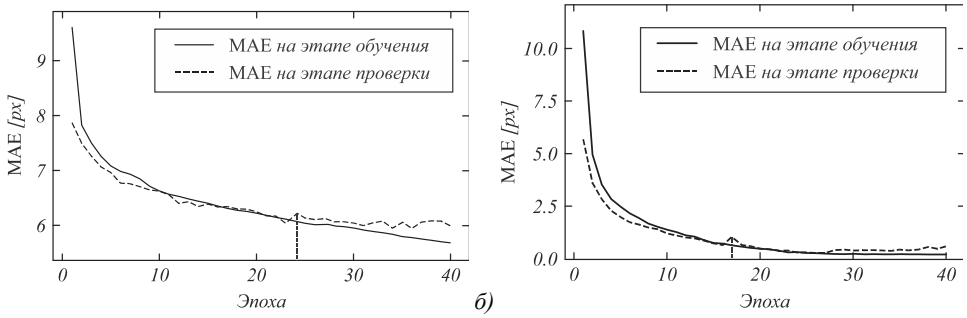


Рис. 8. График обучения: (а) сверточной нейронной сети со случайными весами и операцией soft-argmax для вычисления субпиксельных координат, (б) нейронной сети с энкодером на базе ConvNeXt-Tiny и операцией soft-argmax для вычисления субпиксельных координат

Заключение

В работе выполнен практический сравнительный анализ методов на базе сверточных нейронных сетей для субпиксельной локализации центров пиков яркости на биспиральных наноскопических изображениях флуоресцирующих нанокристаллов. Рассмотрены три схемы предсказания координат: прямая регрессия, вычисление координат по тепловым картам с помощью операции soft-argmax, а также комбинации операции soft-argmax с отдельной головой для вычисления субпиксельного смещения. При этом в качестве энкодеров использовались сверточная нейронная сеть со случайно инициализированными весами, а также предобученные модели ResNet-18, EfficientNet-B0 и ConvNeXt-Tiny. Результаты показали преимущество использования схемы с

комбинацией операции soft-argmax и современных предобученных энкодеров. Лучшая точность была достигнута у модели, в которой использовалась преодобученная нейронная сеть ConvNeXt-Tiny в качестве энкодера, а регрессия координат центров пиков яркости выполнялась с помощью операции soft-argmax. Средняя абсолютная ошибка данной модели составила 0,6024 пикселя до наступления переобучения на семнадцатой эпохе. Для EfficientNet-B0 с soft-argmax ошибка составила 0,9152 пикселя, для ResNet-18 – около 2,66 пикселя, тогда как сверточная нейронная сеть со случайно инициализированными весами в комбинации с операцией soft-argmax обеспечила точность не лучше 6,03 пикселя. Прямая регрессия координат на базе последовательной нейронной сети уступала использованию операции soft-argmax даже при сильных энкодерах. В частности, использование предобученного энкодера ConvNeXt-Tiny в комбинации с прямой регрессией координат обеспечило среднюю абсолютную ошибку 3,081 пикселя. Добавление отдельной головы смещения к soft-argmax в среднем не улучшало точность и иногда ухудшало ее. Так, для схемы с предобученным энкодером ConvNeXt-Tiny, операцией soft-argmax и отдельной головой, вычисляющей субпиксельное смещение, средняя абсолютная ошибка составила 1,3096 пикселя. Дальнейший прогресс ожидается за счет использования более мощных энкодеров, дообучения на реальных и синтетических данных, введения температурных коэффициентов в функцию ошибки и постобработки максимумов, например гауссовой аппроксимации.

Табл. 1. Значения MAE для исследуемых моделей, полученные на тестовых данных до наступления переобучения

Энкодер	Прямая регрессия координат с помощью последовательной нейронной сети	Определение координат с помощью операции soft-argmax	Определение координат с помощью операции soft-argmax и головы, вычисляющей смещение
CNN со случайными весами	9,688 (10 эпох)	6,031 (23 эпохи)	7,831 (14 эпох)
ResNet18	5,171 (12 эпох)	2,6569 (6 эпох)	3,991 (12 эпох)
EfficientNet	4,001 (21 эпох)	0,9152 (29 эпох)	1,569 (9 эпох)
ConvNeXt	3,081 (9 эпох)	0,6024 (17 эпох)	1,3096 (15 эпох)

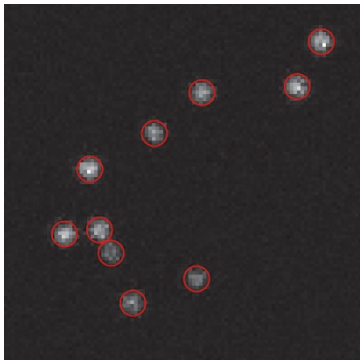


Рис. 9. Пример локализации пиков яркости на биспиральном наноскопическом изображении с помощью нейронной сети с предобученным энкодером ConvNeXt-Tiny и операцией soft-argmax для субпиксельной регрессии координат

Благодарности

Работа выполнена при поддержке АНО «Аналитический центр при Правительстве Российской Федерации», грант № 70-2025-000738.

References

[1] Pavani SRP, Thompson MA, Biteen JS, Lord SJ, Liu N, Twieg RJ, Piestun R, Moerner WE. Three-dimensional single-molecule fluorescence imaging beyond the diffraction limit using a double-helix point spread function. Proceedings of the National Academy of Sciences of the United States of America 2009; 106(9): 2995-2999. DOI: 10.1073/pnas.0900245106.

[2] Pavani SRP, Piestun R. 3D microscopy with a double-helix point spread function. In: Proceedings of SPIE 7184, Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XVI. SPIE; 2009: 71840Z. DOI: 10.1117/12.816585.

[3] Shechtman Y, Sahl SJ, Backer AS, Moerner WE. Optimal point spread function design for 3D imaging. Physical Review Letters 2014; 113(13): 133902. DOI: 10.1103/PhysRevLett.113.133902.

[4] Martens KJA, Jabermoradi A, Yang S, Hohlbein J. Integrating engineered point spread functions into the phasor-based single-molecule localization microscopy framework. Methods 2021; 193: 107-115. DOI: 10.1016/j.ymeth.2020.07.010.

[5] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521(7553): 436-444. DOI: 10.1038/nature14539.

[6] Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016. ISBN: 978-0-262-03561-3.

[7] Weigert M, Schmidt U, Haase R, Sugawara K, Myers G. Star-convex polyhedra for 3D object detection and segmentation in microscopy. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE; 2020: 3655-3662. DOI: 10.1109/WACV45572.2020.9093435.

[8] Hendrycks D, Dietterich T. Benchmarking neural network robustness to common corruptions and perturbations. In: Proceedings of the International Conference on Learning Representations; 2019. arXiv:1903.12261.
[9] Toshev A, Szegedy C. DeepPose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2014: 1653-1660. DOI: 10.1109/CVPR.2014.214.
[10] Luvizon DC, Tabia H, Picard D. Human pose regression by combining indirect part detection and contextual information. Computers & Graphics 2019; 85: 15-22. DOI: 10.1016/j.cag.2019.09.002.
[11] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. IEEE; 2017: 2961-2969. DOI: 10.1109/ICCV.2017.322.
[12] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2016: 770-778. DOI: 10.1109/CVPR.2016.90.
[13] Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research. Vol 97. PMLR; 2019: 6105-6114.
[14] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE; 2022: 11976-11986. DOI: 10.1109/CVPR52688.2022.01167.
[15] Bayas C, von Diezmann A, Gustavsson AK, Moerner WE. Easy-DHPSF 2.0: Open-source software for three-dimensional localization and two-color registration of single molecules with nanoscale accuracy. Protocol Exchange; 2019. DOI: 10.21203/rs.2.9151/v2.
[16] Lew MD, von Diezmann ARS, Moerner WE. Easy-DHPSF: Open-source software for three-dimensional localization of single molecules with precision beyond the optical diffraction limit. Protocol Exchange; 2013. DOI: 10.1038/protex.2013.026.
[17] Zhang Z, Ochirova A, Liu S, Herbert AD, Wu Y, Boucher W, Lee SF, Laue ED, Klenerman D, Ponjavic A. DHPSFU: A Fiji plugin for fast and accurate double helix-PSF 3D single-molecule localisation microscopy. Scientific Reports 2025; 15: 30551. DOI: 10.1038/s41598-025-15623-3.
[18] Gao J, Wang P, Li W, Zhang X, Song C, Liu Z, Han S, Liu H. Imaging and positioning through scattering media with double-helix point spread function engineering. Journal of Biomedical Optics 2023; 28(4): 046008. DOI: 10.1117/1.JBO.28.4.046008.
[19] Carreira J, Agrawal P, Fragkiadaki K, Malik J. Human pose estimation with iterative error feedback. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2016: 4733-4742. DOI: 10.1109/CVPR.2016.512.
[20] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2016: 779-788. DOI: 10.1109/CVPR.2016.91.
[21] Sun X, Xiao B, Wei F, Liang S, Wei Y. Integral human pose regression. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, eds. Computer Vision – ECCV 2018. Lecture Notes in Computer Science. Vol 11210. Springer; 2018: 529-545. DOI: 10.1007/978-3-030-01231-1_33.
[22] Nibali A, He Z, Morgan S, Prendergast L. Numerical coordinate regression with convolutional neural networks. arXiv preprint arXiv:1801.07372; 2018. DOI: 10.48550/arXiv.1801.07372.
[23] Zhou X, Wang D, Krähenbühl P. Objects as points. arXiv preprint arXiv:1904.07850; 2019. DOI: 10.48550/arXiv.1904.07850.
[24] Papandreou G, Zhu T, Chen LC, Gidaris S, Tompson J, Murphy K. PersonLab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In: Computer Vision – ECCV 2018. Lecture Notes in Computer Science. Vol 11218. Springer; 2018: 282-299. DOI: 10.1007/978-3-030-01264-9_17.
[25] Cheezum MK, Walker WF, Guilford WH. Quantitative comparison of algorithms for tracking single fluorescent particles. Biophysical Journal 2001; 81(4): 2378-2388. DOI: 10.1016/S0006-3495(01)75884-5.
[26] Pallotta L, Giunta G, Clemente C. Sub-pixel SAR image registration through parabolic interpolation of the 2D cross-correlation. IEEE Transactions on Geoscience and Remote Sensing 2020; 58(6): 4132-4144. DOI: 10.1109/TGRS.2019.2961245.
[27] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2009: 248-255. DOI: 10.1109/CVPR.2009.5206848.

Сведения об авторах

Наумов Андрей Витальевич, 1974 года рождения, д.т.н., член-корреспондент РАН, Институт спектроскопии РАН, Москва. E-mail: *a.naumov@troitsk.lebedev.ru* Orcid: 0000-0001-7938-9802

Роженцов Алексей Аркадьевич, 1972 года рождения, д.т.н., исполняющий обязанности ректора Поволжского государственного технологического университета, Йошкар-Ола. E-mail: *rozhencovaa@volgatech.net* Orcid: 0000-0003-0043-0070

Баев Алексей Александрович, 1984 года рождения, к.т.н., доцент кафедры проектирования и производства электронно-вычислительных средств Поволжского государственного технологического университета, Йошкар-Ола. E-mail: *baevaa@volgatech.net* Orcid: 0000-0003-1285-8752

Иванов Константин Олегович, 1989 года рождения, к.т.н., доцент кафедры радиотехнических систем Поволжского государственного технологического университета, Йошкар-Ола. E-mail: *ivanovko@volgatech.net* Orcid: 0000-0002-6395-4085

Поступила в редакцию 17 декабря 2025 г. *Окончательный вариант – 09 февраля 2026 г.*

Application of Deep Learning Technologies for Brightness Peak Detection in Bispiral Images of Fluorescent Point Emitters

A.V. Naumov [1], A.A. Rozhentsov [2], A.A. Baev [2], K.O. Ivanov [2]
[1] Institute of Spectroscopy of the Russian Academy of Sciences (ISAN), Fizicheskaya St. 5, Troitsk, Moscow, 108840, Russia;
[2] Volga State University of Technology, Lenin Square 3, Yoshkar-Ola, 424000, Russia

Abstract

The article examines approaches to forming training datasets, features of input data representation, and neural network architectures for solving the problem of localizing diffraction peaks in bi-spiral images of fluorescent point emitters (molecules, quantum dots, and others) in three-dimensional localization microscopy with ultra-high spatial resolution using rotating point spread function optics. The advantage of calculating sub-pixel peak coordinates from heatmaps using the soft-argmax operation and the use of pre-trained encoders with limited data volume is demonstrated. In particular, the mean absolute error (MAE) of localization for the model with a pre-trained ConvNeXt-Tiny encoder and soft-argmax was 0.60 pixels, approximately 5 times lower than with direct coordinate regression using the same encoder (3.08 pixels) and 10 times lower than that of a convolutional neural network with randomly initialized weights (6.03 pixels). For the EfficientNet-B0 and ResNet-18 encoders with soft-argmax, the MAE was 0.92 and 2.66 pixels, respectively.

Keywords: bispiral images, deep learning, artificial intelligence, computer vision, machine learning, nanoscopy, convolutional neural networks, digital microscopy.

Citation: Naumov AV, Rozhentsov AA, Baev AA, Ivanov KO. Application of Deep Learning Technologies for Brightness Peak Detection in Bispiral Images of Fluorescent Point Emitters. Computer Optics 2026; 50(5): 1919. doi: 10.18287/COJ1919.

Acknowledgements: This work was financially supported by the Analytical Center for the Government of the Russian Federation under grant No. 70-2025-000738.

About authors

Andrey Vitalyevich Naumov, born 1974, Doctor of Technical Sciences, Corresponding Member of the Russian Academy of Sciences; Institute of Spectroscopy of the Russian Academy of Sciences (ISAN), Moscow. E-mail: *a.naumov@troitsk.lebedev.ru* ORCID: ORCID: 0000-0001-7938-9802

Aleksey Arkadyevich Rozhentsov, born 1972, Doctor of Technical Sciences, Acting Rector of Volga State University of Technology, Yoshkar-Ola. E-mail: *rozhencovaa@volgatech.net* ORCID: 0000-0003-0043-0070

Aleksey Aleksandrovich Baev, born 1984, PhD in Engineering, Associate Professor, Department of Design and Production of Electronic Computing Equipment, Volga State University of Technology, Yoshkar-Ola. E-mail: *baevaa@volgatech.net* ORCID: 0000-0003-1285-8752

Konstantin Olegovich Ivanov, born 1989, PhD in Engineering, Associate Professor, Department of Radio Engineering Systems, Volga State University of Technology, Yoshkar-Ola. E-mail: *ivanovko@volgatech.net* ORCID: 0000-0002-6395-4085

Received December 17, 2025. The final version – February 09, 2026.