

# **Сравнение нейросетевых и непрерывно-морфологических методов в задаче детекции текста (Text Detection)**

Гайдученко Н.Е., Труш Н.А., Торлак А.В., **Миронова Л.Р.**, **Акимов К.М.**, Гончар Д.А.

October 20, 2018

## **Аннотация**

Данная статья посвящена анализу и сравнению методов распознавания текста на изображениях. Текст - это один из наиболее распространенных способов коммуникации, который для передачи информации может быть представлен в виде документа или являться частью картинки. Несмотря на достигнутые результаты в этой области, задача детекции текста на изображениях требует дальнейшего исследования, особенно для изображений с сложным фоном. В данной статье мы будем рассматривать использование как моделей глубокого обучения, которые активно применяются в машинном зрении, так и непрерывно-морфологических методов обработки изображений. Сравнение будет производиться на датасетах, представляющих разные стороны задачи детекции текстов, с использованием ряда метрик (таких как F-score, etc.) для определения качества работы сравниваемых методов.

**Ключевые слова:** *нейронные сети, непрерывно-морфологические методы, распознавание текста, обучение без учителя.*

## Введение

В последнее время задача детекции текста на изображениях документов и на реальных фотографиях, содержащих надписи, привлекла немалое внимание в научном сообществе. Существуют методы и технологии распознавания строк и букв в сканированных документах, однако они мало применимы к реальным изображениям. Актуальность задачи детекции текста связана со следующими факторами. Во-первых, увеличилось число возможных приложений данной технологии. Например, она может применяться на мобильных устройствах, в роботизированных системах, при поиске картинок в интернете. Во-вторых, текст - это один из наиболее распространенных способов коммуникации, который для передачи информации может быть представлен в виде документа или являться частью картинки. Автоматическое нахождение текста на изображении позволяет увеличить объем получаемой информации. В-третьих, успехи в области компьютерного зрения позволяют применять более продвинутые методы для решения задачи детекции текста на изображениях. В то время как задача оптического распознавания букв на изображениях сканированных документов может считаться изученной, задача детекции текста показывает невысокое качество распознавания, порядка 80%. Данная задача требует учета большого количества факторов, таких как ориентация текста, шрифт, цвет, освещение, фон изображения.

В данной статье проводится сравнительный обзор методов детекции текста на изображениях, рассматриваются как непрерывно-морфологические методы, так и нейросетевые, и выявляется, какие из них в каких случаях показывают наиболее точный результат.

Первый из рассмотренных нами морфологических методов [1] использует геометрический скелет для изображения текста, построенный по бинаризованной картинке, с последующим построением метрики близости и кластеризацией полученного графа и позволяет работать с рукописными документами, так как не опирается на знание о почерке, стиле, языке или о структуре текста.

Во втором [3] используется гауссовский фильтр второго порядка для получения локального представления об ориентации и размере текста. Далее полученные значения собираются в гистограммы для определения общих значений. Полученные распределения бинаризуются для выделения линий текста из фона. Гауссовский фильтр можно эффективно посчитать и получающееся качество близко к более сложным алгоритмам обучения с учителем.

В третьем подходе [4] применяется модификация метода наиболее стабильных экстремальных регионов, посчитанная по 4 цветовым каналам, для выделения букв на картинке. Далее идет анализ связных компонент для выделения регионов с текстом и последующее формирование строк текста. Данный подход может бороться с яркостными эффектами.

Среди нейронных методов мы рассмотрели сверточные сети [5], метод с боксами [6]. А также STPN [7], который хорошо работает на крупномасштабных и на разноязыковых текстах без дальнейшей обработки, в отличие от многих других методов. На вход он принимает изображение любого размера и обнаруживает text line путем плотного продвижения маленького окна. Такой подход позволяет детектировать символы различного размера за счет того, что размер окна меняется.