

# Постановка задачи сравнения нейросетевых и непрерывно-морфологических методов детекции текста

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

December 6, 2018

**Модель СТРПН делится на три основных этапа:**

## 1. Detecting Text in Fine-scale Proposals

СТРПН представляет собой сверточную сеть, которая принимает на вход изображения любого размера. Текст обнаруживается путем плотного скольжения окошка маленького размера и выводит прямоугольники, которые обводят найденный моделью текст. Детектор скользит через карты conv5, где каждый шаг измеряется 16 пикселями. Предсказанные координаты образует секции, содержащие текстовые предложения. Вертикальные координаты каждой секции определяются следующим образом:

$$\begin{aligned} v_c &= (c_y - c_y^a)/h_a, & v_h &= \log(h/h^a) \\ v_c^* &= (c_y^* - c_y^a)/h_a, & v_h &= \log(h^*/h^a) \end{aligned}$$

$v$  - предсказанные координаты

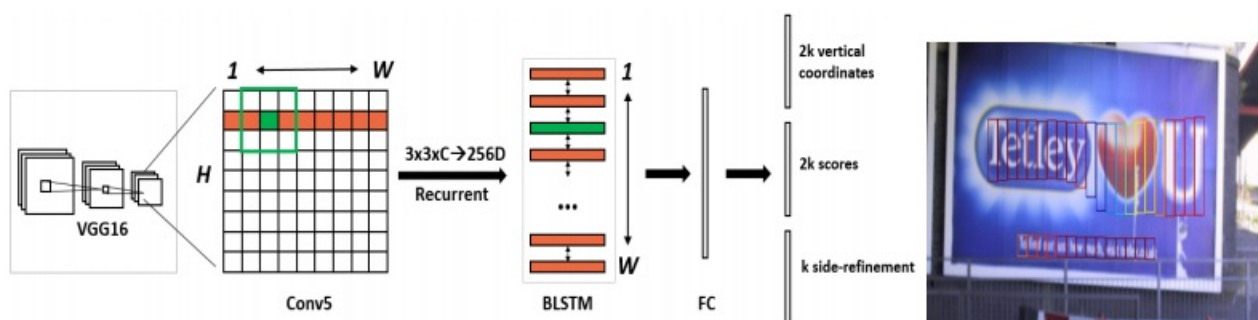
$v^*$  - ground-thruth координаты

$c_y^a$  - центр по оси  $y$

$h_a$  - ширина секции

$c_y, h$  - предсказанные координаты

Таким образом, получаем каждое текстовое предложение в определенной рамке размером  $h \times 16$  как на данном изображении:



## 2. Recurrent Connectionist Text Proposals

Чтобы повысить точность локализации текста, детектируемая строка делится на более мелкие текстовые предложения, каждые из которых обрабатываются отдельно. Для этого, RNN сеть, использующая conv5-карты, принимает на вход сверточную функцию каждого окна как последовательность входов и обновляет его внутреннее состояние в скрытом слое  $H_t$ .

$$H_t = \phi(H_{t-1}, X_t) \quad t = 1, 2, \dots, W$$

$H_t$  - рекуррентное внутреннее состояние, которое вычисляется через текущее состояние  $X_t$  и предыдущее  $H_{t-1}$ .

$X_t$  - входящий conv5 признак из  $t$  скользящего окна(3x3).

$W$ - ширина окна conv5

### 3.Side-refinement

На данном этапе анализируются все детектируемые текстовые секции и принимается каждая текстовая секция у которой соотношение текст/не текст  $> 0.7$ .

$$o = (x_{side} - c_x^a)/w^a, \quad o^* = (x_{side}^* - c_x^a)/w^a$$

$x_{side}$  - предсказанная  $x$  координата до ближайшей горизонтальной стороны(левой или правой)  $c_x^a$  - центр координаты  $w_a$  - ширина, которая зафиксирована.