

УДК 004.89:004.93

Д.Р. Чарковський, Н.Б. Стадник, канд. техн. наук

Тернопільський національний технічний університет імені Івана Пулюя, Україна

МЕТОДИ ДЕТЕКТУВАННЯ ТЕКСТОВИХ ОБЛАСТЕЙ НА ЗОБРАЖЕННЯХ

D.R. Charkovkyi, N.B. Stadnyk, Ph.D.

METHODS FOR DETECTION OF TEXT REGIONS IN IMAGES

Кількість інформації, котра зберігається як зображення, велика, що робить неможливим її опрацювання вручну. Попри те, що область можливого застосування достатньо широка та наявний значний прогрес у машинному навчанні, детектування та поділ на сегменти текстових областей на зображеннях наразі ще є проблемою. Текст, котрий розміщений на зображеннях реальних сцен, може мати різний розмір, колір, стиль, поворот щодо горизонту тощо. Також ймовірним є його перекриття додатковими об'єктами на зображенні чи присутність низької контрастності із основним фоном.

Для детектування та виділення розташування текстових областей на зображеннях застосовуються стандартні методи розпізнавання текстових областей. Усі наявні методи розпізнавання можна поділити на такі три групи:

- методи пов'язаних компонентів;
- текстурні методи;
- методи глибокого навчання.

Для порівняння методів детектування текстових областей на зображеннях було розглянуто такі характеристики:

- швидкість роботи (кількість часу, необхідна для навчання цього методу);
- ресурсна вартість обчислень (вимоги до обчислювальних ресурсів обладнання);
- робота зі складними зображеннями: геометричні перетворення (повороти, розтягування); текстурні спотворення (шум, низька контрастність, складні фони);
- ефективність методу (точність чи інші засоби оцінки).

Методи, які порівнювалися: пов'язаних компонентів - метод максимально стабільних екстремальних областей (MSER), метод перетворення за товщиною штриха символу тексту (SWT); текстурні - дискретно-косинусне перетворення (DCT), гістограма спрямованих градієнтів (HOG), ознаки Хаара; глибокого навчання - згортова нейронна мережа (CNN), повнозгортова нейронна мережа (FCN), мережа архітектури U-Net.

Результати порівняння свідчать, що методи пов'язаних компонентів володіють високою швидкістю роботи, але недостатньо добре працюють із різноманітними спотвореннями зображень та володіють середньою ефективністю. Текстурні методи мають середню швидкість роботи, погано працюють з різними спотвореннями і мають низьку ефективність. Методи глибокого навчання є найповільнішою групою через тривалий час навчання, але добре працюють з різними спотвореннями зображень, а також мають високу продуктивність. Отже, методи, що ґрунтуються на глибокому навчанні, становлять велику цінність для розгляду.

CNN пред'являють менше вимог до обладнання порівняно з іншими мережами, проте є недостатніми для визначення розташування текстових областей на зображеннях. FCN з несиметричною структурою та мережі архітектури U-Net знаходяться на одному рівні. Однак симетрична архітектура мережі U-Net надає більше можливостей для розпізнавання текстових областей через наявність більшої кількості шарів підвищуючої дискретизації для опрацювання карт ознак. Тому архітектура U-Net була обрана для реалізації алгоритму розпізнавання написів на зображеннях.