⚡️🚀המאמר היומי של מייק 30.08.24: ⚡️🚀

Platypus: A Generalized Specialist Model for Reading Text in Various Forms

חוזרים לסקירות אחרי שבוע של חופשה עם מאמר בנושא שלא סקרתי די הרבה זמן והוא Optical Character Recognition או OCR בקצרה. מטרת OCR היא לזהות טקסט בתמונה או במסמך כאשר הטקסט יכול להופיע בצורות ומגוונות. מודלי OCR הקודמים בדרך כלל התמקדו בזיהוי של סוג של טקסט (נגיד נוסחה, טקסט מודפס או כתב יד). המחברים מציעים גישה שמאחדת את מומחי ה-OCR ה״צרים״ לזיהוי סוג ספציפי של טקסט - כלומר מסוגלת לזהות כל סוג של טקסט בתמונה כולל המקרים שיש כמה סוגים של טקסט בתמונה.

בנוסף ב-OCR יש 3 משטרי הפעלה. הראשון זה RAT או Recognize All Text שמטרתו לזהות את כל הטקסטים בתמונה. השני הוא PPR או Point Prompt Recognition שמטרתו לזהות את הטקסט סביב נקודה נתונה (סוג של עוגן) בתמונה. האחרון הוא Box Prompt Recognition או BPR שמיועד לזיהוי של טקסט בתוך מלבן נתון בתמונה (כמו שיש לנו Bounding Boxes בזיהוי אובייקטים בתמונה אבל בכיוון ההפוך).

אז המחברים מאמנים מודל המורכב מהאנקודר (שהופך תמונה לאמבדינג) הדקודר האוטורגרסיבי. הדקודר מקבל כקלט את סוג הטקסט בתמונה (מודפס או כתב יד). בנוסף הדקודר מקבל את סוג המשימה (RAT, PPR או BPR) עם כל הפרטים הנחוצים לביצוע משימה (כלומר קואורדינטות של הפאץ'). בנוסף המודל מקבל גרנולריות של זיהוי הטקסט (כלומר word-level או line-level שהראשון הוא זיהוי מילה בודדת והשני הוא זיהוי טקסט שלם). הפרטים האלו מוזנים כאמור לדקודר שמטרתו לגנרט את הטקסט המופיע בתמונה.

זה כל הפרטים המעניינים - מאמר די קליל….

https://arxiv.org/abs/2408.14805