

مطالعه در مورد مدل های فارسی انگلیسی فعلی برای ocr در سطح word:

معرفی مدل TrOCR

مدل TrOCR یک مدل encoder-decoder است که شامل یک ترنسفورمر تصویری به عنوان encoder و یک ترنسفورمر متنی به عنوان decoder می باشد. این مدل توسط Li و همکاران در مقاله ای با عنوان "TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models" معرفی شده و اولین بار در این مخزن [microsoft/trocr-base-handwritten](https://huggingface.co/microsoft/trocr-base-handwritten) منتشر شده است.

توضیحات مدل TrOCR

- **Encoder**** تصویر: **Encoder**** تصویر از وزن های پیش آموزش داده شده مدل BEiT بهره می برد.

- **Decoder**** متن: **Decoder**** متن نیز از وزن های مدل RoBERTa استفاده می کند.

نحوه عملکرد

تصاویر به مدل به صورت توالی ای از قطعات با اندازه ثابت (رزولوشن x1616) ارائه می شوند که به طور خطی جاسازی شده اند و همچنین این قطعات با استفاده از encoder تصویری BEiT پردازش می شوند. سپس، ویژگی های استخراج شده از رمزگذار به decoder متنی RoBERTa ارسال می شوند تا توالی های متنی تولید شوند.

نتایج و کاربردها

مدل TrOCR توانسته است در تشخیص متن از تصاویر (OCR) به نتایج برجسته ای دست یابد و در کاربردهای مختلفی از جمله تشخیص متون دست نویس و چاپی مورد استفاده قرار گیرد. برخی از ویژگی ها و کاربردهای این مدل شامل موارد زیر است:

- ****سرعت بالا****: مدل TrOCR بهینه سازی شده و می تواند با سرعت بالایی عملیات OCR را انجام دهد.

- ****دقت بالا****: استفاده از وزن های پیش آموزش داده شده مدل های BEiT و RoBERTa باعث افزایش دقت در تشخیص متون شده است.

- **انعطاف‌پذیری:** این مدل قابلیت استفاده در کاربردهای مختلف از جمله تشخیص متون دست‌نویس، چاپی و حتی متون پیچیده‌تر را داراست.

منابع و مستندات

- مقاله مرجع مدل: [TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models](https://arxiv.org/abs/2109.10282)

- مخزن مدل: [microsoft/trocr-base-handwritten](https://huggingface.co/microsoft/trocr-base-handwritten)

معرفی دیتاست IAM

در باره مجموعه داده

مجموعه داده‌های فرم‌های دست‌نویس IAM شامل فرم‌هایی از متن انگلیسی دست‌نویس است که می‌تواند برای آموزش و آزمایش تشخیص‌دهنده‌های متن دست‌نویس و انجام آزمایش‌های شناسایی و تأیید نویسنده استفاده شود.

محتوا

مجموعه داده شامل فرم‌های کاملی از متن دست‌نویس بدون محدودیت است که با وضوح dpi300 اسکن شده و به‌عنوان تصاویر PNG با 256 سطح خاکستری ذخیره شده‌اند. فرم‌ها به دایرکتوری‌های جداگانه تقسیم می‌شوند به طوری که همه فرم‌ها در هر دایرکتوری توسط یک شخص نوشته می‌شوند.

نتیجه‌گیری

مدل TrOCR یک ابزار قدرتمند و دقیق برای تشخیص متن از تصاویر است که با استفاده از تکنیک‌های پیشرفته ترنسفورمر می‌تواند به دقت و سرعت بالایی دست یابد. این مدل در کاربردهای مختلفی مانند تشخیص متون دست‌نویس و چاپی بسیار موثر است و می‌تواند به عنوان یک راحل جامع OCR مورد استفاده قرار گیرد.

چکیده مقاله

تشخیص متن یک مسئله پژوهشی طولانی‌مدت برای دیجیتالی‌سازی اسناد است. روش‌های موجود معمولاً بر اساس CNN برای درک تصویر و RNN برای تولید کاراکترها ساخته شده‌اند. علاوه بر این، معمولاً یک مدل زبان دیگر به عنوان یک مرحله پس‌پردازش برای بهبود دقت کلی نیاز است. در این مقاله، ما یک روش تشخیص متن انتها به انتها با استفاده از مدل‌های ترنسفورمر پیش‌آموزش‌داده‌شده تصویری و متنی، به نام TroOCR، پیشنهاد می‌دهیم که از معماری ترنسفورمر برای هر دو درک تصویر و تولید متن کلمه‌ای بهره می‌برد. مدل TroOCR ساده اما موثر است و می‌تواند با داده‌های مصنوعی بزرگ‌مقیاس پیش‌آموزش داده شده و با مجموعه داده‌های برجسب‌گذاری‌شده انسانی تنظیم دقیق شود. آزمایش‌ها نشان می‌دهند که مدل TroOCR از مدل‌های فعلی در تشخیص متون چاپی، دست‌نویس و صحنه پیشی می‌گیرد.