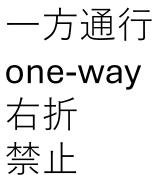
DeepSeek-OCR の Free OCR による文字読み取り





- 最新技術の特徴 ・多言語(日・英など)混在
- ・斜め文字
- ・多様なフォント、手書き

なぜ多言語・多様な文字に強いのか

「Free OCR」= レイアウトを無視した純粋なテキスト抽出

• プロンプト「<image>¥nFree OCR.」を使用すると、DeepSeek-OCR は文書の構造(段落、表、見出し)を無視して、画像内の全テキストをそのまま抽出

二段階処理による多様な文字の認識

- 第一段階:SAM-base (80M)
 - ・一文字ずつの形状、線の太さ、筆圧を認識
 - **斜めの文字、かすれた印刷、手書き**の個別パターンを捉える
- 第二段階:CLIP-large(300M)
 - 画像全体の文脈から文字の意味を理解
 - **多言語が混在**していても、**それぞれの言語の特**徴を識別
- 100言語、3000万枚のPDFによる学習 多様なフォント、言語、書式を幅広く学習済み

3つのコア技術

1 SAM (Segment Anything Model)

- Meta AIが開発した画像認識モデル
- Window Attentionにより文字の細部(形、太さ、書き方)を抽出
- Free OCRでの役割: 画像内の各文字の資格的特徴を認識
- 2 CLIP (Contrastive Language-Image Pre-training)
- OpenAIが開発した視覚と言語を統合するモデル
- 画像とテキストを同じ埋め込み空間(意味的に類似したものが近くに配置される多次元ベクトル空間)で処理
- ・ Free OCRでの役割: 文字全体の文脈を把握し、認識精度を向上
- ③ MoEデコーダ (Mixture-of-Experts)
- DeepSeek3B-MoE:6モデル内に64個の小さなネットワーク(専門家)を持ち、入力データごとにその中から最適な6個だけを動かす仕組み
- 全体では3Bパラメータを持つが、実際に計算するのは570Mパラメータ分だけなので高速
- ・ Free OCRでの役割: 圧縮された視覚情報からテキストを復元

「Free OCR」プロンプトで可能になる新しい応用

- 「Free OCR」プロンプトにより、レイアウト 解析を省略したテキスト抽出を実行
- 名刺、レシート、手書きメモなどの単純な文書 処理に適する
- 構造解析が不要な場面に適する