https://huggingface.co/vidore/colqwen2-v1.0

ColPali ⚡ Safetensors ⊕ English vidore-experimental vidore 📄 arxiv:2004.12832 📄 arxiv:2407.01449 📄 arxiv:2106.09685 🏛 License: mit

● Model card ·≣ Files and versions 🟡 Community **2**

✎ Edit model card

≡

ColQwen2: Visual Retriever based on Qwen2-VL-2B-Instruct with ColBERT strategy

This is the base version trained with batch_size 256 instead of 32 for 5 epoch and with the updated pad token

ColQwen is a model based on a novel model architecture and training strategy based on Vision Language Models (VLMs) to efficiently index documents from their visual features. It is a Qwen2-VL-2B extension that generates ColBERT- style multi-vector representations of text and images. It was introduced in the paper ColPali: Efficient Document Retrieval with Vision Language Models and first released in this repository

This version is the untrained base version to guarantee deterministic projection layer initialization.

ColPali's Architecture

Version specificity

This model takes dynamic image resolutions in input and does not resize them, changing their aspect ratio as in ColPali. Maximal resolution is set so that 768 image patches are created at most. Experiments show clear improvements with larger amounts of image patches, at the cost of memory requirements.

Downloads last month
31,586

✦ **Inference API** ⓘ

Unable to determine this model's pipeline type. Check

⬡ **Model tree for** vidore/colqwen2-v1.0 ⓘ

Base model
⚙ Finetuned
⚙ **Finetuned** (8)

▦ **Spaces using** vidore/colqwen2-v1.0 2

⤳ manu/ColPali-demo ● zb9/colqwen-embed

▤ **Collection including** vidore/colqwen2-v1.0

ColQwen2 Models ▤ Collection
Pre-trained checkpoints for the ColQwen... · 4 item