Image Search with 🤗 datasets

Can we build an image search app for British Library book images (and should we?)

What are in these pictures?

What does this picture show?

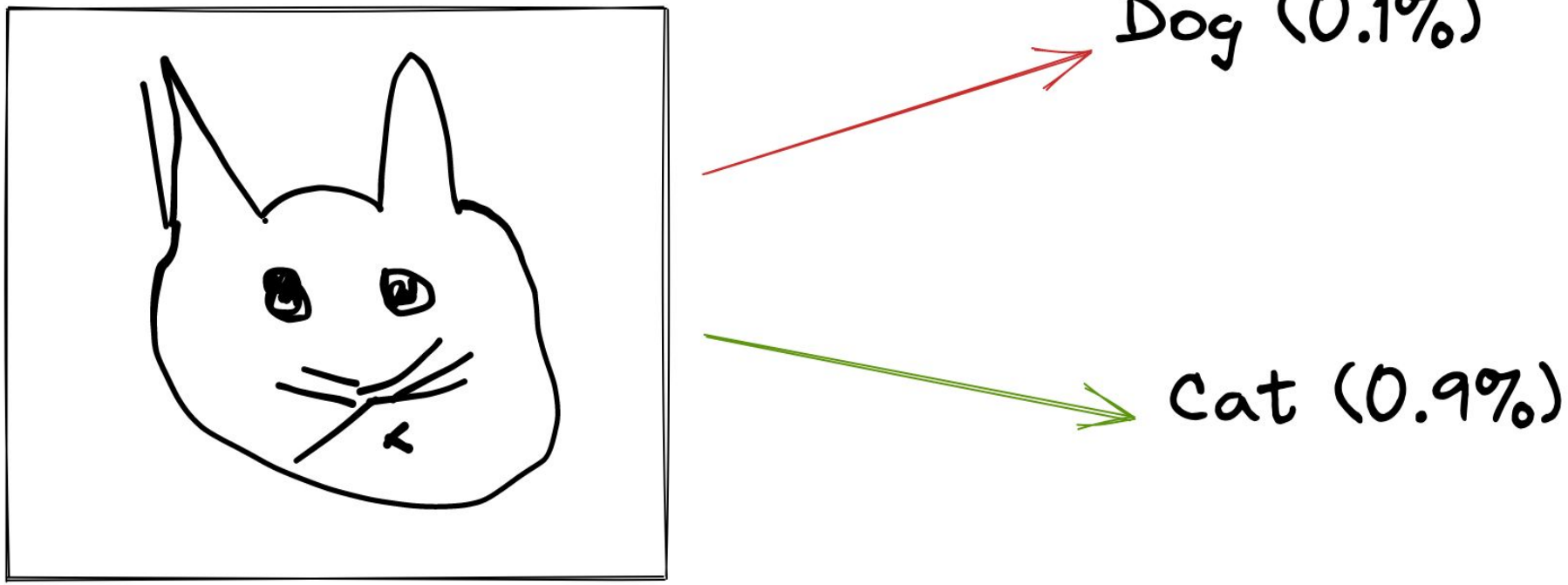
https://www.theguardian.com/uk-news/2016/jan/03/like-a-beautiful-painting-image-of-new-years-mayhem-in-manchester-goes-viral

What are in these pictures?

Dog

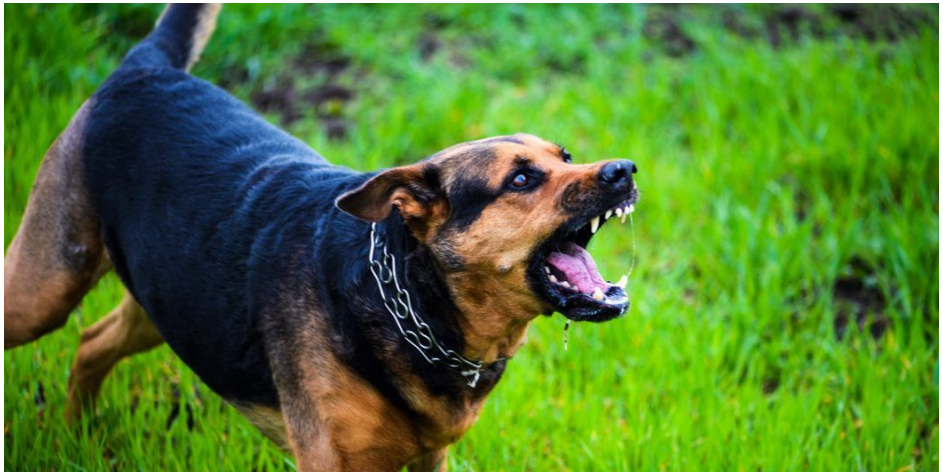
Dog

Labels suck



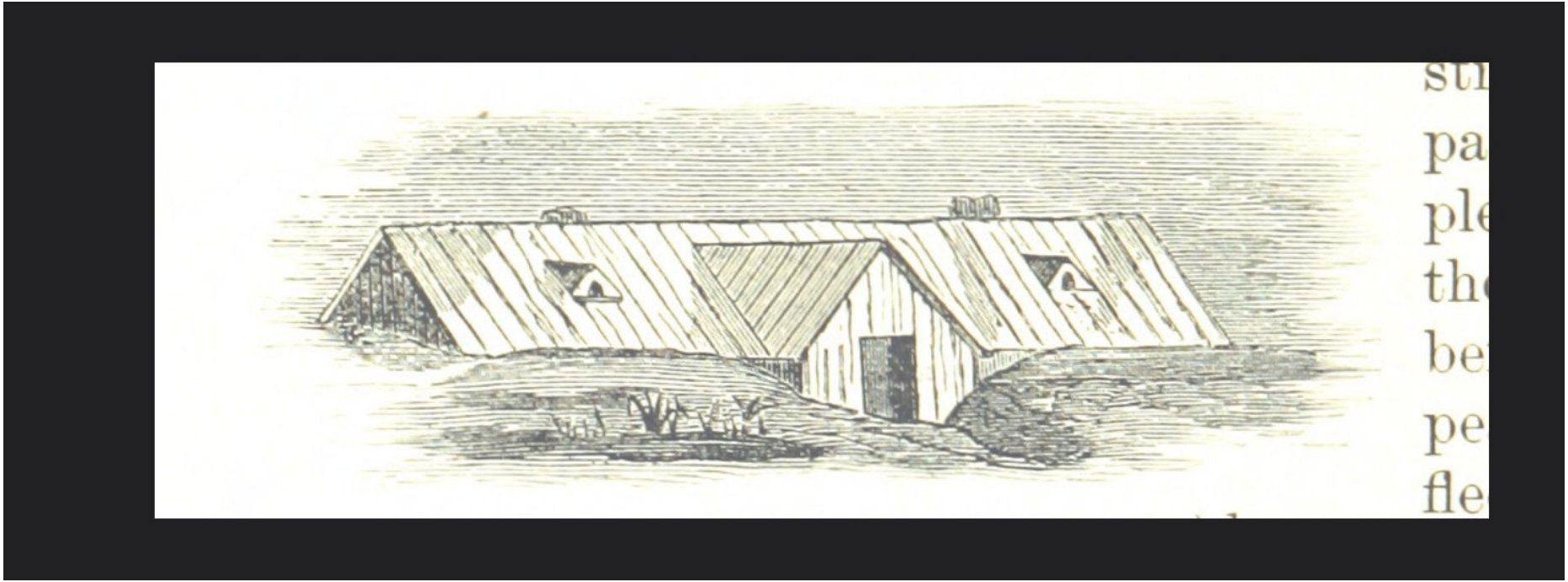
Rich text descriptions

A photograph of a Corgi that looks super happy/slightly baked



A photograph of an angry/scared dog with bared teeth

Describing a rich image collection



British Library + Follow

British Library digitised image from page 44 of "Illustrated Poems and Songs for Young People. Edited by Mrs. Sale Barker"

Image taken from:

Title: "Illustrated Poems and Songs for Young People. Edited by Mrs. Sale Barker"

2,817 views | 6 faves | 0 comments

Taken on December 2, 2013

No known copyright restrictions

Show EXIF

This photo is in 1 gallery

https://www.flickr.com/photos/britishlibrary/11179272775/

← Back to blog

Image search with 🤗 datasets

Published March 16, 2022.

Update on GitHub

davanstrien
Daniel van Strien guest

Open in Colab

🤗 datasets is a library that makes it easy to access and share datasets. It also makes it easy to process data efficiently -- including working with data which doesn't fit into memory.

When datasets was first launched, it was associated mostly with text data. However, recently, datasets has added increased support for audio as well as images. In particular, there is now a datasets feature type for images. A previous blog post showed how datasets can be used with 🤗 transformers to train an image classification model. In this blog post, we'll see how we can combine datasets and a few other libraries to create an image search application.

https://huggingface.co/blog/image-search-datasets

What is 🤗 datasets?

- Hugging Face is a company focused on democratizing machine learning
- Hugging Face datasets is a library for efficiently working with large data (with a particular focus on machine learning)
- Hugging Face has a datasets hub for sharing datasets
- This means we can download datasets with one line of code and get it back in a consistent format i.e.

  dataset = load_dataset('blbooksgenre')

- datasets was originally focused more on text but has branched out into other domains (audio, image etc.)

What is sentence-transformers?

“SentenceTransformers is a Python framework for state-of-the-art sentence, text and image embeddings. The initial work is described in our paper Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.”

Source: https://www.sbert.net/index.html

We can also understand what it does with an example. For the following sentences:

- I like pasta
- I like spaghetti
- I like bread
- I like dungeon synth

We can think of the first three being ‘closer’ semantically since they are all about food, and the last one is about music. Sentence transformers gives us various tools of computing ‘embeddings’. These ‘embeddings’ can represent a sentence in a way which will hopefully mean that the food sentences will be ‘closer’ to each other. In practice this means we could then cluster (or search) for food related sentences and get back the first three before the last one without relying on specific words.

Connecting Text and Images (CLIP)

Given a batch of N (image, text) pairs, CLIP is trained to predict which of the N × N possible (image, text) pairings across a batch actually occurred. To do this, CLIP learns a multi-modal embedding space by jointly training an image encoder and text encoder to maximize the cosine similarity of the image and text embeddings of the N real pairs in the batch while minimizing the cosine similarity of the embeddings of the N 2 − N incorrect pairings.

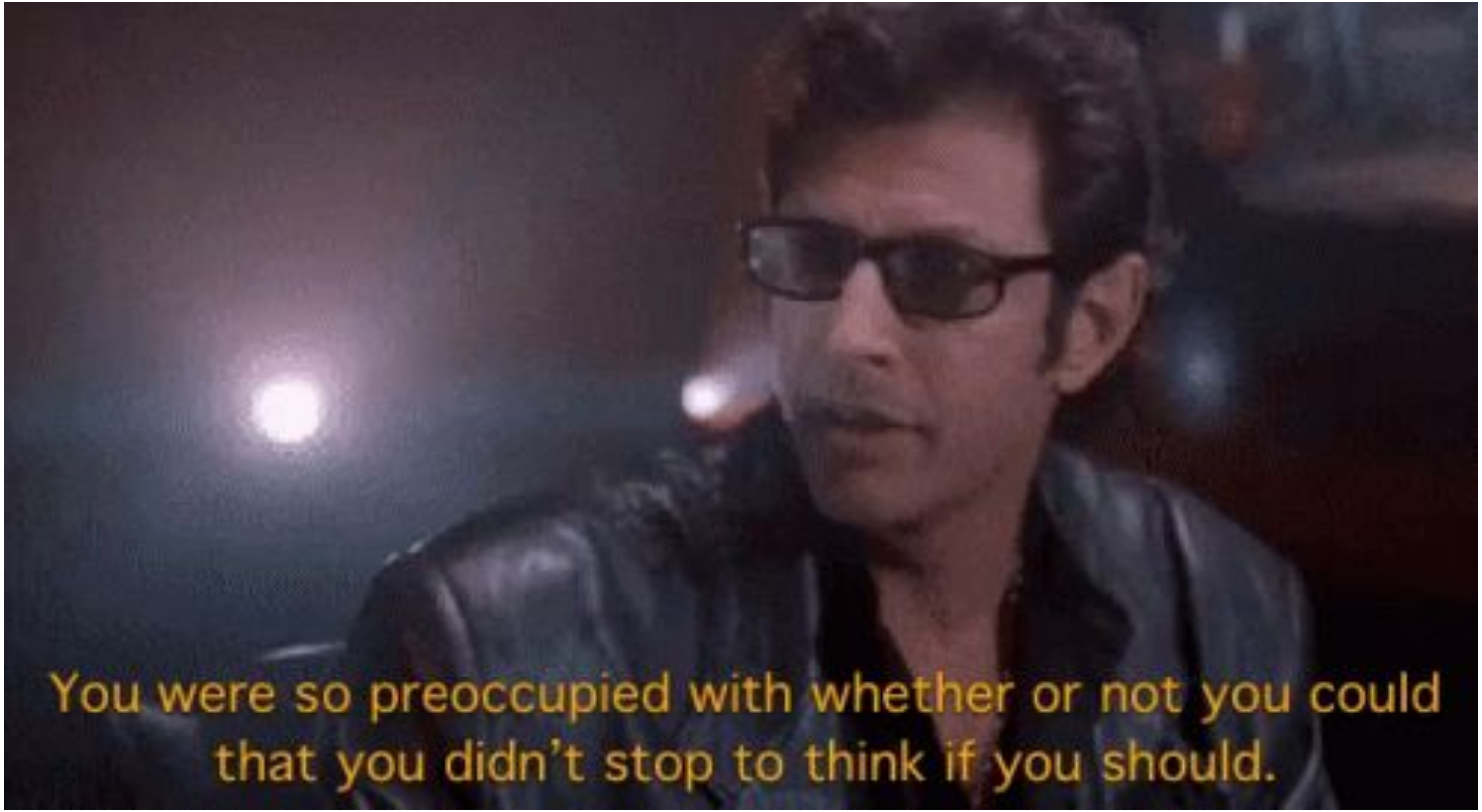
https://arxiv.org/pdf/2103.00020.pdf

Connecting Text and Images (CLIP)

Putting this all together

How can we put this together (notebooks)

What should we do with this?



What should we do with this? aka does anyone actually read a model card 😅?

Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru

Trained machine learning models are increasingly used to perform high-impact tasks in areas such as law enforcement, medicine, education, and employment. In order to clarify the intended use cases of machine learning models and minimize their usage in contexts for which they are not well suited, we recommend that released models be accompanied by documentation detailing their performance characteristics. In this paper, we propose a framework that we call model cards, to encourage such transparent model reporting. Model cards are short documents accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions, such as across different cultural, demographic, or phenotypic groups (e.g., race, geographic location, sex, Fitzpatrick skin type) and intersectional groups (e.g., age and race, or sex and Fitzpatrick skin type) that are relevant to the intended application domains. Model cards also disclose the context in which models are intended to be used, details of the performance evaluation procedures, and other relevant information. While we focus primarily on human-centered machine learning models in the application fields of computer vision and natural language processing, this framework can be used to document any trained machine learning model. To solidify the concept, we provide cards for two supervised models: One trained to detect smiling faces in images, and one trained to detect toxic comments in text. We propose model cards as a step towards the responsible democratization of machine learning and related AI technology, increasing transparency into how well AI technology works. We hope this work encourages those releasing trained machine learning models to accompany model releases with similar detailed evaluation numbers and other relevant documentation.

https://arxiv.org/abs/1810.03993

What should we do with this? aka does anyone actually read a model card 😅?

Out-of-Scope Use Cases

Any deployed use case of the model - whether commercial or not - is currently out of scope. Non-deployed use cases such as image search in a constrained environment, are also not recommended unless there is thorough in-domain testing of the model with a specific, fixed class taxonomy. This is because our safety assessment demonstrated a high need for task specific testing especially given the variability of CLIP's performance with different class taxonomies. This makes untested and unconstrained deployment of the model in any use case currently potentially harmful.

Certain use cases which would fall under the domain of surveillance and facial recognition are always out-of-scope regardless of performance of the model. This is because the use of artificial intelligence for tasks such as these can be premature currently given the lack of testing norms and checks to ensure its fair use.

Since the model has not been purposefully trained in or evaluated on any languages other than English, its use should be limited to English language use cases.

Source: https://github.com/openai/CLIP/blob/main/model-card.md#model-use

What should we do with this? aka does anyone actually read a model card 😅?

Out-of-Scope Use Cases

Any deployed use case of the model - whether commercial or not - is currently out of scope. Non-deployed use cases such as image search in a constrained environment, are also not recommended unless there is thorough in-domain testing of the model with a specific, fixed class taxonomy. This is because our safety assessment demonstrated a high need for task specific testing especially given the variability of CLIP's performance with different class taxonomies. This makes untested and unconstrained deployment of the model in any use case currently potentially harmful.

Certain use cases which would fall under the domain of surveillance and facial recognition are always out-of-scope regardless of performance of the model. This is because the use of artificial intelligence for tasks such as these can be premature currently given the lack of testing norms and checks to ensure its fair use.

Since the model has not been purposefully trained in or evaluated on any languages other than English, its use should be limited to English language use cases.

Source: https://github.com/openai/CLIP/blob/main/model-card.md#model-use