# 1 Appendix B: Fine-tuning ColPali

Edited version of Merve Noyan's notebook

This notebook is a very minimal example to fine-tune ColPali on UFO documents and queries a dataset synthetically generated.

Then we will show a very minimal example on how to retrieve infographics.

[1]:
```
import os

os.environ["CUDA_DEVICE_ORDER"] = "PCI_BUS_ID"
os.environ["CUDA_VISIBLE_DEVICES"] = "1"
os.environ["NCCL_P2P_DISABLE"] = "1"
os.environ["NCCL_IB_DISABLE"] = "1"

import pandas as pd
import torch
from colpali_engine.loss import ColbertPairwiseCELoss
from datasets import DatasetDict, load_dataset
from huggingface_hub import login
from peft import LoraConfig, get_peft_model
from transformers import BitsAndBytesConfig, ColPaliForRetrieval,␣
↪ColPaliProcessor, Trainer, TrainingArguments
```

[2]:
```
secret = pd.read_csv("secret.config", header=None)
HF_TOKEN = secret[1][1]
del secret
login(HF_TOKEN)
```

## 1.1 Loading the Model

Fine-tuning ColPali takes around 48 GB of VRAM, which is way too much for an RTX 4090. To overcome memory limits, we can apply QLoRA to only train an adapter and load the model in a lower precision (4-bit). Furthermore, we reduce the batch size and compensate for it with gradient accumulation.

[3]:
```
model_name = "vidore/colpali-v1.2-hf"

bnb_config = BitsAndBytesConfig(
    load_in_4bit=True,
    bnb_4bit_quant_type="nf4",
    bnb_4bit_compute_dtype=torch.bfloat16,
)
model = ColPaliForRetrieval.from_pretrained(
    model_name,
    torch_dtype=torch.bfloat16,
    quantization_config=bnb_config,
    device_map="cuda:0",
```

```
).eval()

lora_config = LoraConfig(
    r=8,
    lora_alpha=8,
    lora_dropout=0.1,
    target_modules=["down_proj", "o_proj", "k_proj", "q_proj", "gate_proj",␣
 ↪"up_proj", "v_proj"],
    init_lora_weights="gaussian",
)
lora_config.inference_mode = False
model = get_peft_model(model, lora_config)
processor = ColPaliProcessor.from_pretrained(model_name)
```

```
Loading checkpoint shards:   0%|          | 0/2 [00:00<?, ?it/s]
```

### 1.2 Load the dataset

We will use this dataset created by Daniel van Strien. In this blog post he explains very thoroughly how to create a dataset for retrieval tasks.

[4]:
```
ds = load_dataset("davanstrien/ufo-ColPali")
ds = ds["train"].train_test_split(test_size=0.1, seed=42)
train_ds = ds["train"]
test_ds = ds["test"]

ds
```

[4]:
```
DatasetDict({
    train: Dataset({
        features: ['image', 'raw_queries', 'broad_topical_query',
'broad_topical_explanation', 'specific_detail_query',
'specific_detail_explanation', 'visual_element_query',
'visual_element_explanation', 'parsed_into_json'],
        num_rows: 2018
    })
    test: Dataset({
        features: ['image', 'raw_queries', 'broad_topical_query',
'broad_topical_explanation', 'specific_detail_query',
'specific_detail_explanation', 'visual_element_query',
'visual_element_explanation', 'parsed_into_json'],
        num_rows: 225
    })
})
```

We need to get rid of dataset items where our text query column is None.

[5]:
```
train_ds = train_ds.filter(lambda example: example["specific_detail_query"] is␣
↪not None)
train_ds  # should be less than 2018
```

[5]:
```
Dataset({
    features: ['image', 'raw_queries', 'broad_topical_query',
'broad_topical_explanation', 'specific_detail_query',
'specific_detail_explanation', 'visual_element_query',
'visual_element_explanation', 'parsed_into_json'],
    num_rows: 1956
})
```

The dataset contains documents about UFOs and queries that might be related to document. Take alook at examples shortly.

[6]:
```
print(train_ds[0]["specific_detail_query"])
display(train_ds[0]["image"])
```

```
1976 Ufo sightings in Norway and Denmark
```

# Har det hänt nåt?

**SAMMANSTÄLLNING AV THORVALD BERTHELSEN**

***Det skall redan nu sägas, att 1976 kommer att betraktas som ett bra UFO-år. D v s ett år med högre UFO-aktivitet än normalt. Senast detta inträffade var 1973.***

***Tendensen tycks vara likadan i både Norge och Danmark, men resultaten från övriga världen kan vi bedöma först senare.***

***Pressläggningen av detta nummer sker den 12 december och fortfarande strömmar det in en mängd rapporter, som först skall utredas, och inte hinner komma med.***

Denna sammanställning har blivit ganska omfångsrik och intressanta rapporter, som borde ha behandlats i fristående artiklar, har måst "klämmas" in här.

Retroaktivt material får inte plats denna gång utan sparas till kommande sammanställningar, såvida intensiteten inte fortsätter — då blir det trångt! Andra artiklar av nyhetsmässigt värde för våra läsare pockar på utrymme och allt måste med för att inte mista nyhetens behag.

Ett varmt tack till fältforskare, lokala UFO-Sverigegrupper, distriktschefer och övriga medarbetare, som bidragit med arbete och material på ett snabbt och effektivt sätt och möjliggjort dessa uppskattade sammanställningar.

### LYSANDE METALLFÄRGAT FÖREMÅL

**Västergötland, Fristad** 2 oktober 1976 klockan 21.00

Fem personer iakttog ett lysande föremål över plastfabriken. Mats Rydén berättar:

— Det var ett lysande metallfärgat föremål, som ljudlöst kom glidande. Det hade den traditionella tefatsformen, d v s två tallrikar lagda mot varandra.

— Farkosten tycktes komma från Bredaredshållet, men avståndet är omöjligt att uppskatta. Den var synlig i någon minut och försvann sedan. Vi fick den uppfattningen, att den gick rakt emot oss. Den sänkte sig sedan, innan den försvann.

En stund senare iakttog Anita Tancred på Trandaredsområdet i Borås ett kraftigt ljussken. Det var så starkt att hon först trodde, att en eldsvåda brutit ut på vindsvåningen i huset mittemot.

### LJUSPUNKT PÅ RAK KURS

**Gästrikland, Torsåker** 9 oktober 1976 klockan 17.30

Sven Johansson i Kratte masugn iakttog ett föremål av ungefär aftonstjärnans storlek på vägrät linje från söder mot norr. Han iakttog det under ett par minuter och det lyste med ett klart fast sken.

### ÅTERKOM OCH SKIFTADE FÄRG

**Västergötland, Borås** 12 oktober 1976 klockan 19.30

Jerry Lindgren, boende i Norrby, iakttog ett föremål som förflyttade sig från Trandared mot Ryda och tillbaka igen.

Det lyste med en ljus färg, men när det återkom skiftade det färg till rött och violett. Det sänkte sig och tycktes då avge ljusstrålar mot marken.

### TROR SIG HA SETT SAMMA FÖREMÅL FLERA GÅNGER

**Ångermanland, Sollefteå** 14 oktober 1976 klockan 19.45

Två personer iakttog, som de säger, ett egendomligt föremål över staden. Det beskrivs som ovalt och hade ett eldrött ljussken.

Efter den första observationen såg de föremålet ytterligare tre gånger under kvällen. Vid dessa tillfällen rörde det sig med stor hastighet.

### FARTYGSBESÄTTNING IAKTTOG TVÅ FÖREMÅL

**Östersjön, Utö—Åbo** 16 oktober 1976 klockan 01.00

Två klara och tydliga ljusfenomen iakttogs från m/s Sirius kommandobrygga, då fartyget var på väg från Utö till Lohm i Ålands skärgård.

Enligt lotsen Eskil Öhman kom föremålen från sydväst i riktning mot nordost. Till att börja med var storleken som en stjärna, men den ökade sedan i storlek.

Några minuter efter det första fenomenet uppenbarade sig ett annat likadant, som hade samma kurs, men lyste ännu starkare.

Från fartyget följde man fenomenen såväl med ögonen som med kikare. Besättningen är van vid att navigera nattetid men har aldrig tidigare sett något liknande.

### INTE VI, SÄGER FLYGET OM DETONATION

**Södermanland, Stockholms södra förorter** 21 oktober 1976 klockan 15.40

Vår representant Bo Sigerlöv i Grödinge rapporterar:

— Alla talar om den stora smällen, men ingen vet vad det egentligen var. Det var en dov knall, som fick fönsterrutorna att skallra och skrämde upp människor i hela kommunen.

— Det var en ovanligt kraftig detonation, som hördes över hela Stockholm och även hemma hos mig, utanför Södertälje. Knallen tycks ha uppstått över området Bredäng—Västertorp och polisens sambandscentral blev nedringd av oroliga människor, som trodde att krig brutit ut.

— Alla är ense om att smällen kom uppifrån och inga spår av en sprängning eller detonationskrater har kunnat hittas.

— På de närmaste flygflottiljerna, Nyköping, Västerås, Uppsala och Norrköping, säger man sig inte ha orsakat smällen. En förutsättning för att det skall kunna vara en flygplansljudbang och bli så kraftig, är att flygplanet flyger på otillåten låg höjd.

— Ett allvarligt fel inträffade vid 20-tiden i en transformatorstation vid Storsvängen i Västertorp samma dag och tusentals människor drabbades av strömavbrott ända till följande dag.

### STORT FÖREMÅL PÅ LÅG HÖJD ÖVER BOLLMORA

**Södermanland, Bollmora (S Sthlm)** 21 oktober 1976 klockan 17.40

Många människor i området iakttog denna torsdagskväll ett märkligt föremål. Fru Lena Eriksen berättar för vår representant Christer Nordin:

— Jag har aldrig tidigare iakttagit ett UFO, men vad skall man tro när man sett detta? Jag var på besök hos min mor när hon gjorde mig uppmärksam på en märklig farkost, som svävade fram över hustaken. Det var ett stort föremål på ungefär 1 500 meters höjd.

— I sakta fart svävade det fram. Det var stort som en helikopter och lyste med en cyklamenröd färg. Farkosten var rundad och försedd med två rader fyrkantiga fönster, som gick runt om.

— Från föremålets översida steg det upp någonting som syntes som en ljus rökpelare. Det avlägsnade sig mot Nacka-hållet och jag kunde inte uppfatta något ljud.

From this dataset we will have the following columns to create the documents and the queries:

- `image` contains our documents.
- `specific_detail_query` contains the textual queries.

[7]:
```
def collate_fn(examples):
    texts = []
    images = []

    for example in examples:
        texts.append(example["specific_detail_query"])
        images.append(example["image"].convert("RGB"))

    batch_images = processor(images=images, return_tensors="pt").to(model.
↪device)
    batch_queries = processor(text=texts, return_tensors="pt").to(model.device)
    return (batch_queries, batch_images)
```

### 1.3 Trainer

The trainer uses a ColBERT contrastive hard-margin loss. This loss is implemented in ColPali engine, it expects batch document and query embeddings, so essentially we need to process the documents and queries separately, then pass them to the model separately, then send to loss calculation.

Note that, since we are defining a custom loss, we have to subclass transformers Trainer to be able to pass it to the model.

[8]:
```
class ContrastiveTrainer(Trainer):
    def __init__(self, loss_func, *args, **kwargs):
        super().__init__(*args, **kwargs)
        self.loss_func = loss_func

    def compute_loss(self, model, inputs, num_items_in_batch=4,␣
↪return_outputs=False):
        query_inputs, doc_inputs = inputs
        query_outputs = model(**query_inputs)
        doc_outputs = model(**doc_inputs)
        loss = self.loss_func(query_outputs.embeddings, doc_outputs.embeddings)
        return (loss, (query_outputs, doc_outputs)) if return_outputs else loss

    def prediction_step(self, model, inputs):
        query_inputs, doc_inputs = inputs  # unpack from data collator
        with torch.no_grad():
            query_outputs = model(**query_inputs)
            doc_outputs = model(**doc_inputs)

            loss = self.loss_func(query_outputs.embeddings, doc_outputs.
↪embeddings)
```

```
            return loss, None, None
```

[9]:
```
training_args = TrainingArguments(
    output_dir="./colpali_ufo",
    num_train_epochs=1,
    per_device_train_batch_size=2,
    gradient_accumulation_steps=8,
    gradient_checkpointing=False,
    logging_steps=20,
    warmup_steps=100,
    learning_rate=5e-5,
    save_total_limit=1,
    report_to="wandb",
    dataloader_pin_memory=False,
)
```

[10]:
```
trainer = ContrastiveTrainer(
    model=model, train_dataset=train_ds, args=training_args,␣
 ↪loss_func=ColbertPairwiseCELoss(), data_collator=collate_fn
)

trainer.args.remove_unused_columns = False
```

We are training on a small dataset (little less than 2k examples) for one epoch so the training is fairly short (around 8 mins).

[11]:
```
trainer.train()
```

```
wandb: WARNING The `run_name` is currently set to the same
value as `TrainingArguments.output_dir`. If this was not intended, please
specify a different run name by setting the `TrainingArguments.run_name`
parameter.
wandb: Using wandb-core as the SDK backend.  Please refer to
https://wandb.me/wandb-core for more information.
wandb: Currently logged in as: polpastells
(clic). Use `wandb login --relogin` to force relogin
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
Could not estimate the number of tokens of the input, floating-point operations
will not be computed
```

```
<IPython.core.display.HTML object>
```

[11]:
```
TrainOutput(global_step=122, training_loss=0.05558847440559356,
metrics={'train_runtime': 465.0577, 'train_samples_per_second': 4.206,
'train_steps_per_second': 0.262, 'total_flos': 0.0, 'train_loss':
0.05558847440559356, 'epoch': 0.9979550102249489})
```

### 1.4 Load and test fine-tuned model

Let's try the fine-tuned model. You can simply test by passing in text-image pairs and check the scores for the ones that are actually pairs of each other, and also the scores of those that are irrelevant (i.e. all scores except for the scores of the matching ones).

[12]:
```
print(test_ds[0]["specific_detail_query"])
display(test_ds[0]["image"])
print(test_ds[1]["specific_detail_query"])
display(test_ds[1]["image"])
print(test_ds[2]["specific_detail_query"])
display(test_ds[2]["image"])
```

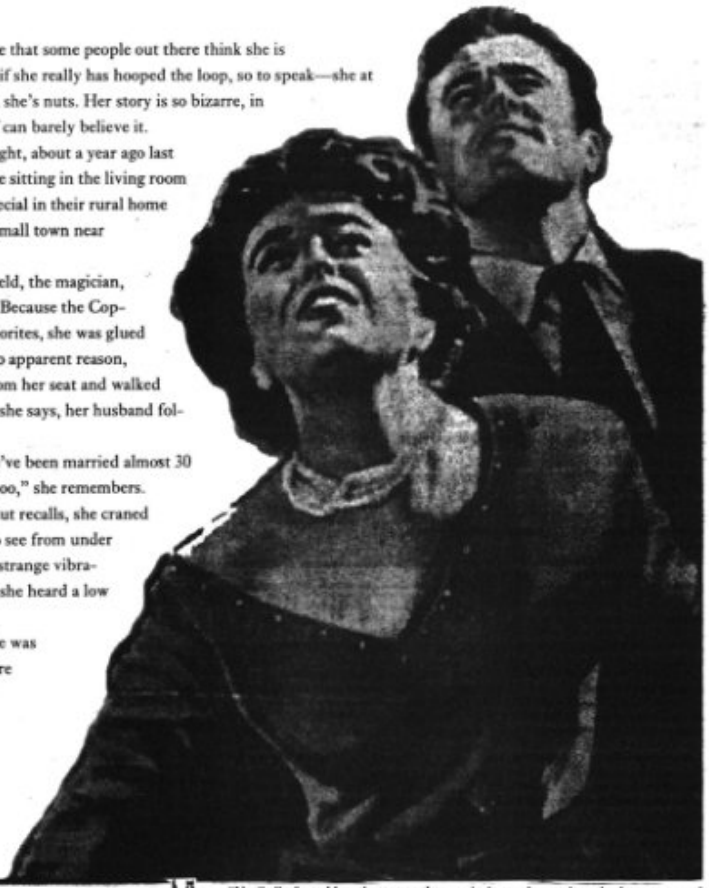
```
David Copperfield special
```

NASHVILLE SCENE, Nashville, TN - Oct. 26, 1995 CR: B. Kinnard

# OUT OF THIS WORLD

## UFO FlyBys in Middle Tennessee

BY JOYSA M. WINTER



LaVere Pisut is well aware that some people out there think she is crazy. If they're right—if she really has hooped the loop, so to speak—she at least *knows* people think she's nuts. Her story is so bizarre, in fact, that sometimes Pisut herself can barely believe it.

It happened one cozy spring night, about a year ago last April. Pisut and her husband were sitting in the living room watching a David Copperfield special in their rural home about 10 miles outside Baxter, a small town near Cookeville.

Pisut was excited that Copperfield, the magician, was performing some new tricks. Because the Copperfield program is one of her favorites, she was glued to the screen. Nevertheless, for no apparent reason, Pisut says, she abruptly got up from her seat and walked outside to her front porch. Soon, she says, her husband followed her outside as well.

"For him to follow me out—we've been married almost 30 years—that was kind of unusual too," she remembers.

While walking out the door, Pisut recalls, she craned her neck toward the sky, trying to see from under the eave of the porch. She felt a "strange vibration" through her shoulders, and she heard a low rumbling noise.

Her husband asked her what she was looking at, and for a moment, there was nothing to see. Until...

PHOTOS BY SUSAN HICKS

**They saw the light** *Jeff Hart (left) and Marc "Pugsley" Pisut say they witnessed a strange light above the treetops near Baxter in 1992. "I swear, we weren't drunk and we weren't stoned," Pisut says.*

"Until all of a sudden, there were these lights up above the house, either square or triangular in formation," Pisut says, pointing up to a barren spot just off her porch—a rare window of sky among acres of canopied trees. "It was very dark that night, no moon, and I couldn't see stars through it. That's how I know it was solid."

Pisut doesn't know how long she and her husband sat on the porch watching the object—which was as big as her house—hover in the air. But she guesses they stood there for about 20 minutes. And now, thinking back on it, all she can do is despair over the things she wishes she had done.

"Why didn't we turn out the porch light to see it better? Or grab a flashlight and shine it up, because that's how low it was. Or we could have run off the porch to get a better look. I don't know why we didn't do these things!"

Once the object flew away across the trees, Pisut says, there was a great flash of white light and then total silence. It was a queer stillness that blanketed the countryside.

"This was spring, and spring is not quiet," she says. "We had 150 chickens and roosters, 11 geese, five dogs, and there was no sound. I don't even know where the dogs were—and they always come up when we're outside. No frogs, no peep from the goat yard next to my house. We're next to a pond but no noise from the crickets. Absolute, dead silence."

When it was all over, and Pisut had regained some of her senses, she ran inside and called her sister, who lives next door on the other side of the wooded thicket.

"LaVere called me at about 10 till 10," says Gaylene Fields, sitting on a step on her sister's front porch, staring at the blank spot in the sky where stars are now twinkling, the same spot where all this bizarreness started. "I asked her where it was headed, and my son and I got in the car and drove west toward Granville. We could see an orangey glow at the horizon, so brilliant that we thought it had crashed. Halfway there, we smelled this acrid, chemical smell we had never smelled before."

But they say they lost sight of the object. Once they had arrived in Granville, Fields and her son found some Jackson County police and a sheriff who followed them back to her sister's house. An hour later, two Putnam

(continued on page 6)

Snake with legs and feet reported in Africa, 1899

## A Waco Snake.

Waco Telephone: J. W. Boynton brought to the Telephone yesterday afternoon what was a genuine curiosity and the existence of which has been denied by some people. It was nothing more nor less than a snake with legs. The snake which was a small one, not more than sixteen inches long, was what is known as a "thunder snake," the body being covered by alternate splotches of black and red, intermingled in a manner which made a very pretty effect, almost causing you to forget that you were looking at a serpent. The snake had two legs, each about two inches long, and the boys who killed him claim that he pulled himself along very rapidly by their aid. The legs were set opposite one another under his body about four inches from his head. Each leg had four toes or feelers.

Above: Fort Worth Morning Register August 12th 1899

**SNAKE WITH LEGS AND FEET IS REPORTED TO BE**

**FOUND IN AFRICA**

The Sacramento Bee July 20th 1950

"A snake with four legs,feet and joints is reported to have been killed by the district commissioner in Kalabo, a distant area in the North Rhodesian bush. Many strange tales of unusual snakes have come from this district in the past. Whistling snakes have been reported and once the natives insisted they had found a snake which could sing."

Document mentioning Dr. Hynek and the OVNI

Donald y Grant Jaroslaw

Mount Clemens, Michigan, muy cerca del lago St. Clair y de la Base Aérea de Selfridge. Como a las 14:30 observaron un objeto suspendido sobre el lago durante 10 minutos. En ese momento tomaron las 4 fotos con una cámara *Polaroid Swinger*, antes de que el objeto partiera a una *"velocidad mucho mayor"* que la de cualquier aparato convencional. *"Unos cinco minutos después que desapareció el objeto"* apareció un helicóptero sobrevolando el área.

Los muchachos mostraron las fotos a su madre y ésta, convencida de que eran auténticas, las llevó a los diarios locales. Los periodistas entrevistaron al Mayor **Raymond Nyls**, de la Base Silfridge, quien supuestamente declaró que a esa hora un helicóptero había sobrevolado el área. *"No estaba interesado en este reporte hasta que vi las fotografías* -dijo Nyls a los reporteros-. *Son las mejores fotos que jamás he visto de un OVNI. Hasta se puede apreciar lo que parece ser una antena en la parte de atrás"*.

Las fotografías parecían buenas ante los ojos de Nyls. No presentaban manchas ni cambios de posición que indicaran que se trataba de un modelo. Además, el uso de una cámara *Polaroid* impedía un trucaje en cuarto oscuro y los testigos parecían honestos y no tener

Ampliación de la maqueta utilizada por los hermanos Jaroslaw

interés de publicidad personal. El mismo Dr. **Hynek** investigó el caso y declaró: *"No tengo datos que me hagan dudar de la autenticidad de las fotos; hasta este momento, con toda honestidad, no puedo decir que se trate de un fraude, aunque tal posibilidad no debe eliminarse por completo"*.

La investigación efectuada por Hynek lo llevó a encontrar algunos datos curiosos que, sin embargo no despertaron sus sospechas. En primer lugar, el OVNI no fue captado por el radar de Selfridge, aunque el helicóptero si. Su explicación fue que el OVNI volaba a poca altura (¿y el helicóptero?). Posteriormente halló una pista, que de haber tenido un sentido más crítico y de haber hecho una buena investigación, lo hubieran llevado a encontrar la verdad. Al ser estudiadas las fotos se comprobó que la del helicóptero era la tercera; la cuarta y la quinta eran del OVNI, lo mismo que las dos primeras. Al ser cuestionados, los muchachos explicaron que seguramente vieron el helicóptero mientras observaban el OVNI, que se encontraban tan emocionados que luego no se acordaron bien. ¡Hynek se tragó tan burdo engaño!

Nueve años después tendríamos la respuesta a este caso. En 1976 Hynek recibió la siguiente carta:

Paris 67 a

[13]:
```
images = [test_ds[0]["image"], test_ds[1]["image"], test_ds[2]["image"]]
texts = [test_ds[0]["specific_detail_query"],␣
 ↪test_ds[1]["specific_detail_query"], test_ds[2]["specific_detail_query"]]

# process
batch_images = processor(images=images).to(model.device)
batch_queries = processor(text=texts).to(model.device)

# infer
with torch.no_grad():
    image_embeddings = model(**batch_images).embeddings
    query_embeddings = model(**batch_queries).embeddings

# Score the queries against the images
scores = processor.score_retrieval(query_embeddings, image_embeddings)
```

The matching text-image scores are on the diagon of the scores below, as you can see, they're matched correctly!

[14]:
```
scores
```

[14]:
```
tensor([[12.9375,  4.6250,  7.4375],
        [ 9.9375, 19.3750,  9.1875],
        [ 8.7500,  7.7188, 17.3750]], dtype=torch.bfloat16)
```