AGXK-20 Multimedia Retrieval System

Andreas Bucher
Gaudenz Halter
Xiao'ao Song
Kevin Steijn

AGENDA

What can you expect of this presentation?

Introduction & Project Management

What was our plan and approach?

Feature Acquisition

What Feature Engineering approaches did we use?

Application Architecture & Demo

How does our final application look and feel?

Challenges, Limitations & Lessons Learned

What were the challenges and limitations? What did we learn?

1. Introduction & Project Management

1. Introduction & Project Management

Overview

GOAL

Interactive video retrieval system

STANDARD QUERY TYPES

- Visual Known-Item Search
- Textual Known-Item Search
- Ad-hoc Video Search

DATASET

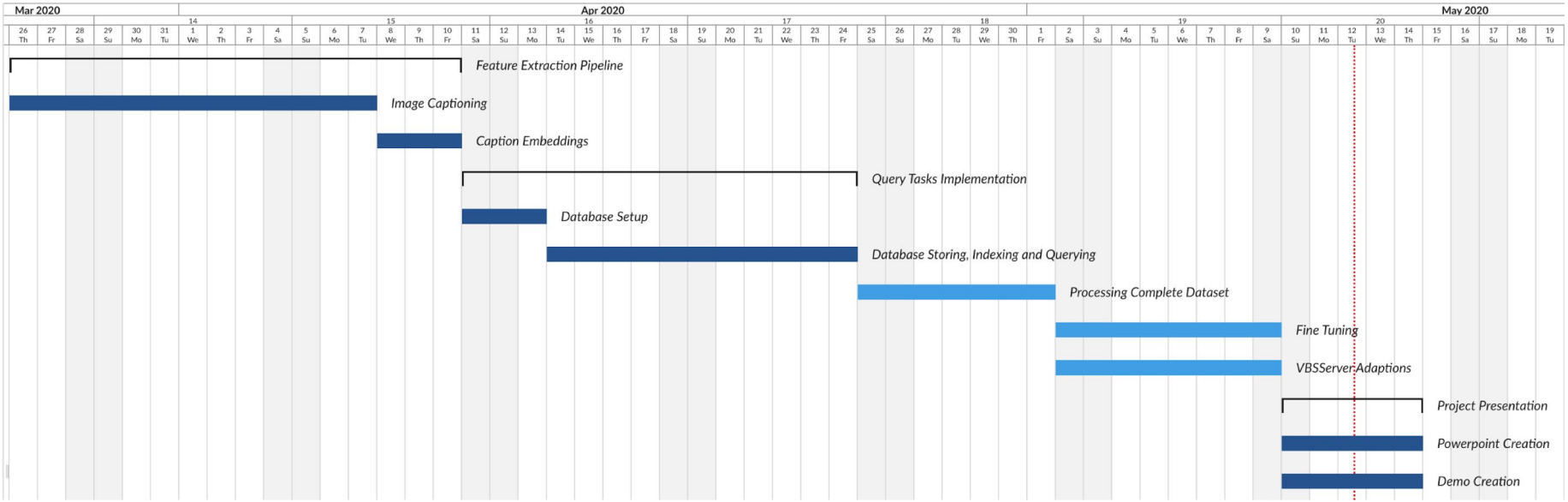
A subset of Vimeo Creative Commons Collection 1 (V3C1) of ~10%

EVALUATE

Video Browser Showdown (VBS) Server

1. Introduction & Project Management

Project planning and procedure




object: woman
object: landscape
2. Feature Acquisition
object: rock

2. Feature Acquisition

Shot Detection: Dataset of videos (subset of V3C1)

Thumbnails

Frames from source video

200x133 resolution

Keyframes

Frames from source video

Original resolution of video

Videos

Source Video

Info.json - Vimeo generated

Description of video - User submitted

“ {"extractor": "vimeo", "protocol": "https", ... , "fps": 29, "format_id": "http-360p"}]}“

“ See the creation of the Hydroponic Farm . . . . . and http://www.boswyckfarms.org.“

2. Feature Acquisition

Shot Detection: Scenedetect Library

Purpose

Uses changes in color and intensity to detect changes in scene

Detectors

Content detector

Threshold detector

Output

csv file with values for each frame

csv file with timestamps for start and end of scene

Potential development

New detector in development might perform better, such as Sparse Scene detector

Adjusting the detector based on the video may improve results

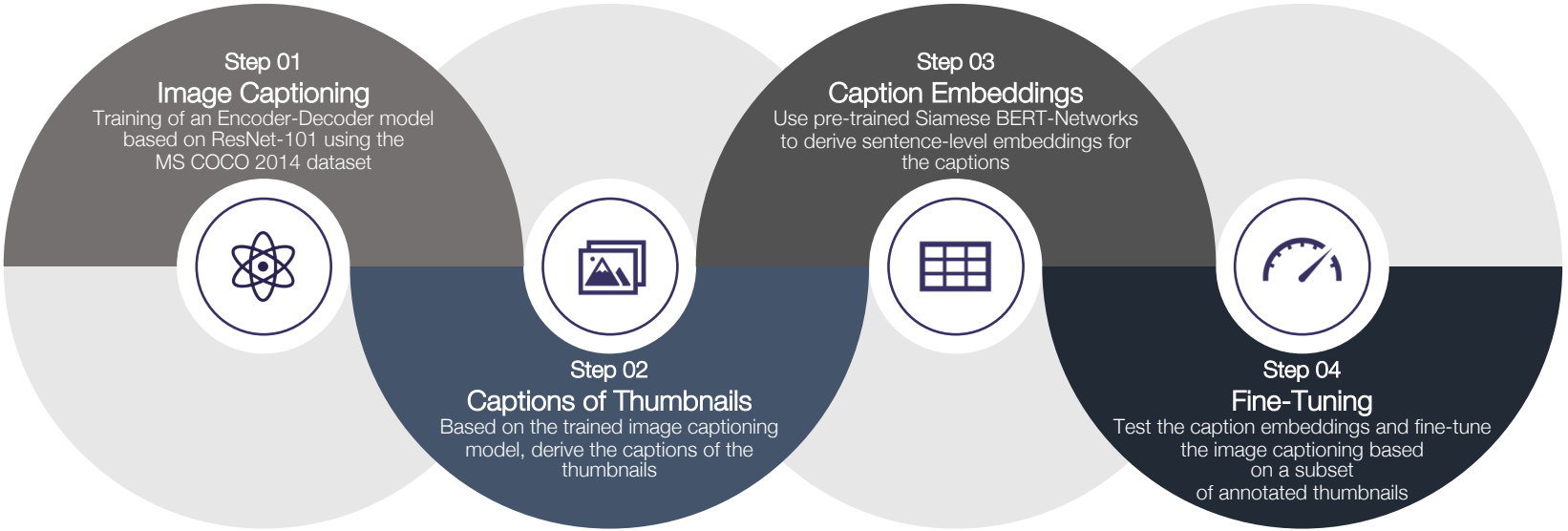
PySceneDetect

Creating
the CDSC Hydroponic Farm

0	5.30
5.30	34.56
.	.
.	.
.	.
247.31	250.98
250.98	283.58

2. FEATURE ACQUISITION

Image Captioning & Caption Embeddings - Feature Acquisition Flow



2. FEATURE ACQUISITION

Image Captioning using an Encoder-Decoder architecture

Encoder-Architecture

Pre-Trained ResNet-101 model, which is further fine-tuned during training

Decoder-Architecture

LSTM-decoder with additional linear layers to compute attention weights

Dataset

MS COCO 2014 with about 82k test and 41k validation images (each with five annotations) [1]

Training

About 38h training on an instance with 16GB GPU (P5000) and 30GB RAM

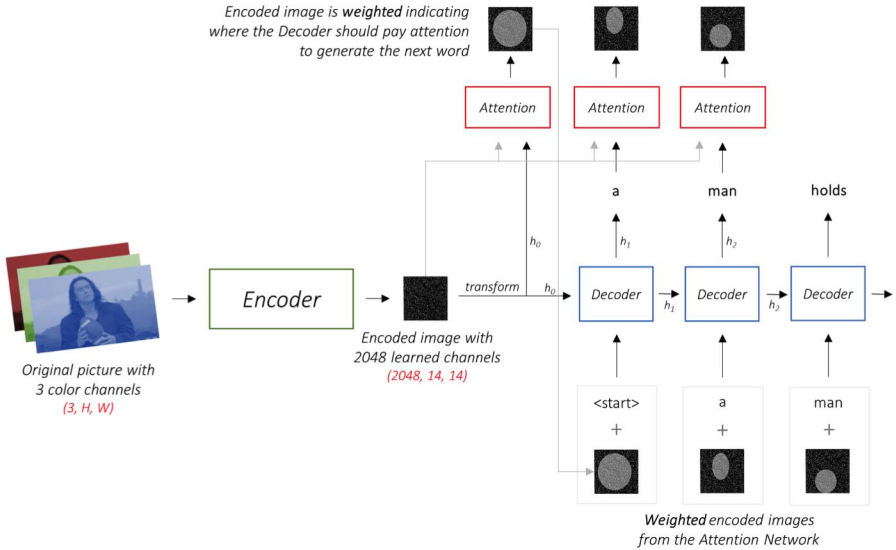
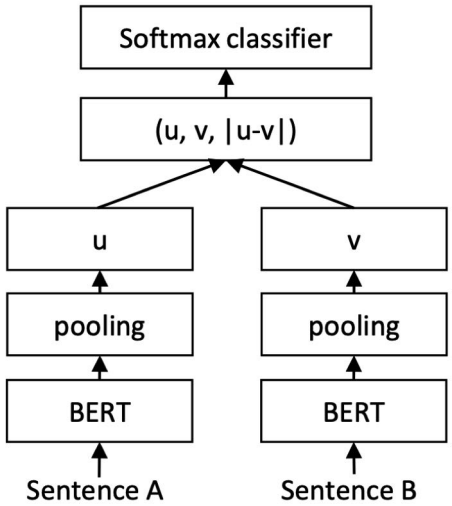


Image Captioning with Encoder-Decoder architecture [2]

Sources: [1] Lin et al. (2015), [2] Vinodababu, S. (2019)

2. FEATURE ACQUISITION

Caption Embeddings using Siamese BERT-Networks



Sentence-BERT[3]

Basic Idea behind this approach

- Use pre-trained Siamese BERT-networks to derive sentence-level embeddings from the generated image captions
- Similar sentences/ captions are embedded in a similar vector space
- Compare a search query against the caption embeddings using cosine similarity
- Transformer-based sentence embeddings allow for context-awareness

Sources: [3] Reimers, N., Gurevych, I. (2019)

2. FEATURE ACQUISITION

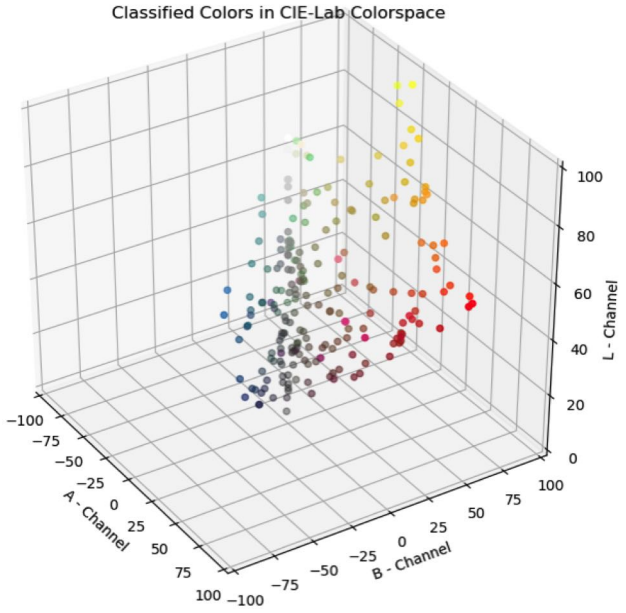
Color Classification

Goal:

Filtering shots by color name

Method:

- Extract k color patches from an image
    - Superpixel segmentation
    - KMeans clustering
- For each, find the closest color in a color dictionary
    - Since the CIELab-colorspace is perceptually uniform, we use the euclidean distance
- Add the color as tag



2. FEATURE ACQUISITION

Color Classification: Processing Steps

Input Image

Superpixel Extraction

k=5

k=10

k=20

KMeans - Clustering

2. FEATURE ACQUISITION

Color Classification: Results

k = 5

'Cream', 'Silk grey', 'Vermilion', 'Olive drab', 'Grey beige'

k = 10

'Signal white', 'Silk grey', 'Strawberry red', 'Quartz grey', 'Brown beige', 'Tomato red', 'Olive drab', 'Grey white', 'Olive yellow', 'Pebble grey'

2. FEATURE ACQUISITION

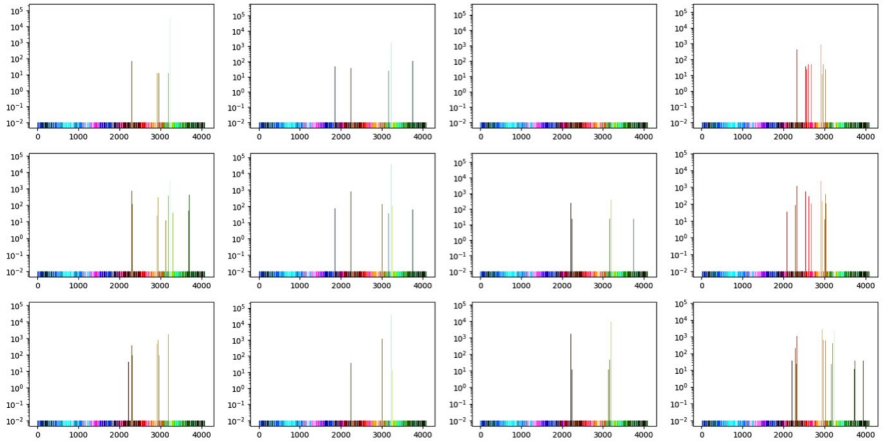
Color Histograms

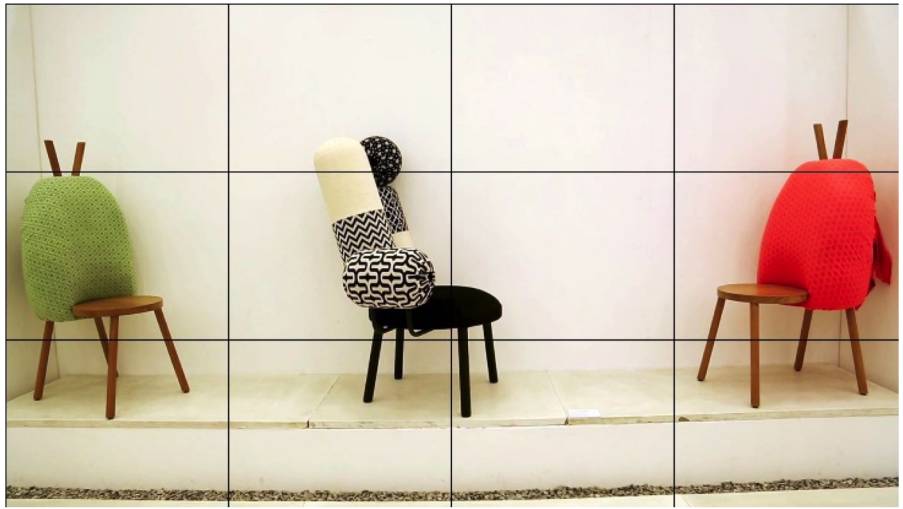
Goal

- Finding similar images by color
- Filter images by sketching

Method

- Split the image into a grid of subimages
- for each cell, compute the color histogram
- On lookup, compute the distance between the histograms for each cell





2. FEATURE ACQUISITION

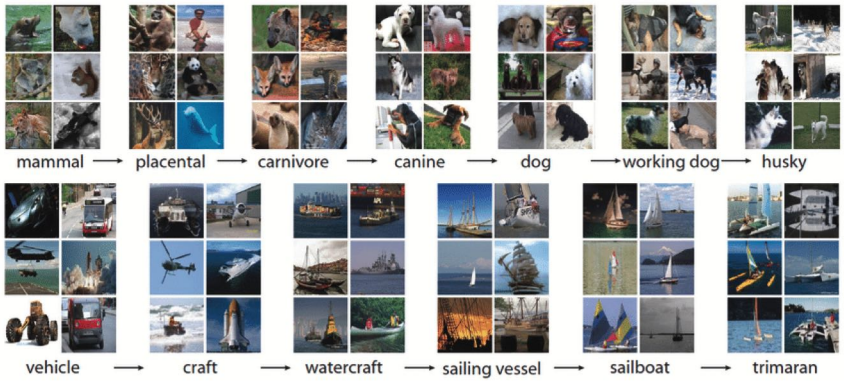
Object Recognition

Goal

- Finding images by object in the image

Method

- Extract Objects with an Xception model
- Add the names with a probability over a threshold as tags
- The model has been pre-trained on 1000 labels of the image-net dataset
- Fallback strategy if the caption would have failed



3. Application & Demo

3. Application

System Overview

Framework

Back-End: Flask WebApp
Front-End: Bootstrap + jQuery

Possible Queries:

Caption embedding
Keywords (objects, color names)
Sketch drawings
Similar images (color histogram)

User Interactions

Sub-Queries
Collaborative bookmarks
Search History

Database

SQLite: keywords, shots
HDF5: feature vector comparison

3. Application

User Interface

Basic Elements

1. Search bar
2. Sketch area
3. Results list
4. Bookmarks
5. History

Query / SubQuery

- Query looks in the whole database
- Subquery only looks in the already retrieved results

Submitting to VBS Server

Clicking on a bookmarked image sends it as result to the server

3. Application

Collaborative Bookmarks

Basic Idea

Bookmarks are synchronized between different users of the application

If a user clicks on an image in the results, it gets stored in the bookmarks

If one finds a good image, other can see them, too.
And can use them as a starting point for their own query.

AGXK-20

127.0.0.1:5000

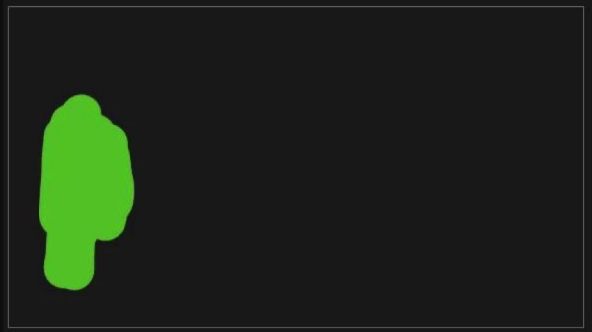
80%

AGXK-20

tree

Query | Sub Query | Embedding | Keywords

Sketch



Stroke

Clear

Query | Sub Query

Weniger

#61cb31

Abbrechen | Wählen

Bookmarked

History

Fotos

Type here to search

DEU

2:56 PM
5/11/2020

4. Conclusion

4. Conclusion

Challenges, Limitations and Lessons Learned

01 Challenges

- Image captioning model is trained on “static” images, not on thumbnails
- High requirements on computing infrastructure for model training
- Knowledge in many different areas required
- Organizational challenges due to curren situation

02 Limitations

- Sometimes, the generated image captions are inaccurate
- Sentence embeddings emphasize particular words of the captions
- Color Histograms can be quite sensitive to specific regions in the colorspace
- Colors haven’t been extracted per-object

03 Lessons Learned

- Sketching becomes handy for known-visual-item search
- Searching is an iterative process

4. Conclusion

Summary

Summary

- Searching by a variety of information
    - Textual: Image captions, color names, objects
    - Color: Similar images, sketching
- Collaborative
- Sketching comes handy for known-visual-item search