Welcome

AMLD 2021 - Sketchy

VISIUM

The team

Senior Machine Learning Engineer

Gaetan

Senior Machine Learning Engineer

Thibault

Junior Machine Learning Engineer

Olivier

Junior Machine Learning Engineer

Pauline

VISIUM

Links for the workshop

Github Repository (with links to Collab notebooks):
https://github.com/VisiumCH/AMLD-2021-Sketchy

Interactive Web Application:
https://amld.visium.ch

Todays' program

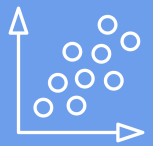
1

Retrieval Systems

2

Sketch-based Image Retrieval

3



Encoder and Embeddings

4

Attention Module

VISIUM

Part 1

1

Retrieval Systems

2

Sketch-based Image Retrieval

3

Encoder and Embeddings

4

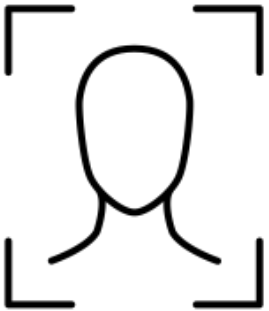
Attention Module

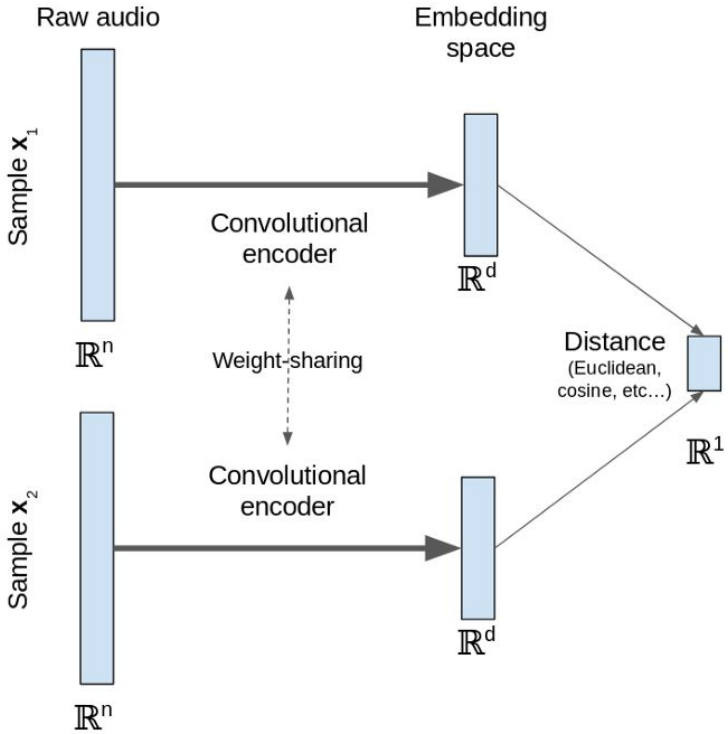
VISIUM

Item Retrieval Systems

Speaker Identification

Face Recognition





VISIUM

Part 2

1

Retrieval Systems

2

Sketch-based Image Retrieval

3

Encoder and Embeddings

4

Attention Module

Sketch-Based Image Retrieval

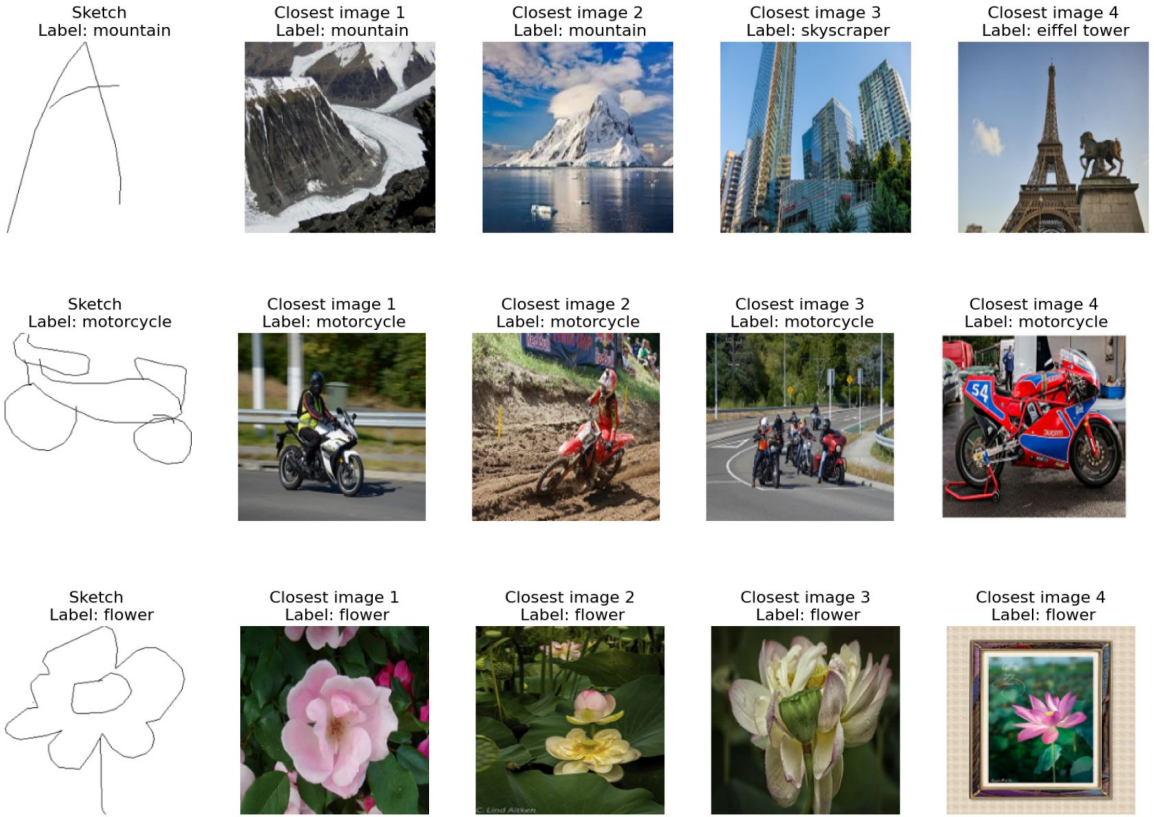
Image Database

Draw Sketch

Retrieve Image

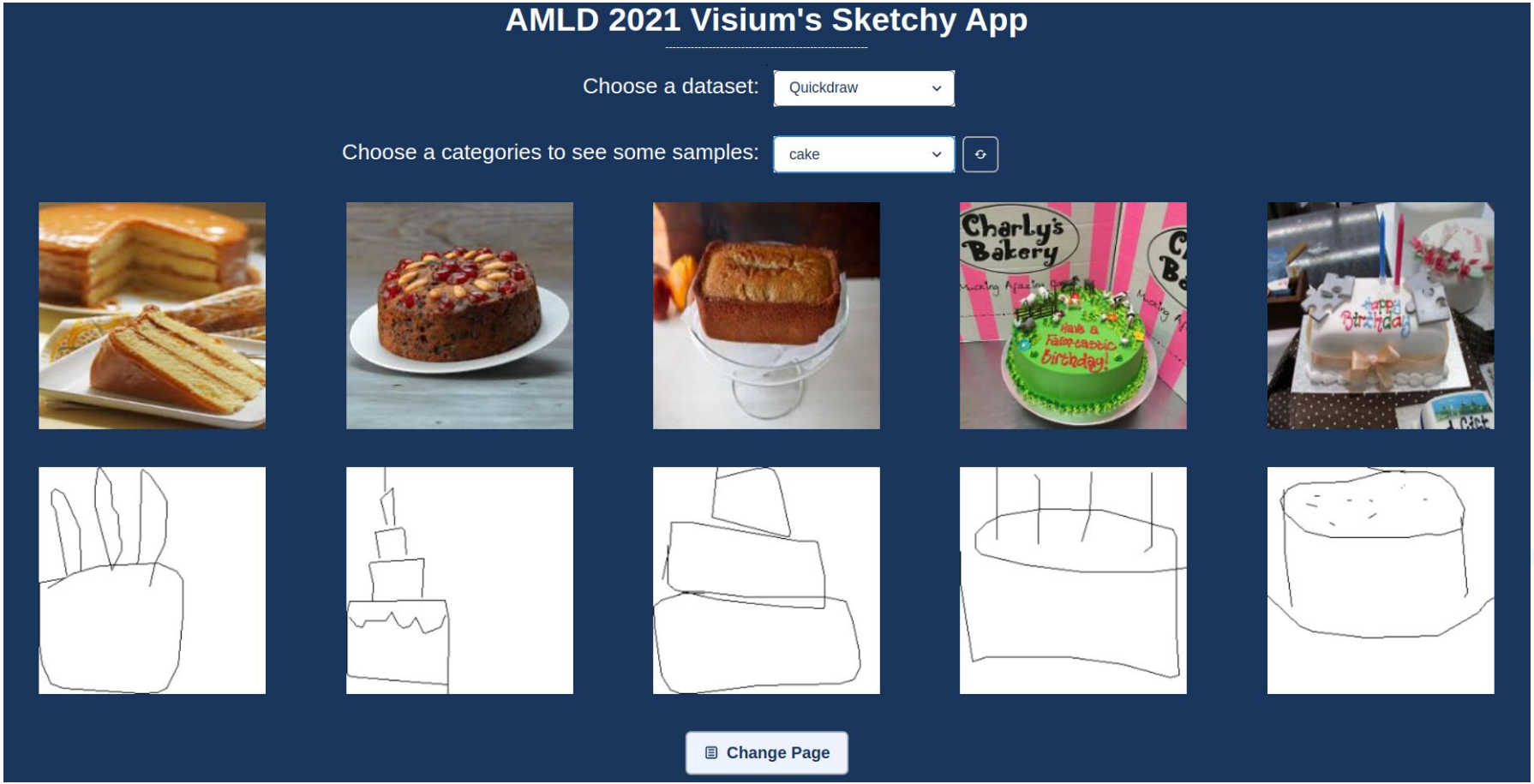
VISIUM

Sketch-Based Image Retrieval

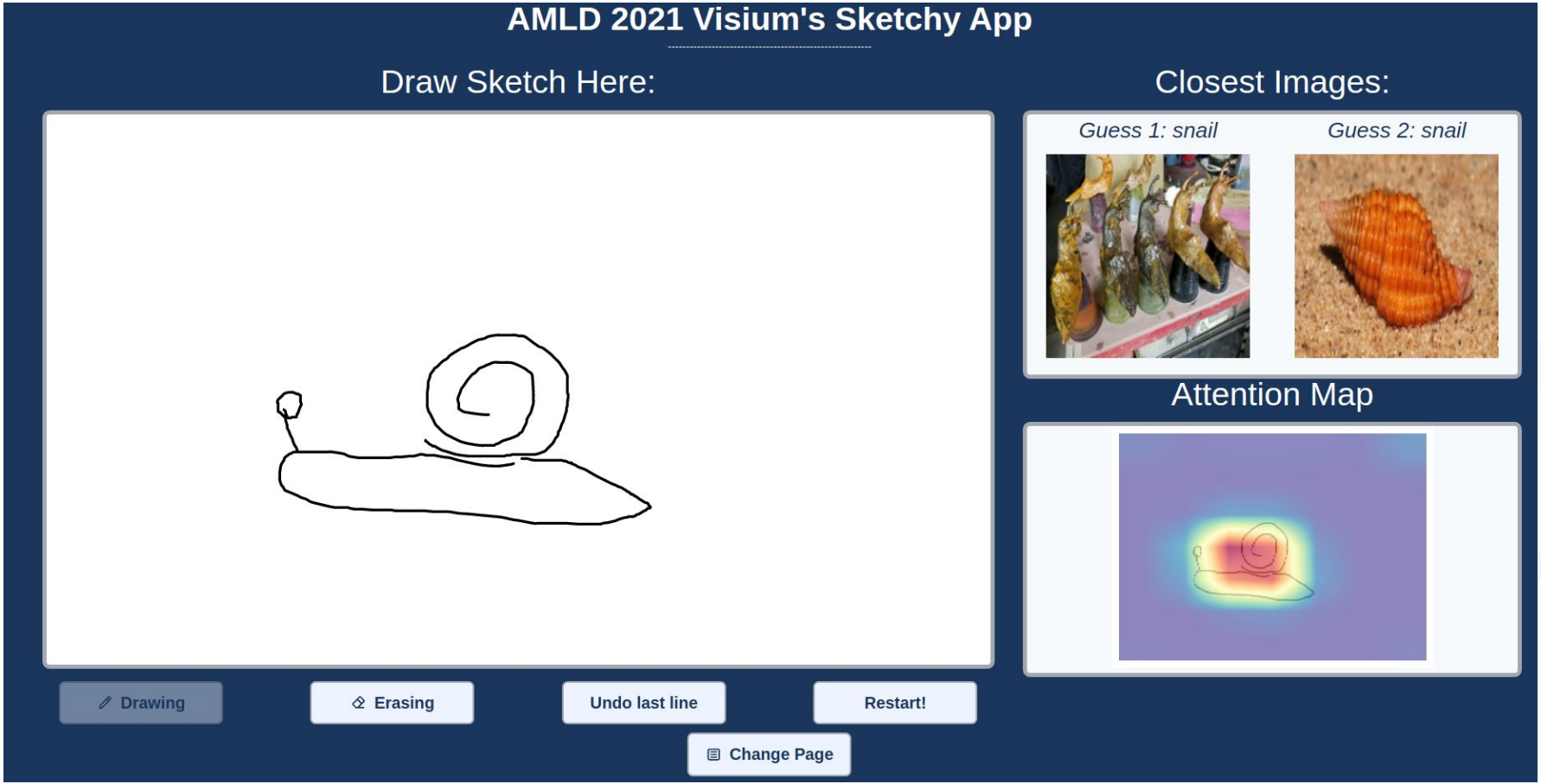


VISIUM

Sketch-Based Image Retrieval (amld.visium.ch)



Sketch-Based Image Retrieval



Part 3

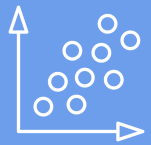
1

Retrieval Systems

2

Sketch-based Image Retrieval

3



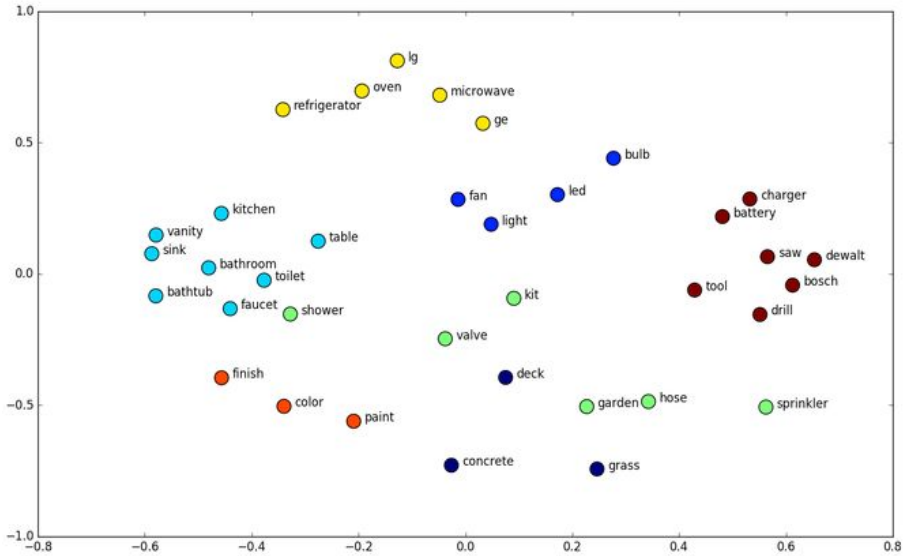
Encoder and Embeddings

4

Attention Module

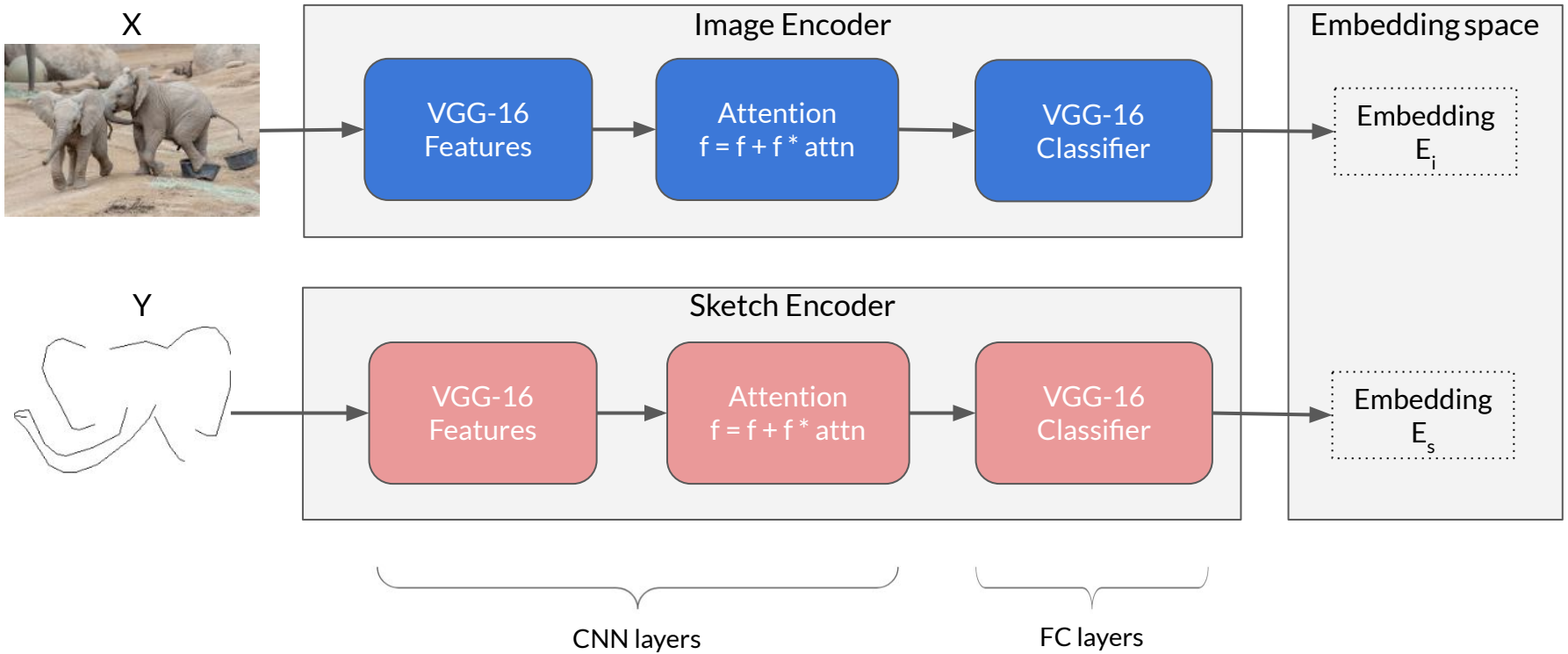
Embeddings

- Embeddings are a low dimensional representation of the semantic of higher dimensional data



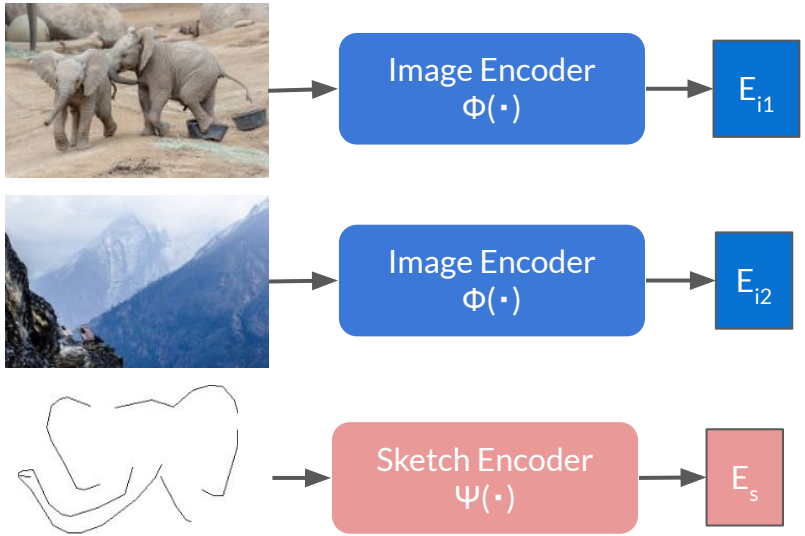
VISIUM

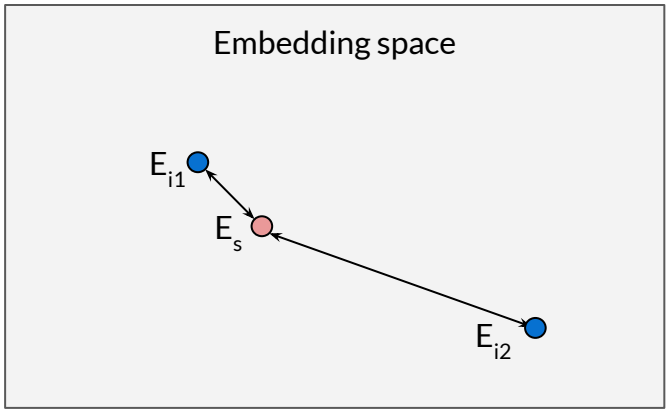
Encoder


X
Image Encoder
VGG-16
Features
Attention
f = f + f * attn
VGG-16
Classifier
Embedding space
Embedding
E_i
Y
Sketch Encoder
VGG-16
Features
Attention
f = f + f * attn
VGG-16
Classifier
Embedding
E_s
CNN layers
FC layers

Encoder and Embeddings

- Networks map images and sketches to ***Embeddings.***
- The closest image in the lower dimensional space is retrieved.





E_{i1} is retrieved

Embeddings

AMLD 2021 Visium's Sketchy App

Embeddings: Images and Sketches in 3D after TSNE projection

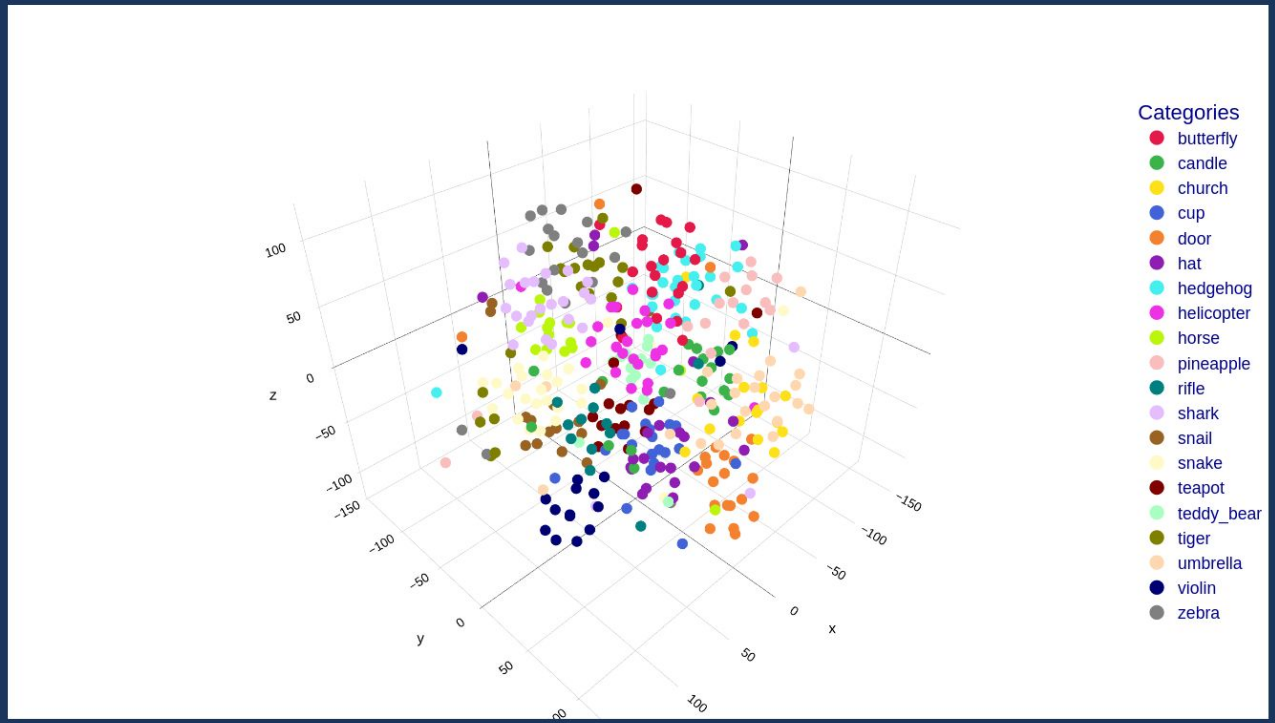
Options

2D 3D

PCA TSNE UMAP

My Sketch

Clicked image



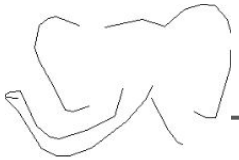
Change Page

Encoder and Embeddings Training

Image Encoder
Φ(·)

E_{i1}

positive



Sketch Encoder
Ψ(·)

E_s

anchor

Triplet Margin Loss

Image Encoder
Φ(·)

E_{i2}

negative

Make **positive and anchor close** to each other and **far from negative**.

$$\delta_- = ||\Psi(a) - \Phi(n)||_2$$

$$\delta_+ = ||\Psi(a) - \Phi(p)||_2$$

$$\lambda(\delta_+, \delta_-) = max\{0, \mu + \delta_+ - \delta_-\}$$

$$L_t = \frac{1}{N} \sum_{i=1}^{N} \lambda(\delta_+^i, \delta_-^i)$$

VISIUM

Encoder and Embeddings Training

Force **sketch and image embeddings to same area in embedding space**.
Reverse gradient of classifier.

$$l_t(e) = tlog(f(R_{\lambda_d}(e))) + (1-t)log(1-f(R_{\lambda_d}(e)))$$

$$L_d = \frac{1}{3N}\sum_{i=1}^{N}(l_0(\Psi(a_i)) + l_1(\Phi(p_i)) + l_1(\Phi(n_i)))$$

Encoder and Embeddings Training

Image Encoder
Φ(·)

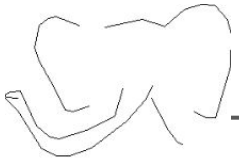
E_{i1}

GRL λ_d

Domain
Classifier f(·)

Domain
Loss

positive



Sketch Encoder
Ψ(·)

E_s

anchor

Triplet
Margin Loss

negative

Image Encoder
Φ(·)

E_{i2}

Train with both losses.

$$L_{tot} = w_d * L_d + w_t * L_t$$

VISIUM

Part 4

1

Retrieval Systems

2

Sketch-based Image Retrieval

3

Encoder and Embeddings

4

Attention Module

VISIUM

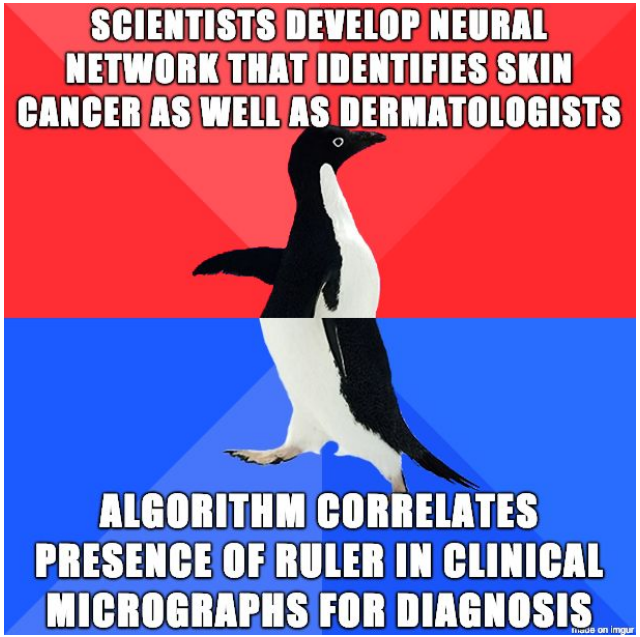
(Soft)-Attention

- Powerful concept in deep learning
- Help localise and highlight important features
- Assign weight depending on region of feature map
  - $x = x + x * attn(x)$

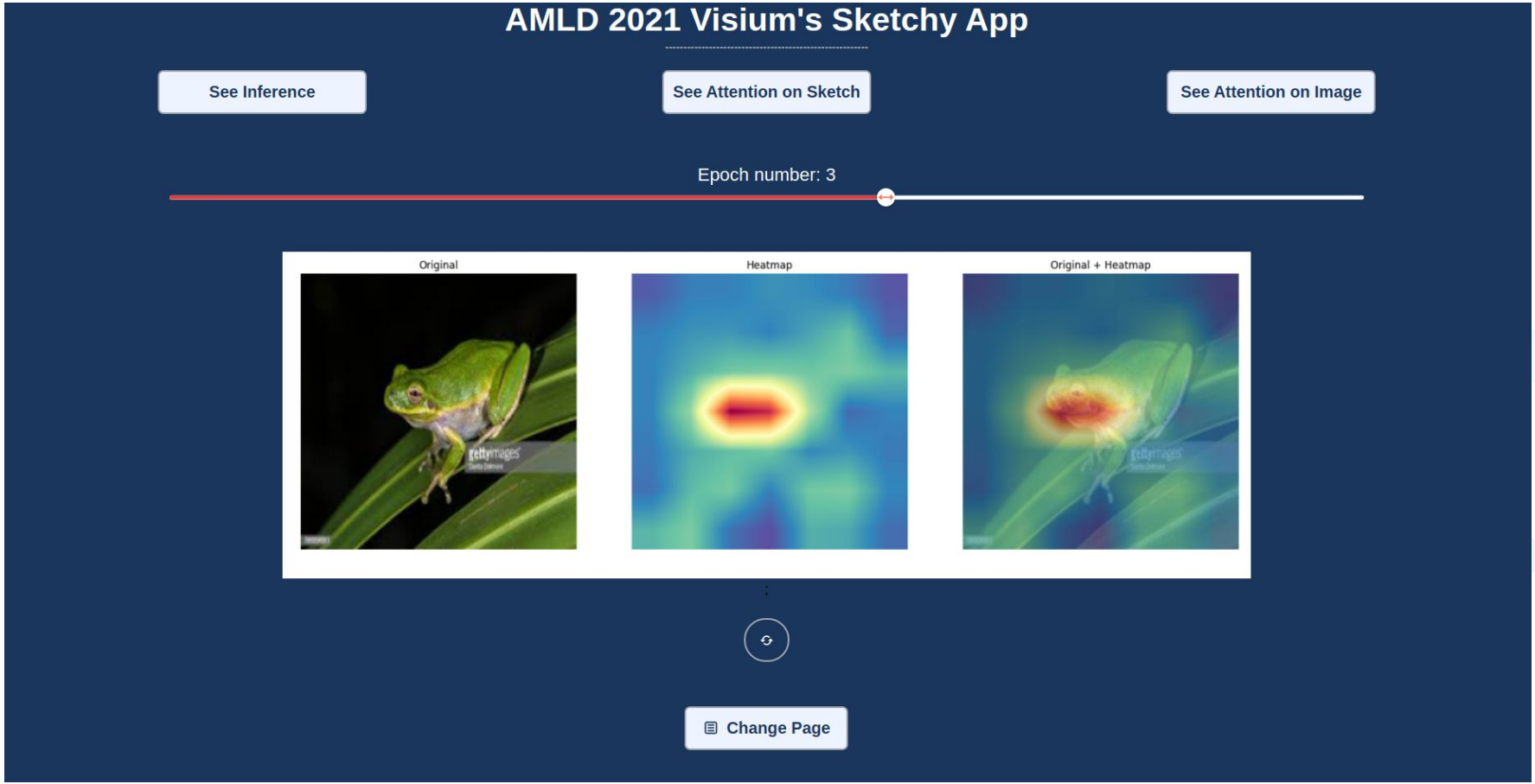
VISIUM

Attention

Visualise attention for sanity check



Attention in Web App



Jupyter Notebooks

Github Repository (with links to Collab notebooks):
https://github.com/VisiumCH/AMLD-2021-Sketchy

1. Training:
    a. Implement the network structure
    b. Understand and implement the different losses

2. Performance:
    a. Compute the performance metrics
    b. Implement the inference pipeline and visualize results

Takeaways

- Item retrieval systems fundamentals
- Always break down complex problems into digestible parts

To go further:

- You can also include other losses to change the behaviour of the model. For example, **[1]** include a semantic loss based on word embedding similarities

[1]: Doodle To Search: Practical Zero-Shot Sketch-based Image Retrieval, Dey, Sounak and Riba, Pau and Dutta, Anjan and Lladós, Josep and Song, Yi-Zhe

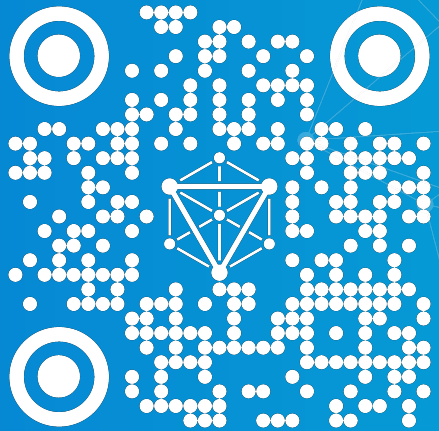
EPFL Innovation Park, 1015 Lausanne

Technopark Zurich, 8005 Zurich

www.visium.ch

+41 21 691 55 30

gr@visium.ch, tc@visium.ch



VISIUM