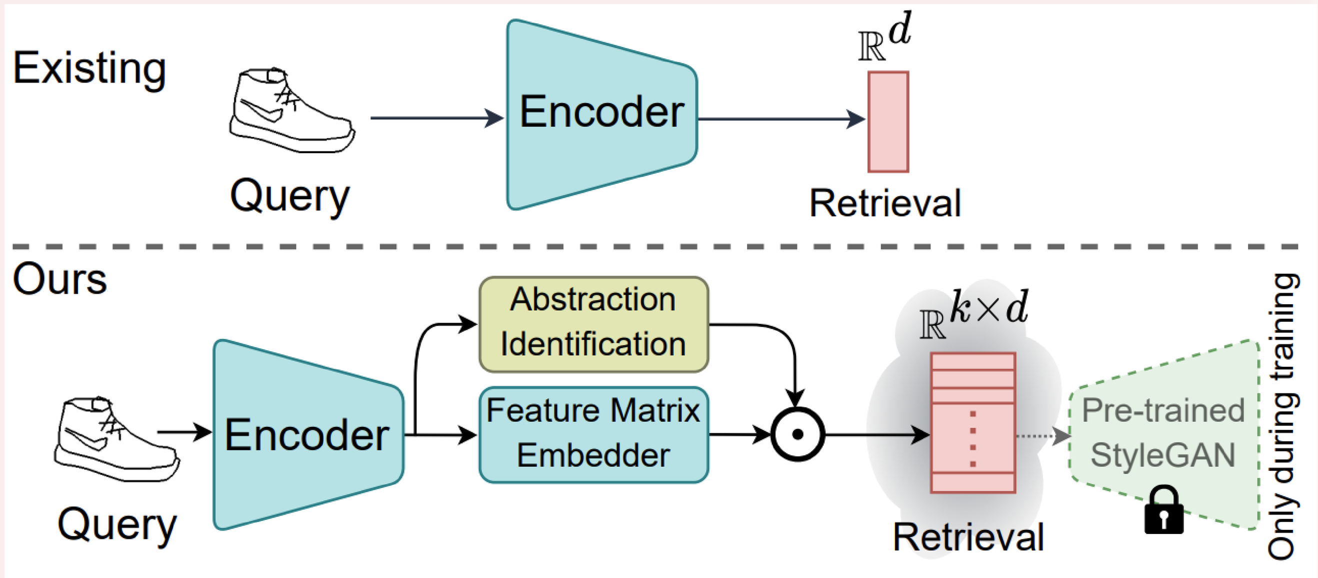
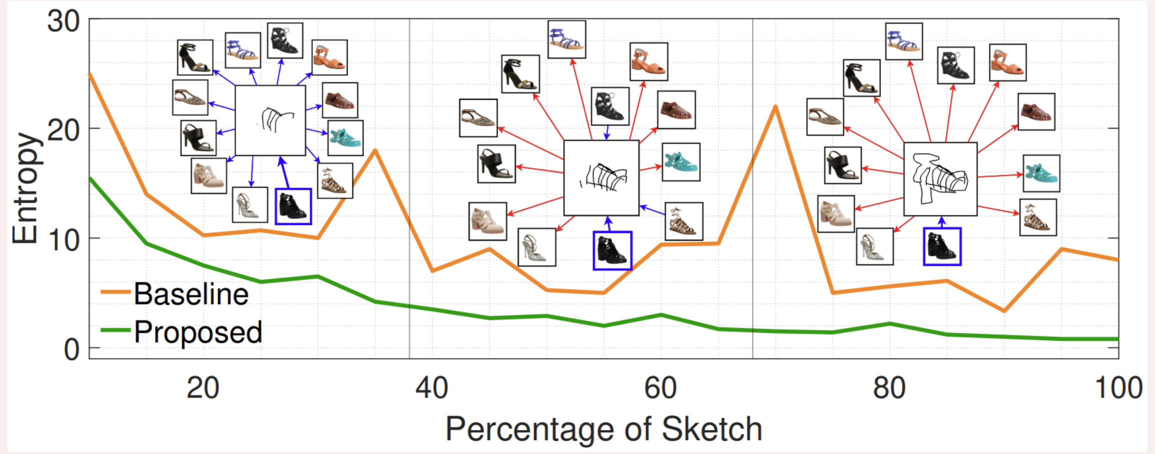
Got 2 Minutes? Start here ⬇

Summary



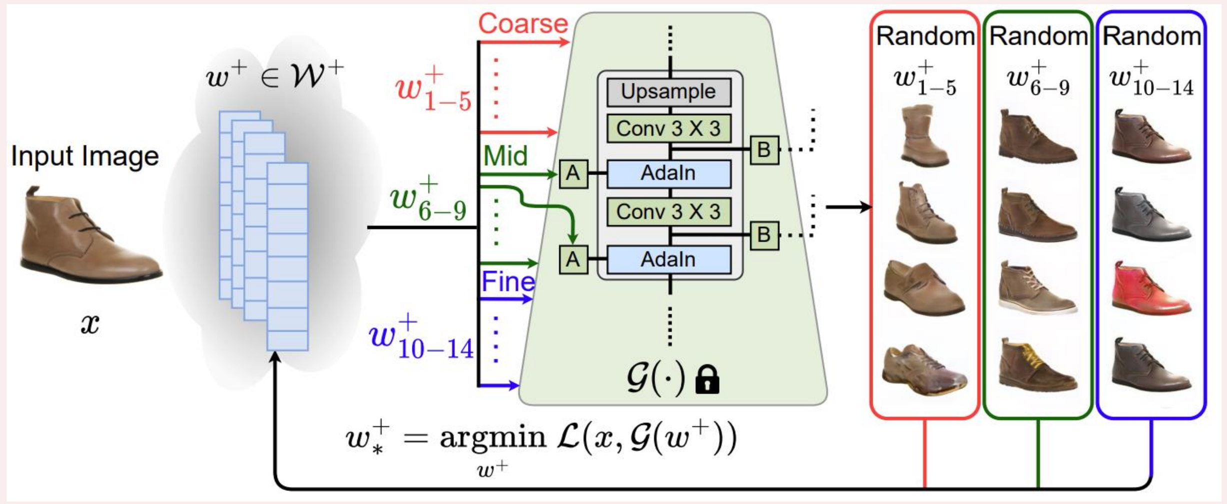
- Unlike existing feature vector embedding, we learn a **feature matrix representation**, regularised by a **pre-trained StyleGAN**.
- Tackles the **abstraction problem as a whole** for the task of FG-SBIR.
- Proposes a **differentiable `Acc.@q` surrogate loss** for retrieval.

Problems



- Freehand sketches exhibit **varied levels of abstraction**, formed by many interlacing factors such as drawing skill, style, culture & subjectivity.
- The gold-standard **Triplet loss** is solely concerned with bringing the best match to top-1, *regardless* of how good or bad the input sketch is.
- Existing methods uses a **fixed embedding matrix** to perform retrieval.

Solutions



- We tailor abstraction understanding in our system's DNA by specifically designing **feature-level and retrieval granularity-level innovations**.
- We utilise the **information-rich StyleGAN latent space** to guide learning of abstraction-aware feature embedding.
- An **abstraction-level selector** to dynamically assess levels of input abstraction, and select appropriate levels in our feature matrix embedding.
- Proposed **`Acc.@q` surrogate loss** allows the retrieval metric to accommodate different levels of sketch-abstraction.

How to Handle Sketch-Abstraction in Sketch-Based Image Retrieval?

Subhadeep Koley[1,2], Ayan Kumar Bhunia[1], Aneeshan Sain[1], Pinaki Nath Chowdhury[1], Tao Xiang[1,2], Yi-Zhe Song[1,2]

[1]*SketchX, CVSSP, University of Surrey*

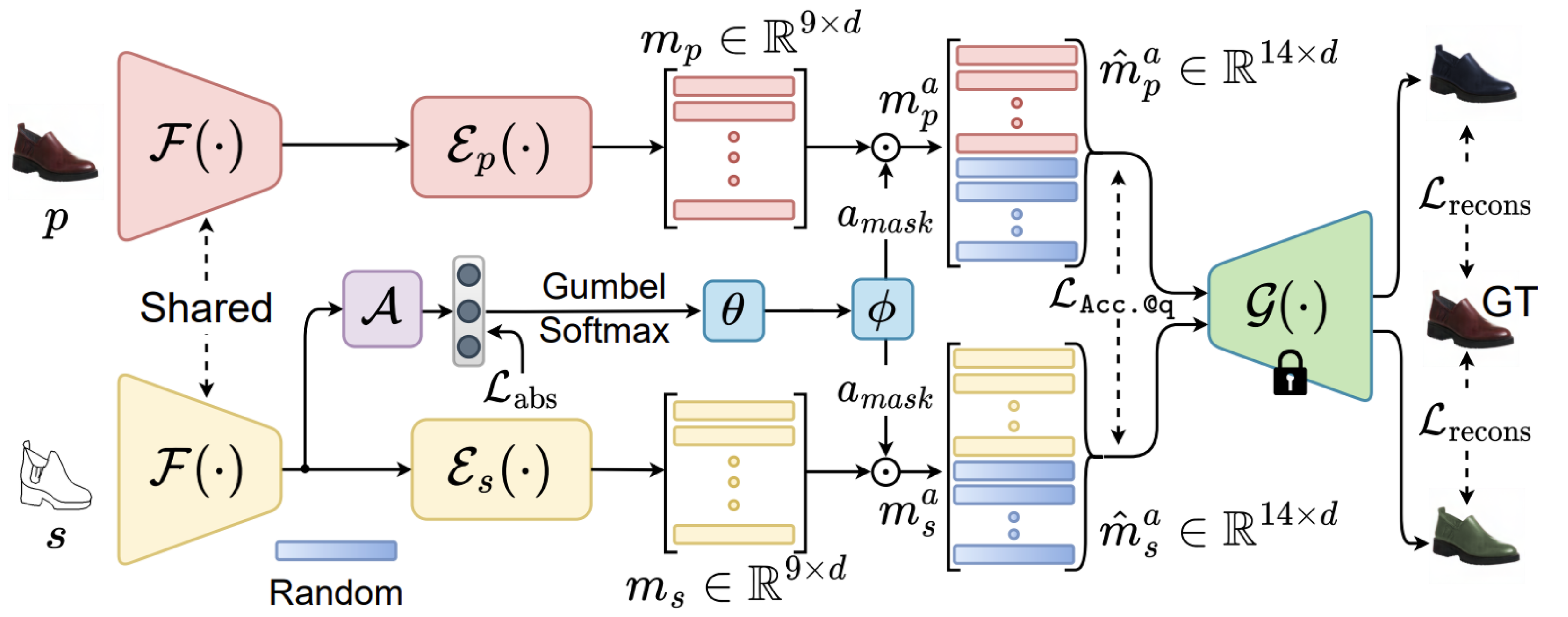
[2]*iFlyTek-Surrey Joint Research Centre on Artificial Intelligence*

SketchX · UNIVERSITY OF SURREY · CVSSP Centre for Vision, Speech and Signal Processing · CVPR JUNE 17-21, 2024 SEATTLE, WA

Proposed Model

- **Salient Components**
  1. Harness the rich semantic embedding of pre-trained StyleGAN model.
  2. Abstraction-level selector to dynamically assess input sketch-abstraction .
  3. Differentiable `Acc.@q` surrogate loss.

- To distil the knowledge from a frozen StyleGAN, we use, **sketch-to-photo** & **photo-to-photo reconstruction losses**, which align the sketch/photo feature matrix embedding to that of the disentangled latent space of StyleGAN.



- **Abstraction identification head** $\mathcal{A}(\cdot)$ takes global average pooled sketch feature from the backbone network as input, and predicts probability distribution over 3 latent groups (*i.e.*, **coarse, medium, or fine**).
- Optimising test-time evaluation metrics (*e.g.*, `Acc.@q`) directly via gradient descent is difficult in practice, as they **invoke non-differentiability**, owing to operations like sorting, counting, etc.
- We present a **surrogate differentiable approximation** of `Acc.@q` via a temperature-controlled Sigmoid function.

Algorithm 1 PyTorch-like code for differentiable `Acc.@q`

```
import torch
import torch.nn.functional as F
import F.pairwise_distance as PD

def accuracy@q_loss(anchor, pos, q):
    # anchor:B×d; pos:B×d; q:1×1(Acc.@q)
    # B←batch size; d←feature dimension
    acc_q = torch.tensor(0) # Accumulate accuracy
    dist = PD(anchor, pos) # B×B

    for i in range(len(anchor)):
        # Target distance for i^th anchor
        tgt_dist = PD(anchor[i], pos[i])# 1×1
        # Relative distance to target
        rel_dist = tgt_dist - PD(anchor[i], pos)# B×1

        signs = sigmoid(rel_dist, 0.01)
        rank_i = torch.sum(signs)# Get rank:1×1
        acc_q += sigmoid(q - rank_i, 1)

    batch_acc = acc_q / len(anchor)
    loss = -1 * batch_acc
    return loss

# Approximate step function with Sigmoid
def sigmoid(tensor, tau):
    tensor = tensor / tau
    y = 1.0 / (1.0 + torch.exp(-tensor))
    return y
```

Experiments & Results

Table 1. Results for standard FG-SBIR task.

Methods	ChairV2			ShoeV2		
	Acc.@1	Acc.@5	MOS $\mu \pm \sigma$	Acc.@1	Acc.@5	MOS $\mu \pm \sigma$
Triplet-SN [86]	47.4	71.4	2.6±0.4	28.7	63.5	2.2±0.2
HOLEF-SN [65]	50.7	73.6	2.7±0.6	31.2	66.6	2.5±0.1
B-pSp	47.8	70.2	2.9±0.2	30.1	63.9	2.8±0.9
B-pix2pix	44.2	66.9	2.8±0.7	26.7	60.2	2.3±0.1
B-CycleGAN	45.1	67.1	2.7±0.3	27.8	62.6	2.5±0.4
Partial-OT [18]	63.3	79.7	3.6±0.7	39.9	68.2	3.5±0.6
CrossHier [56]	62.4	79.1	3.5±0.3	36.2	67.8	3.3±0.1
StyleMeUp [57]	62.8	79.6	3.7±0.2	36.4	68.1	3.3±0.3
On-the-fly [8]	51.2	73.8	3.2±0.5	30.8	65.1	2.8±0.5
SketchPVT [60]	71.2	80.1	3.3±0.7	44.1	70.8	2.9±0.9
Proposed	**72.1**	**80.9**	**4.4±0.5**	**45.3**	**77.3**	**4.1±0.5**