

ViSketch-GPT: Collaborative Multi-Scale Feature Extraction for Hand-Drawn Sketch Retrieval

Giulio Federico^{1,2,*}, Fabio Carrara², Claudio Gennaro² and Marco Di Benedetto²

¹University of Pisa, Pisa (PI), Italy

²Institute of Information Science and Technologies (ISTI-CNR), Pisa (PI), Italy
{name.surname}@isti.cnr.it

The problem we want to solve

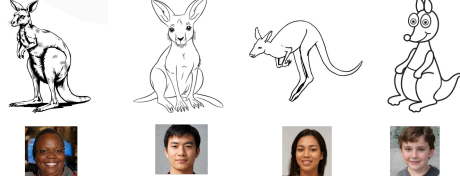
Natural/Realistic Images of a Kangaroo



1

Natural images of a category (e.g., kangaroo) look alike—regardless of the device. A kangaroo is always a kangaroo.

Hand-Drawn Sketches of a Kangaroo



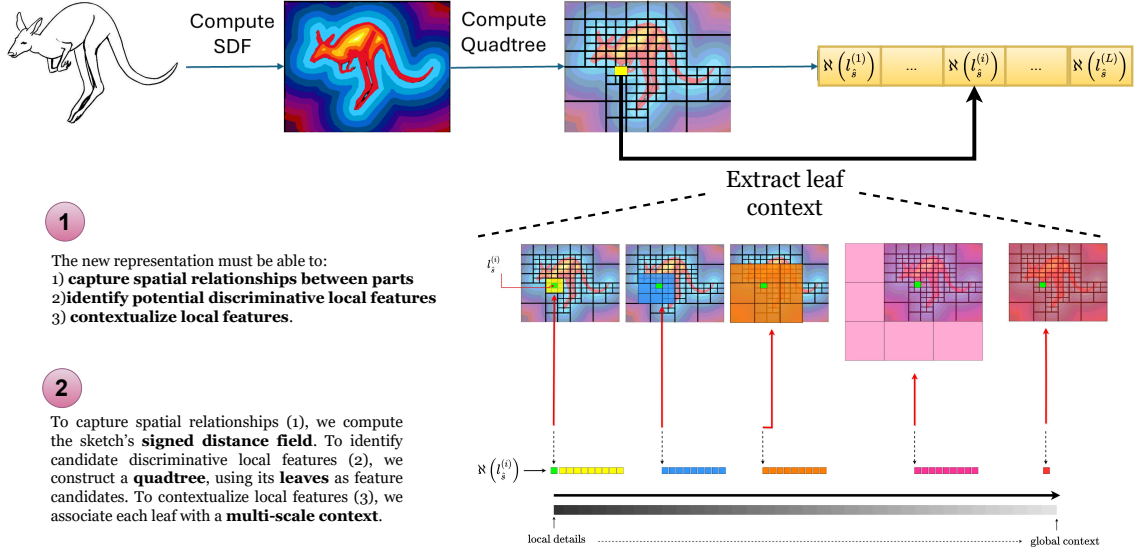
2

Each sketch reflects its author's background, perception, and skill, making the same object appear in highly diverse and subjective ways.

3

In recognition or retrieval tasks, embeddings from models trained on realistic images are highly informative. For sketches, however, **no representation yet captures their essence or handles the wide variability in style and conceptual interpretation.**

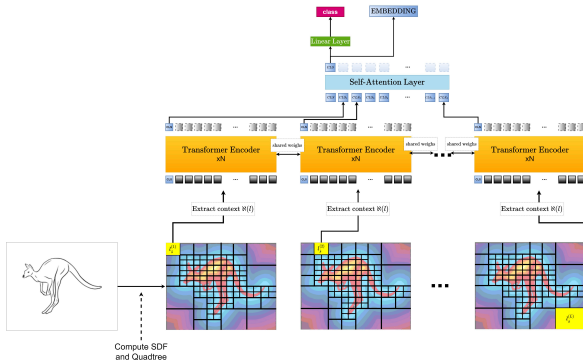
A novel representation for sketches



Sketch Retrieval

1

We trained our new architecture to classify sketches, then tested whether the **final extracted embedding** could be used for retrieval.



2

We conducted experiments on the *QuickDraw* and *TU-Berlin* datasets. Results show that **embeddings extracted from our architecture are significantly more meaningful** for retrieval than those from competing models.

Methods	QuickDraw			TU-Berlin		
	Top-1	Top-5	mAP	Top-1	Top-5	mAP
Bi-LSTM	70.91	89.52	60.11	31.40	59.60	23.71
Sketch-a-Net	74.88	90.10	65.13	37.25	63.50	26.18
DSSA	78.16	91.04	68.10	38.45	66.10	28.77
ResNet18	80.34	91.71	70.98	41.45	67.10	29.33
ResNet50	82.41	92.52	74.84	51.80	74.45	36.94
TCNet	83.59	92.57	76.38	55.30	79.45	38.78
Sketch-BERT	85.47	93.49	78.87	57.25	81.50	41.54
ViSketch-GPT (our)	96.69	99.92	87.35	67.36	89.40	59.99