Sketch-Based Image Retrieval with Style

CSIE5130 Multimedia Analysis & Indexing Final Project

Bo-Ru Lu Chia-Hsuan Li Wei Fang

January 9, 2018

NTU GICE·GINM·EE

Task Definition

Cross domain retrieval using sketch images

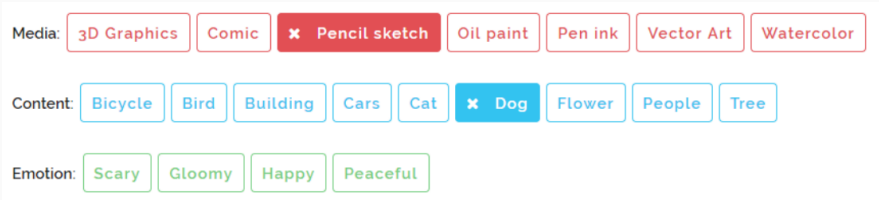
- Styles hold substantial differences
  (Even good classifiers cannot generalize)
- Beyond real images (photography)

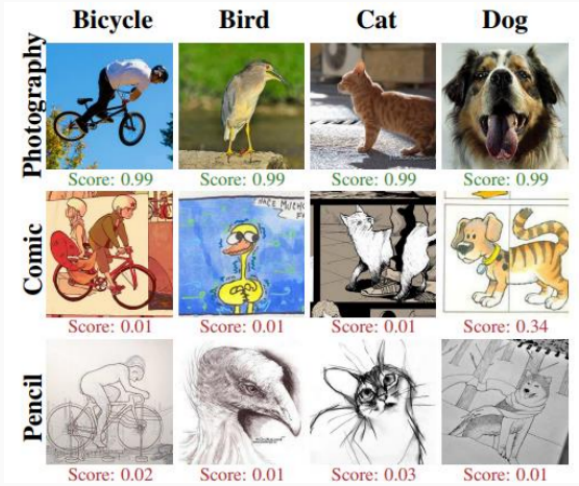
+ Comic $\longrightarrow$

Dataset

BAM! The Behance Artistic Media Dataset

What we use:

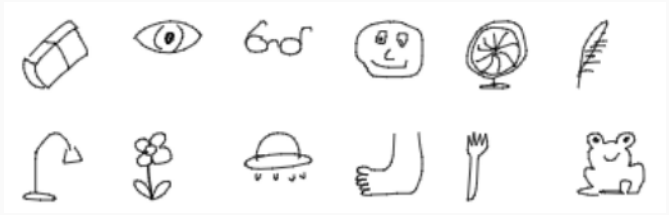




- All 7 styles (media)
- 5 content classes: Bicycle, Bird, Cars, Cat, Dog

The Quick, Draw! Dataset

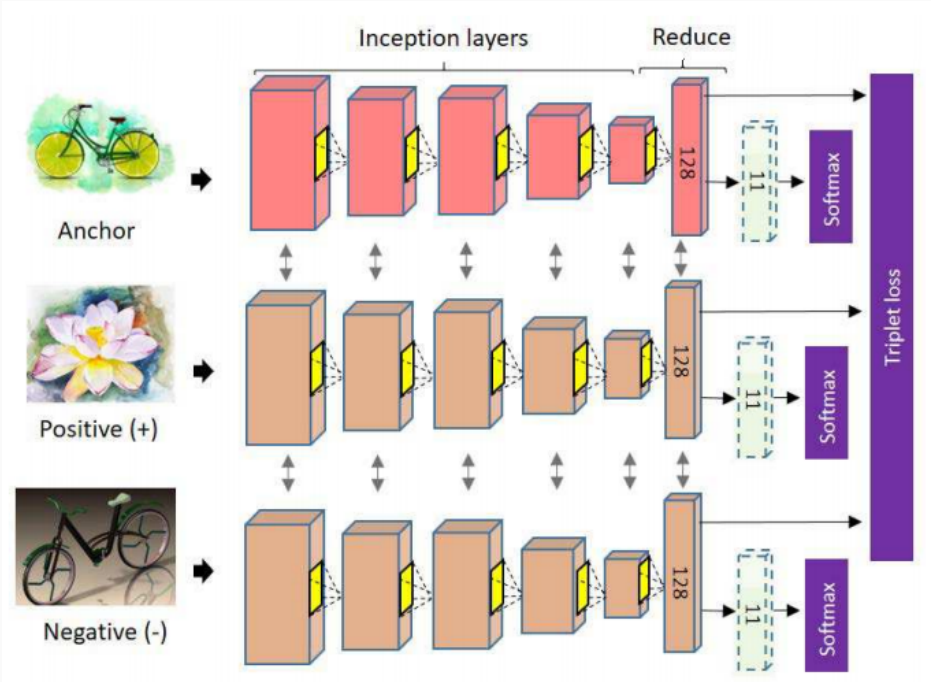
Used as sketch queries



Model

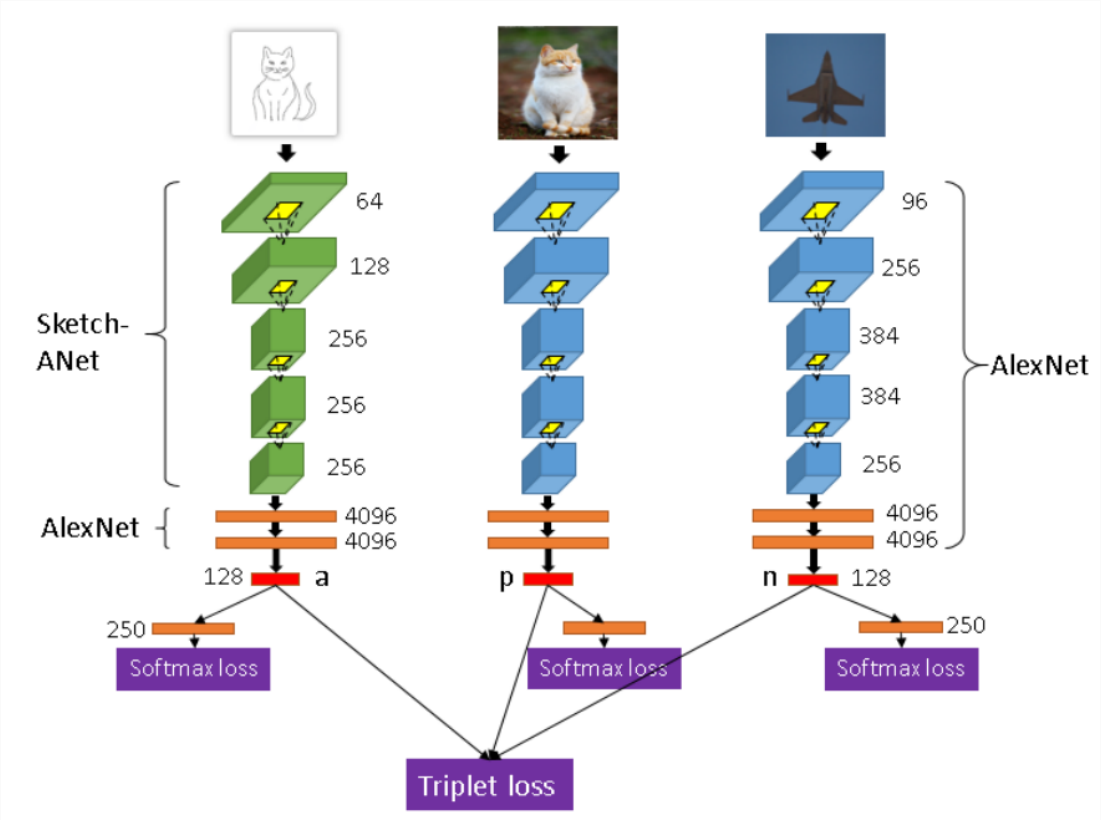
Style Network

- Siamese Network
  - Three branches share weights
- 2-stage training
  - Classification Loss
  - Triplet Loss: decorrelation between semantics and style



Model

Structure Network



- The sketch data we use (Quick, Draw!) is much more abstract and noisy (even after human selection)
- Structure Network: during training we use BAM! images instead of real photos

Model

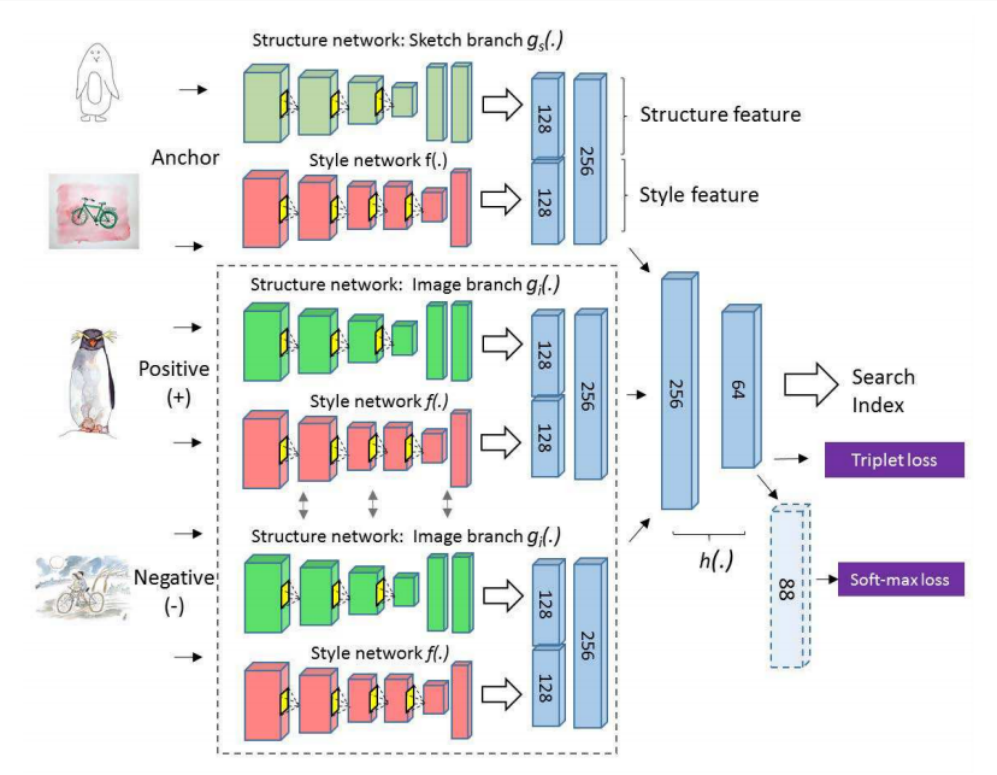
Hierarchical Triplet ConvNet

Modules:

- Style Network
- Structure Network

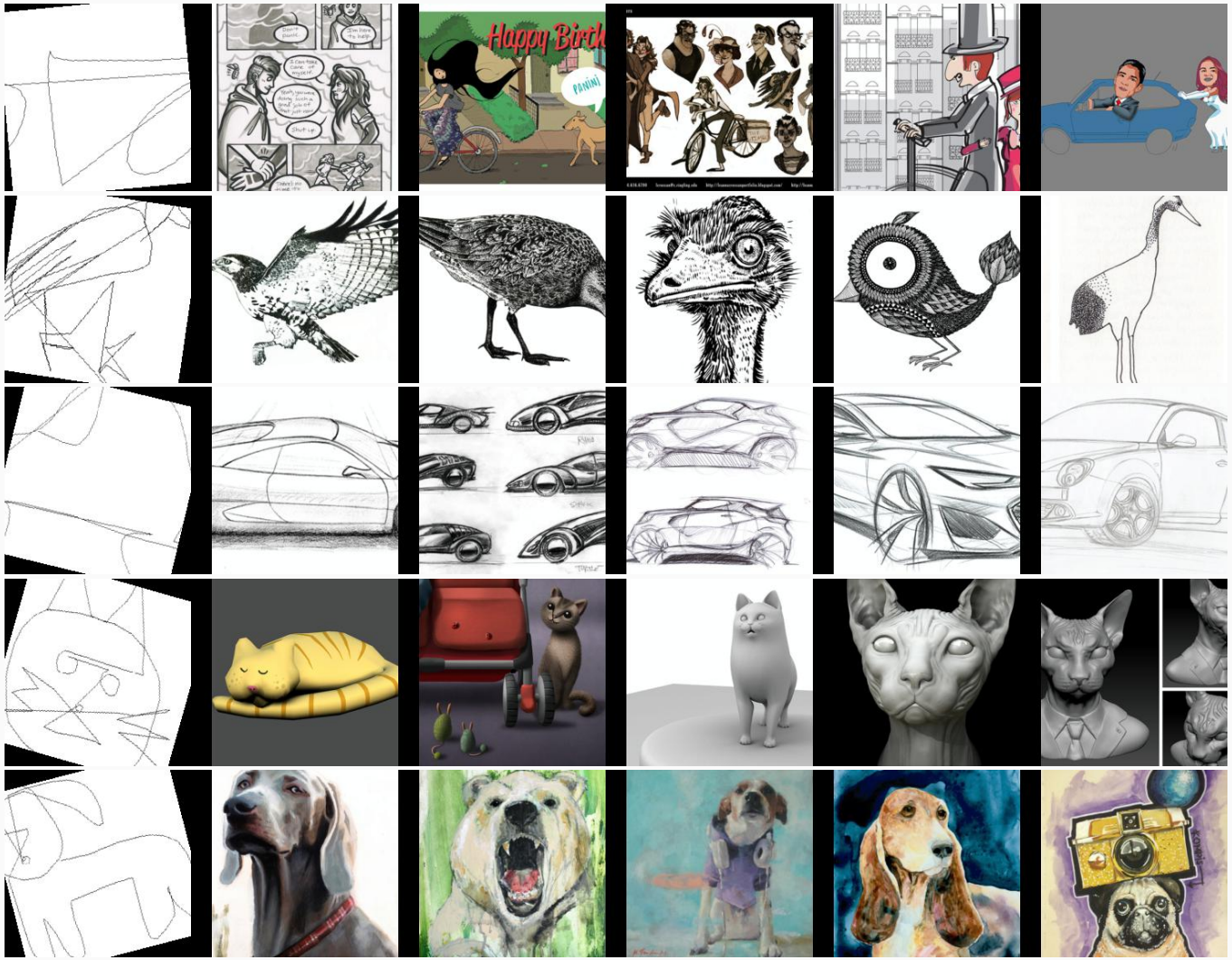
Training

- Classification Loss
  - 5×7 classes
- Triplet Margin Loss
  - Anchor $\longleftrightarrow$ Pos & Neg
  - $\mathcal{L}(a, p, n) = [m + |f(a) - f(p)|^2 - |f(a) - f(p)|^2]$



Sketching with Style: Visual Search with Sketches and Aesthetic Context. ICCV 2017.

Experimental Results



Experimental Results

Sketch mAP

Oilpaint	3D	Pen Ink	Comic	Vector	Graphite	Watercolor	*Overall*
0.934	**0.962**	0.810	0.898	0.906	**0.994**	**0.960**	0.915

Experimental Results

Sketch + Style mAP

	Oilpaint	3D	Pen Ink	Comic	Vector	Graphite	Watercolor	*Overall*
Dog	0.525	0.379	0.384	0.389	0.446	0.520	0.458	0.443
Cat	0.400	0.637	0.668	0.615	0.622	0.755	0.678	0.625
Bird	0.662	0.647	0.594	0.570	0.621	0.678	0.676	0.635
Bicycle	0.414	0.866	0.585	0.552	0.706	0.612	0.652	0.627
Cars	0.687	0.901	0.586	0.632	0.844	0.863	0.771	**0.755**
Overall	0.541	**0.692**	0.559	0.550	0.652	**0.685**	0.646	

Demo