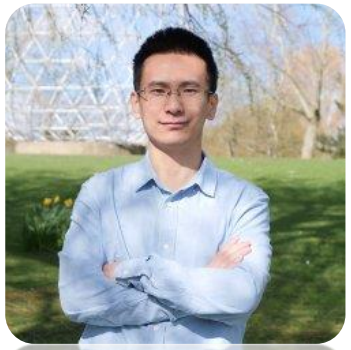
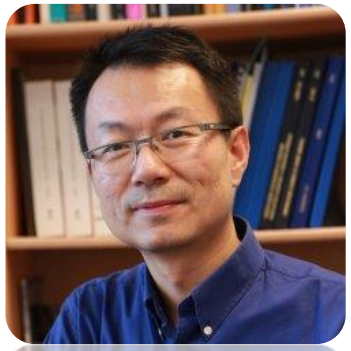
StyleMeUp: Towards Style-agnostic Sketch Based Image Retrieval

Aneeshan Sain [1,2]

Ayan Kumar Bhunia [1]



Yongxin Yang [1,2]



Tao Xiang [1,2]

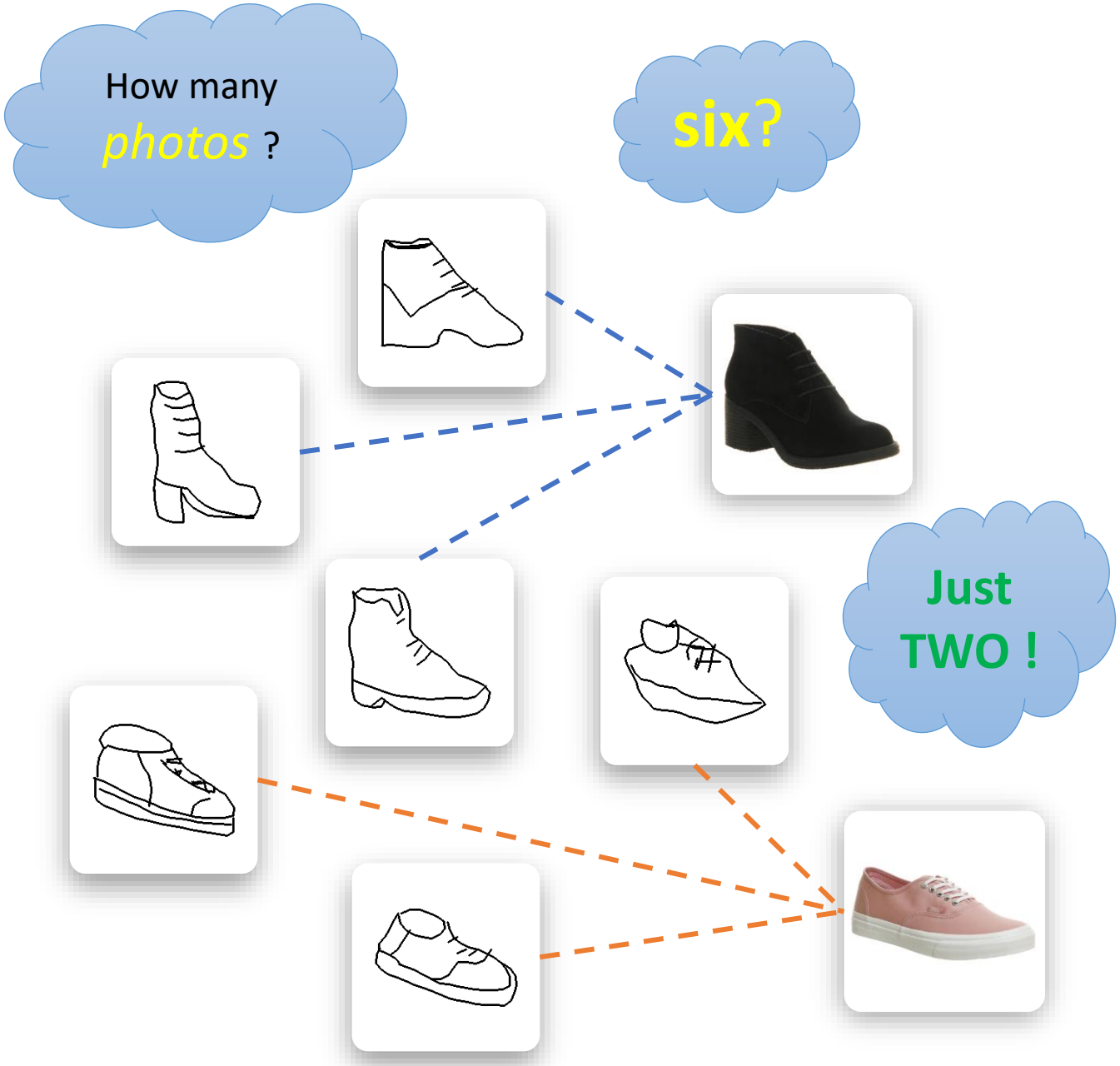
Yi-Zhe Song [1,2]

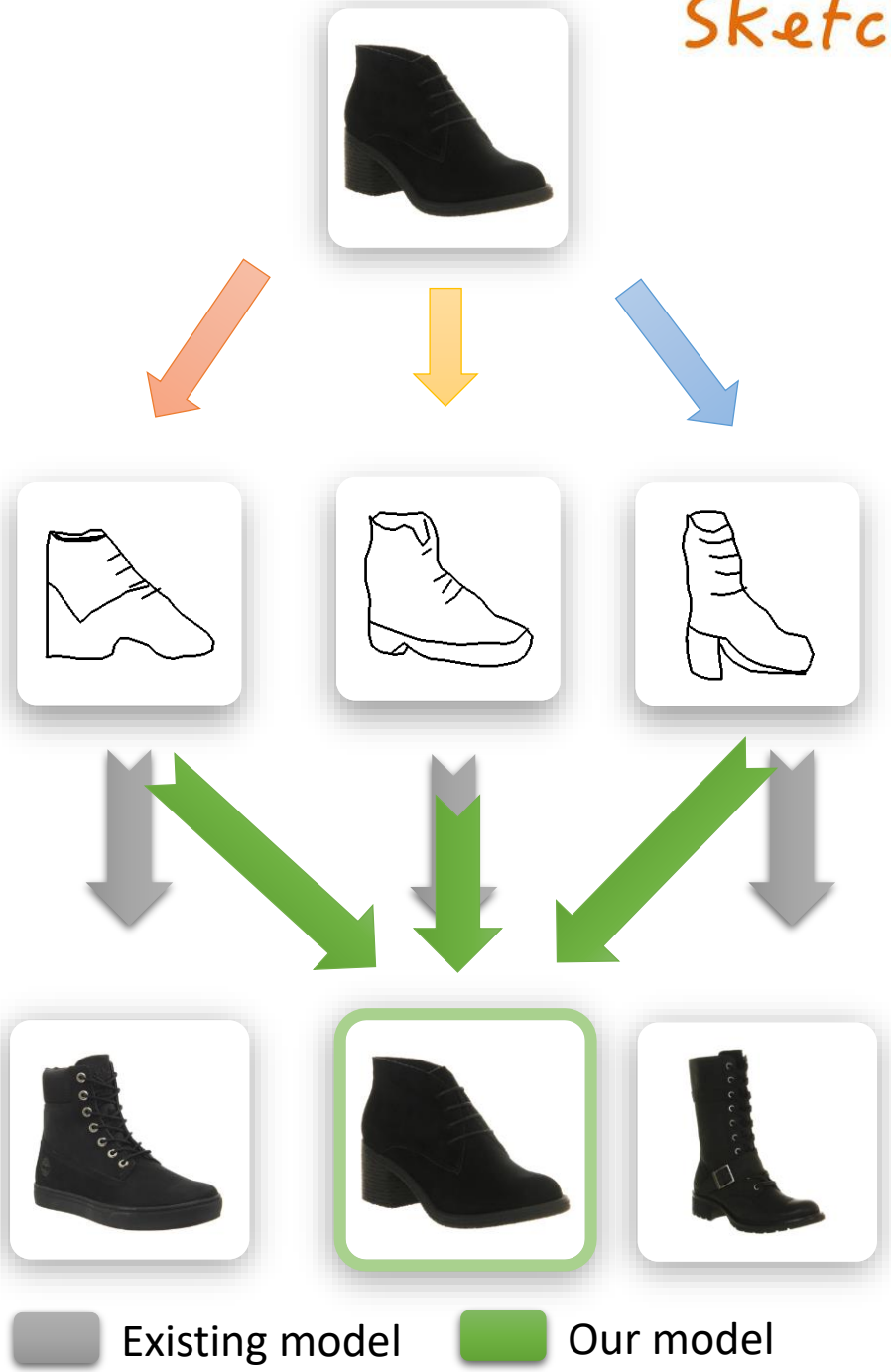
[1] SketchX Lab, CVSSP, University of Surrey, UK

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

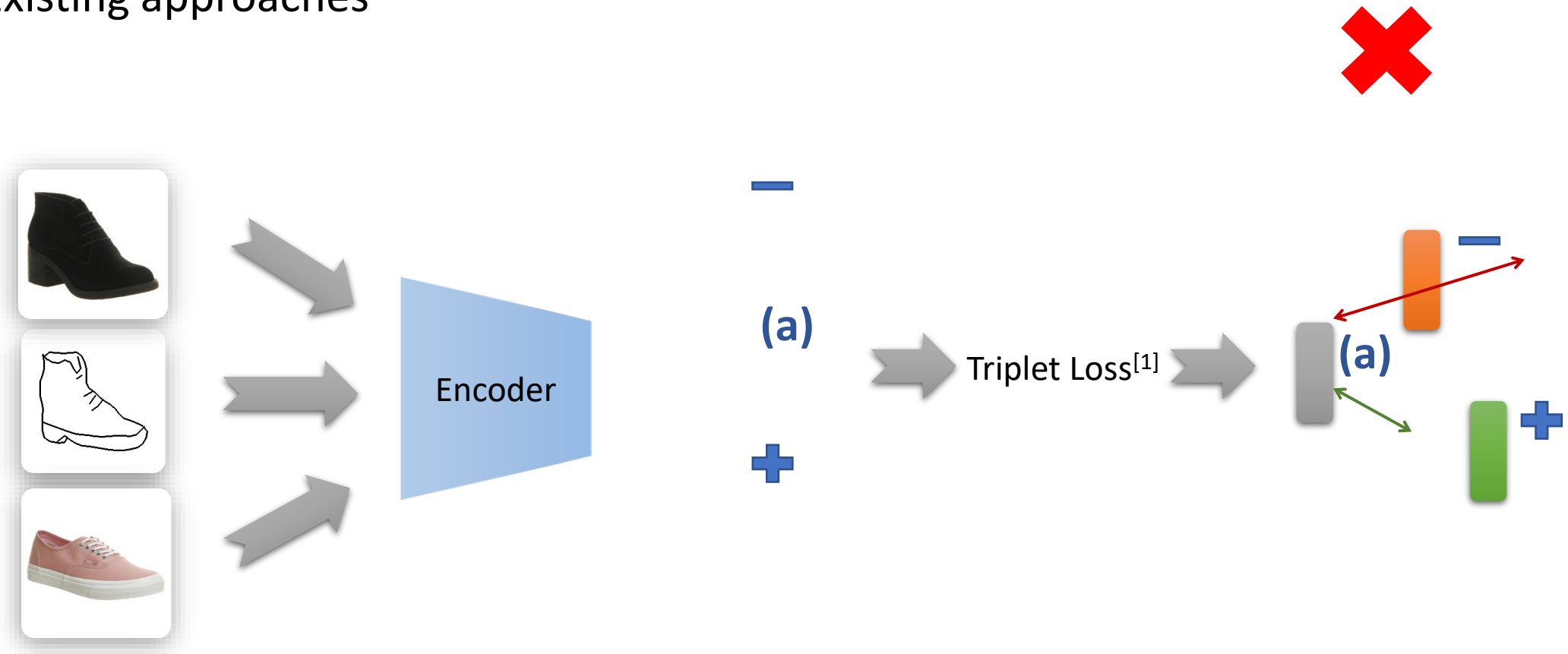
http://sketchx.ai

Style Diversity in Sketches



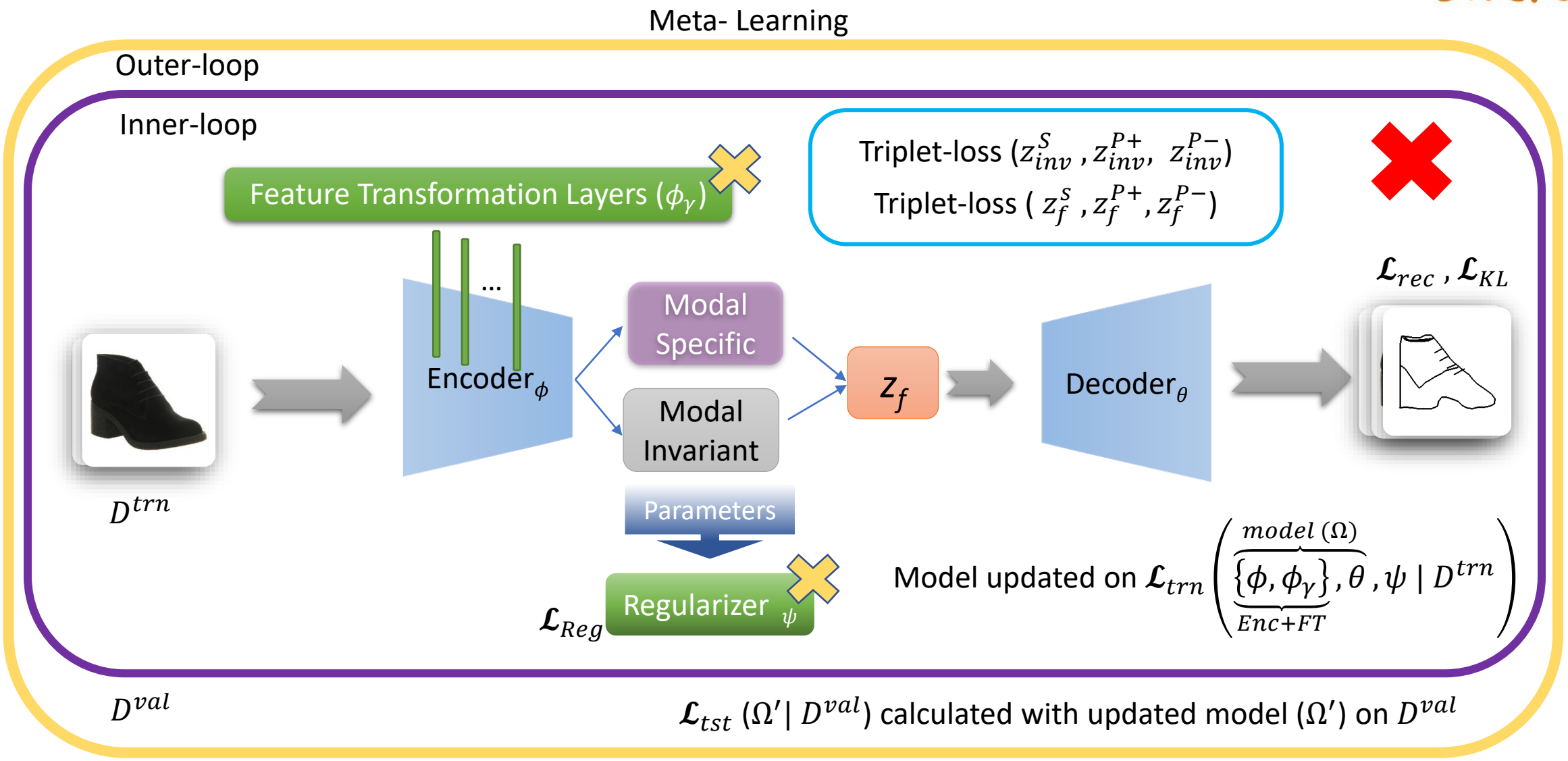


Existing approaches



- ✖ No cross-modal interaction
- ✖ Not useful for disentanglement
- ✖ Therefore, no option of modelling style diversity.

[1] Qian Yu, Feng Liu, Yi-Zhe Song, Tao Xiang, Timothy M Hospedales, and Chen-Change Loy. Sketch me that shoe. In CVPR, 2016.



Our Framework

Quantitative Analysis

Datasets:

Fine-Grained SBIR: QMUL ShoeV2 and ChairV2 [1]

Table 1. Comparative results of our model against other methods on FG-SBIR (D → disentanglement baselines).

	Methods	Chair-V2		Shoe-V2	
		acc.@1	acc.@10	acc.@1	acc.@10
SOTA	Triplet-SN[1]	47.65	84.24	28.71	71.56
SOTA	Triplet-Attn [2]	53.41	87.56	31.74	75.78
SOTA	Triplet-RL[3]	56.54	89.61	34.10	78.82
SOTA	CC-Gen [4]	54.21	88.23	33.80	77.86
D	D-TVAE [5]	49.37	81.63	27.62	70.32
D	D-DVML[6]	52.78	85.24	32.07	76.23
Others	B-Basic-SN	49.58	85.41	29.45	72.83
Others	B-SN-Group	50.35	88.28	30.14	75.62
Others	B-Cross-Modal [7]	52.24	86.58	31.18	73.51
Others	B-Meta-SN	53.57	87.69	32.74	76.92
	Proposed	**62.86**	**91.14**	**36.47**	**81.83**

Categorical SBIR : Sketchy (ext.)[8], TUBerlin (ext.)[8]

Table 2. Comparative results of our model against other methods on SBIR (D → disentanglement baselines).

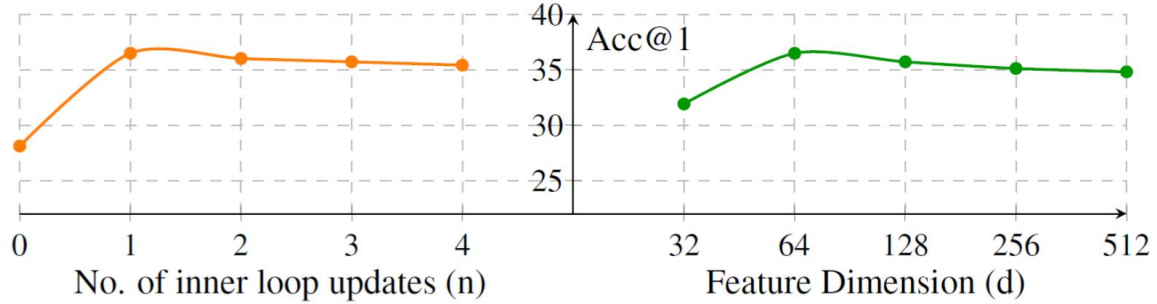
	Methods	Sketchy (ext)		TU Berlin (ext)	
		mAP	P@200	mAP	P@200
SOTA	DSH (64 bit) [8]	0.711	0.858	0.521	0.655
SOTA	GDH (64 bit) [9]	0.810	0.894	0.690	0.728
D	D-TVAE [5]	0.695	0.839	0.507	0.643
D	D-DVML [6]	0.785	0.891	0.648	0.693
Others	B-Basic-SN	0.715	0.861	0.531	0.659
Others	B-SN-Group	0.738	0.872	0.572	0.661
Others	B-Cross-Modal [7]	0.763	0.884	0.622	0.688
Others	B-Meta-SN	0.824	0.897	0.674	0.715
	Proposed	**0.905**	**0.927**	**0.778**	**0.795**

[1] Qian Yu, Feng Liu, Yi-Zhe Song, Tao Xiang, Timothy M Hospedales, and Chen-Change Loy. Sketch me that shoe. In CVPR, 2016.
[2] Jifei Song, Qian Yu, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Deep spatial-semantic attention for fine-grained sketch-based image retrieval. In ICCV, 2017
[3] Ayan Kumar Bhunia, Yongxin Yang, Timothy M Hospedales, Tao Xiang, and Yi-Zhe Song. Sketch less for more: On-the-fly fine-grained sketch-based image retrieval. In CVPR, 2020.
[4] Kaiyue Pang, Ke Li, Yongxin Yang, Honggang Zhang, Timothy M Hospedales, Tao Xiang, and Yi-Zhe Song. Generalizing fine-grained sketch-based image retrieval. In CVPR, 2019.
[5] Haque Ishfaq, Assaf Hoogi, and Daniel Rubin. Tvae: Triplet based variational autoencoder using metric learning. 2018
[6] Xudong Lin, Yueqi Duan, Qiyuan Dong, Jiwen Lu, and Jie Zhou. Deep variational metric learning. In ECCV, 2018
[7] Adrian Spurr, Jie Song, Seonwook Park, and Otmar Hilliges. Cross-modal deep variational hand pose estimation. In CVPR, 2018.
[8] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In ICCV, 2017
[9] Jingyi Zhang, Fumin Shen, Li Liu, Fan Zhu, Mengyang Yu, Ling Shao, Heng Tao Shen, and Luc Van Gool. Generative domain-migration hashing for sketch-to-image retrieval. In ECCV, 2018.

Ablation Studies

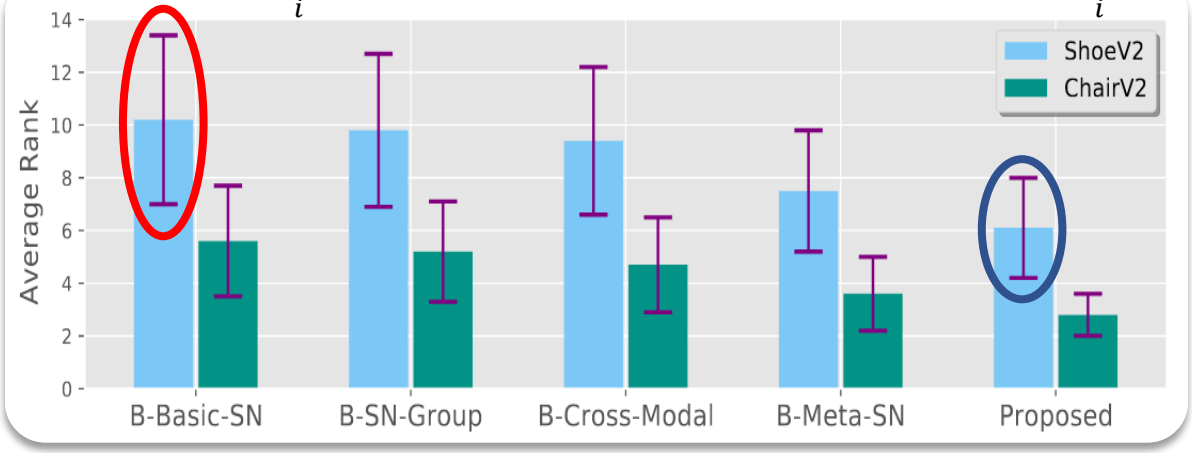
Ablative Study

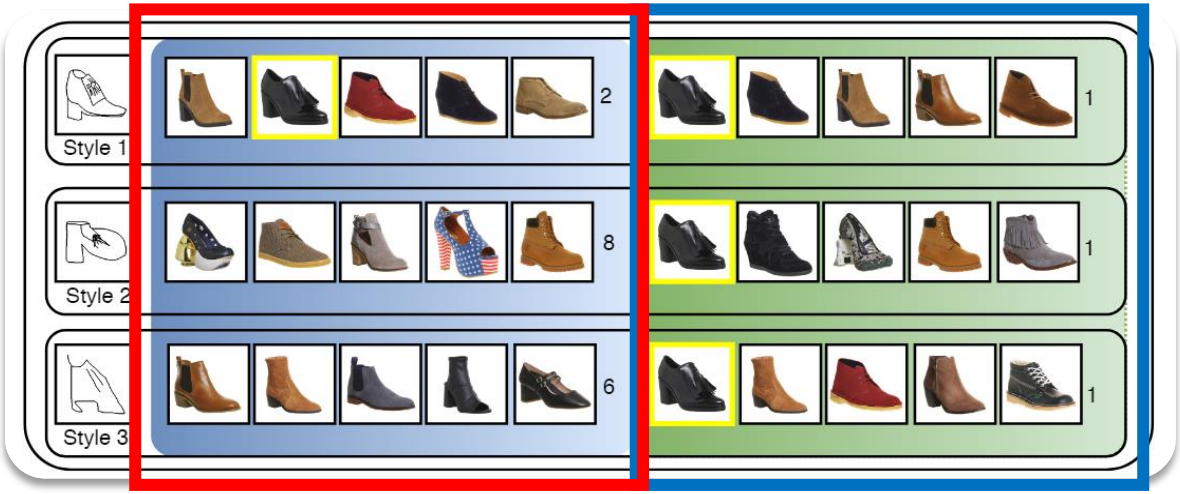
Methods	Shoe-V2		Sketchy (ext)	
	acc.@1	acc.@10	mAP	P@200
w/o Diversity	27.12	69.01	—	—
w/o MFT	33.28	75.34	0.852	0.916
w/o RegD	32.57	73.84	0.837	0.891
Fixed-FT	34.18	79.06	0.878	0.912
Proposed	**36.47**	**81.83**	**0.905**	**0.927**



Varying (a) No. of Inner loop updates (optimal at n=1) (b) feature dimension (optimal at d=64) on QMUL Shoe-V2.

$$rank_{avg} = \frac{1}{mk}\sum_{i} rank_i \; ; V_i = variance\left(rank_j^i\right)\Big|_{j=1}^{k} \; ; \; V_{avg} = \frac{1}{m}\sum_{i} V_i$$



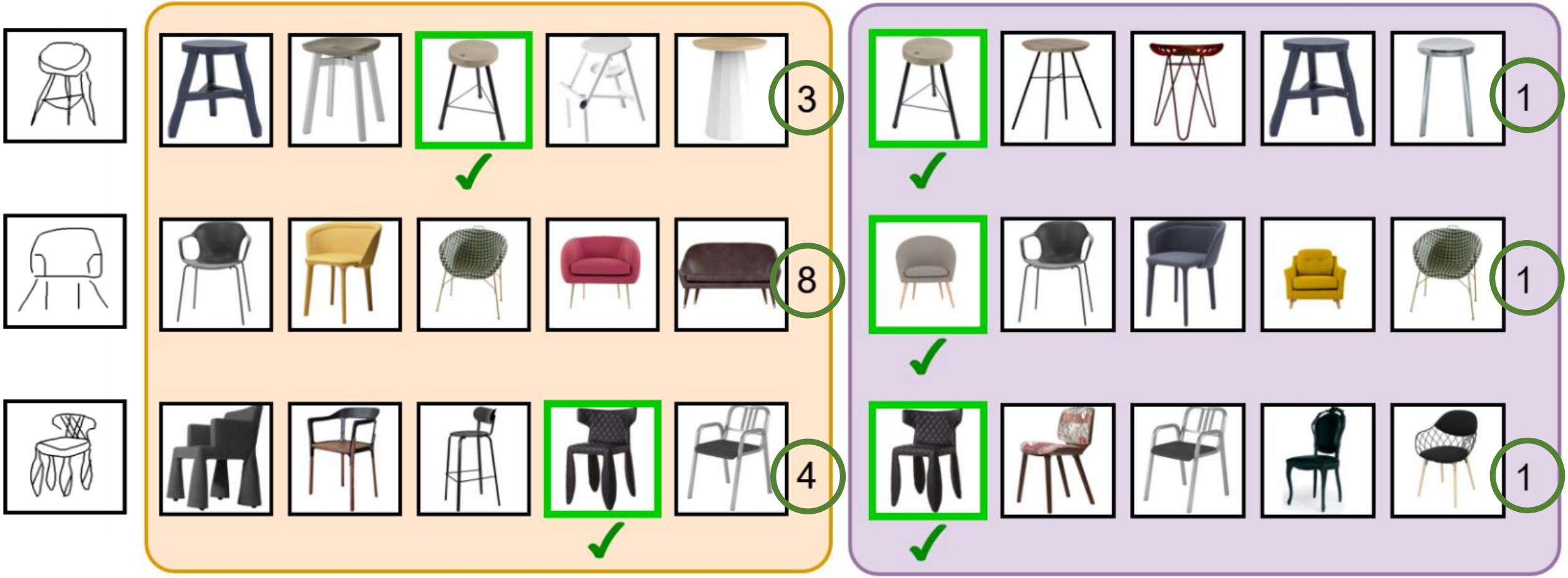


High consistency in retrieval accuracy !

Qualitative result of style-agnostic retrieval

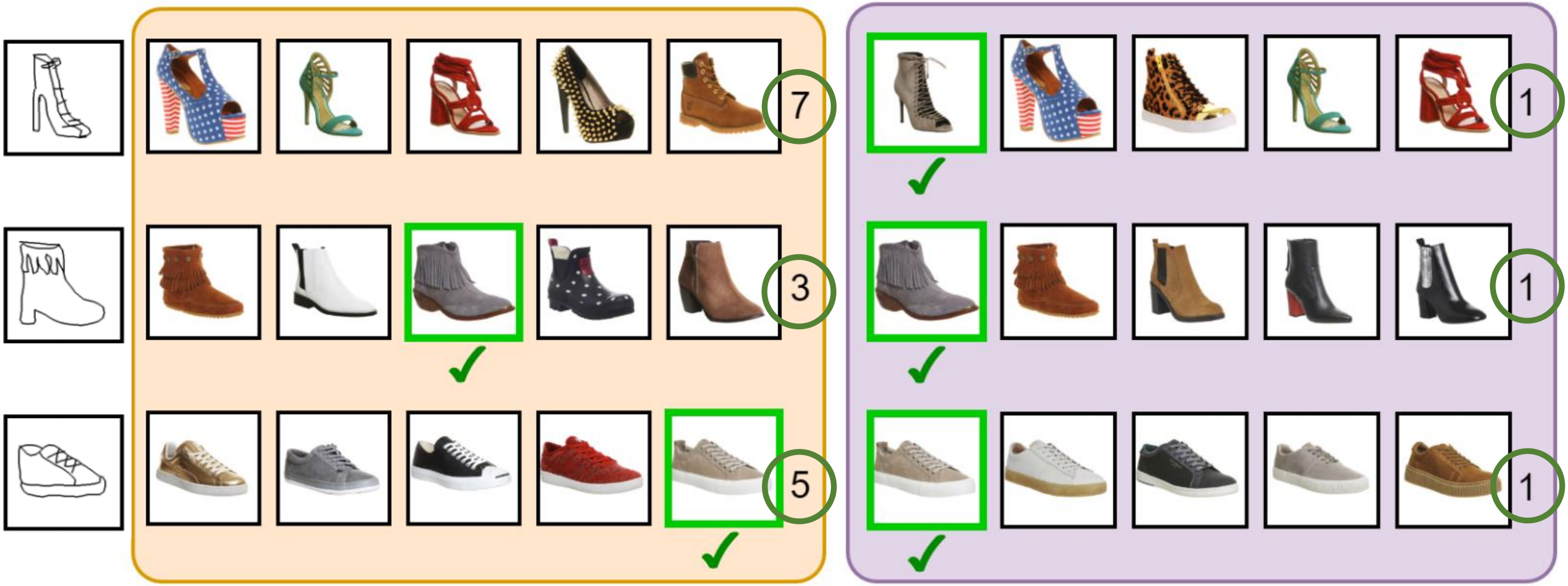
General Qualitative Results on Sketchy dataset

General Qualitative Results on QMUL ChairV2 dataset



General Qualitative Results on QMUL ShoeV2 dataset

SketchX

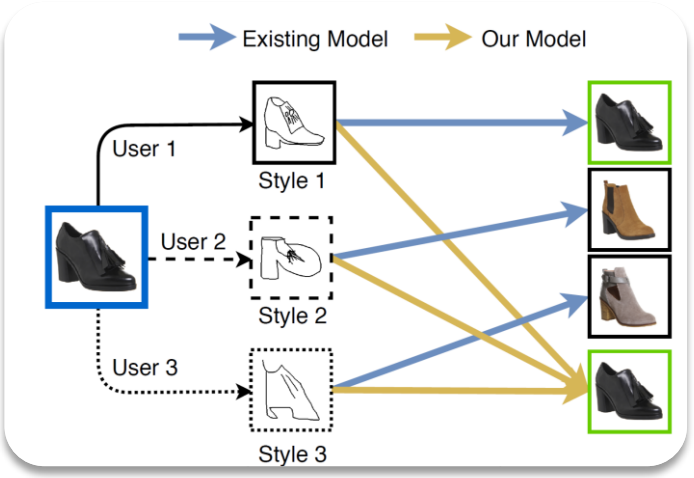


Baseline

Proposed

SketchX

http://sketchx.ai



https://aneeshan95.github.io