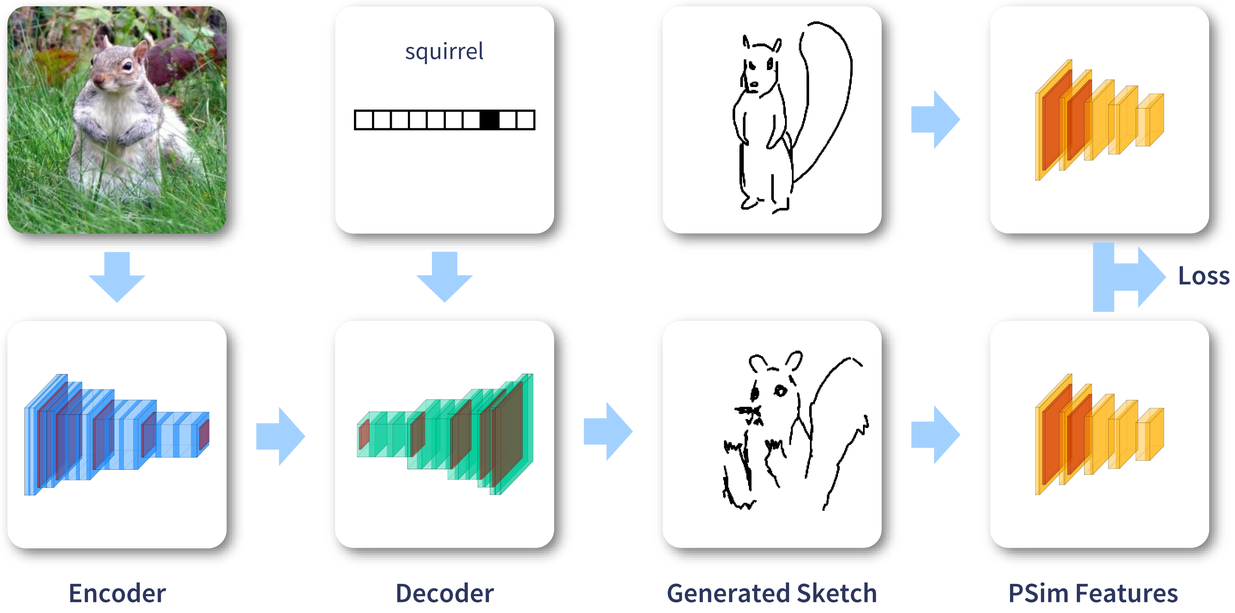
Synthesizing human-like sketches from natural images using a conditional convolutional decoder

Moritz Kampelmühler, Axel Pinz

Graz University of Technology

Correspondence: kampelmuehler@tugraz.at
Project Page: kampelmuehler.github.io/sketches



An encoder-decoder architecture transforms a natural input image into a human-like sketch. Skip connections from encoder to decoder activations foster multi-scale abstraction. Adaptive Instance Normalization (AdaIN) with learned class embeddings replaces Batch Normalization in the decoder to condition on a certain label. Another CNN (PSim) extracts features from generated and ground truth sketches. The loss for the decoder is the cosine similarity of the features of each convolutional layer in that network. Only the decoder and AdaIN embedding parameters are updated during training.

Key Insights

- ImageNet trained CNNs acquire implicit shape representations.
- Adaptive Instance Normalization can be efficiently used as a side input.
- Class related shape priors emerge from label injection.
- NN based perceptual similarity metrics enable domain translation across large gaps.

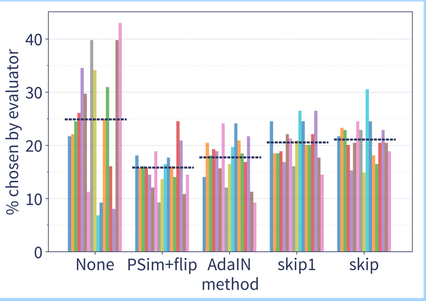
Quantitative Results

method	top-1	top-5	#params
chance	0.8%	3.9%	-
HED	0.4%	3.2%	14.7M
MSE	1.3%	4.9%	17.0M
PSim	37.9%	60.2%	17.0M
+flip	47.1%	69.2%	17.0M
+AdaIN	61.4%	79.9%	18.2M
skip1	62.3%	80.7%	19.2M
skip	**66.6%**	**85.6%**	21.3M
ground truth	91%	98%	-

Classification results of generated sketches in ablation study. Baselines and perceptual similarity loss and added components.

User Study

Example problem set from the user study. RGB image alongside sketches generated by the different methods.



Results of the user study. Hit percentage for each evaluator and method. Blue horizontal lines indicate the means.