Thinking Outside the Pool: Active Training Image Creation for Relative Attributes

Aron Yu[†] Kristen Grauman[†‡]

[†]University of Texas at Austin [‡]Facebook AI Research

Fine-Grained Visual Comparisons

More? Less? Equal?

- compares images exhibiting *subtle visual differences* w.r.t. a target attribute
- trains attribute-specific ranking models using pairwise labels

Curation Limit

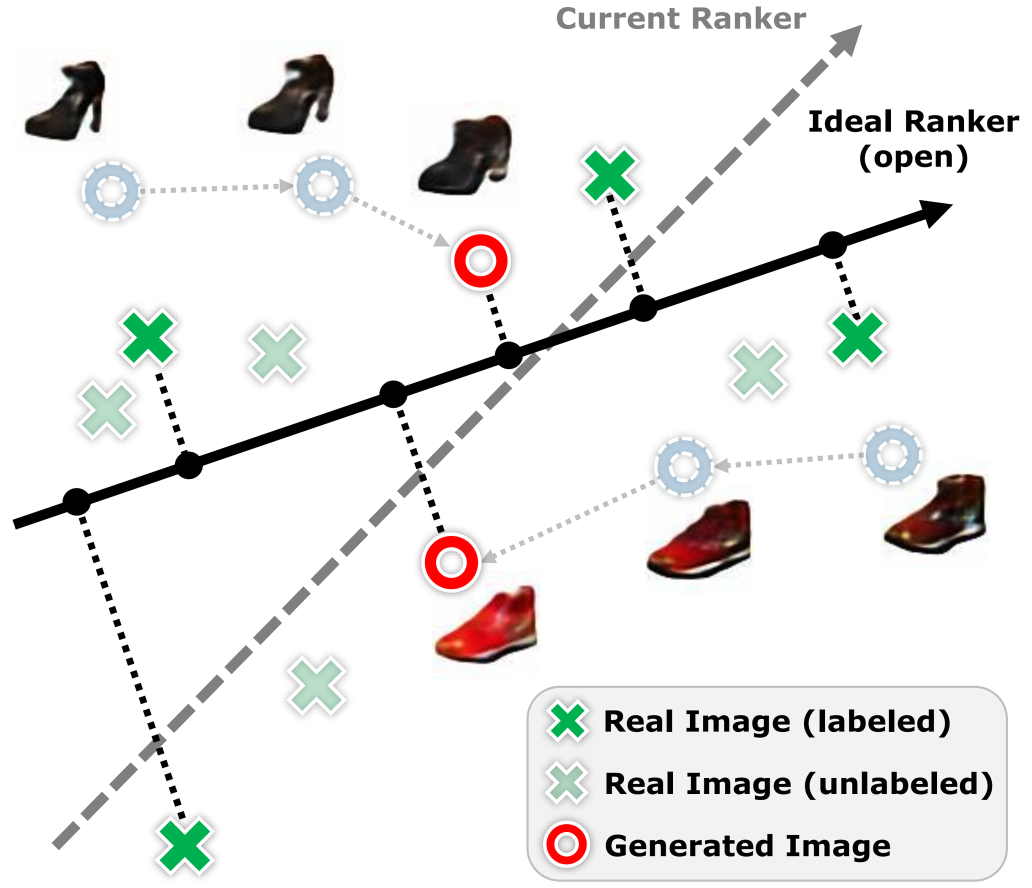
- visual variety and information from web content reach an upper limit
- subtle differences hard to directly curate in large numbers

Pool-Based Active Learning

- suffers from the *"streetlight effect"*
- starts with an existing pool of images [Freytag et al. '14, Vijayanarasimhan et al. '14, ...]

Source: Wikipedia

Think "Outside the Pool"

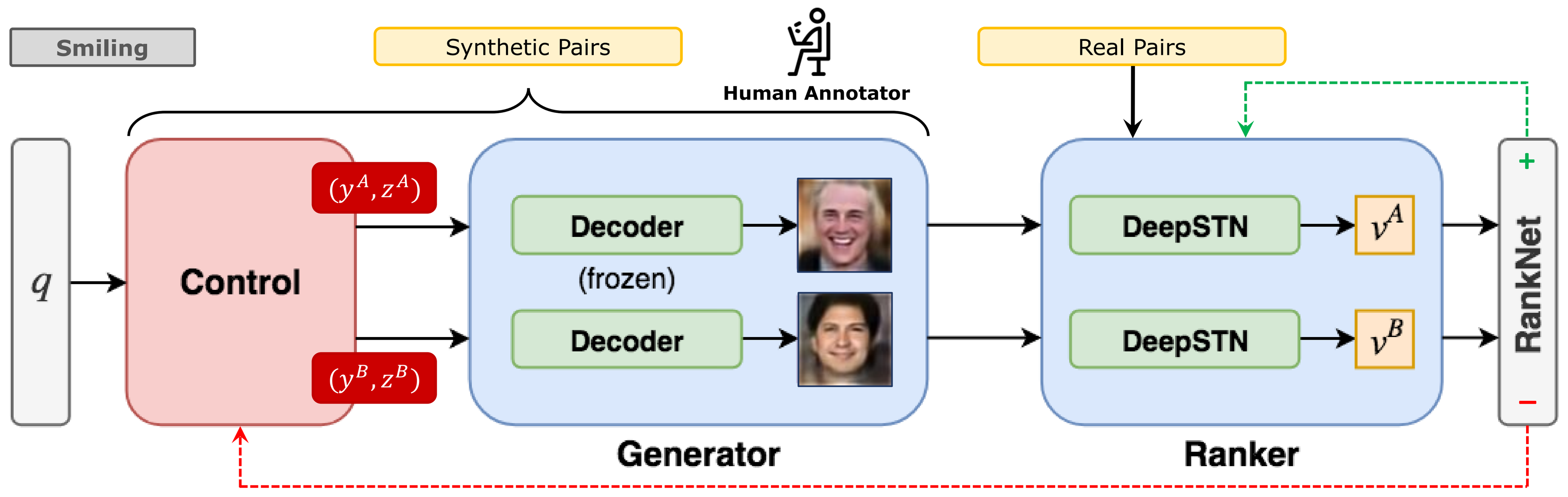


Our Idea

- generate the most ideal training image pairs *directly*
- adversarial learning that allows the model to *actively teach itself*
- have human annotators label the generated image pairs
- iteratively add to the pool of existing training image pairs

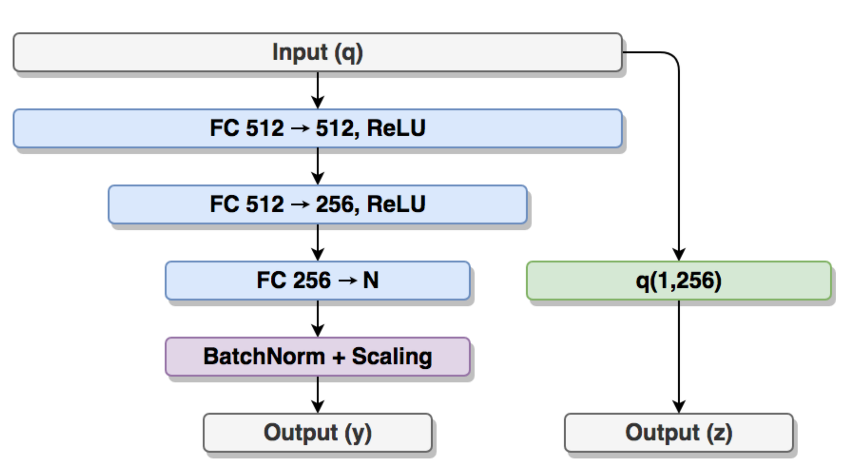
Key Idea: Joint learning of the visual task and training image generation

AcTive Training Image Creation (ATTIC)



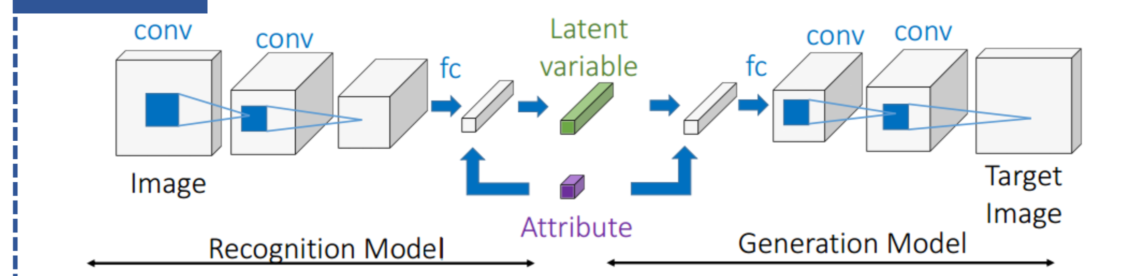
End-to-end *adversarial learning* between Control and Ranker modules

Control



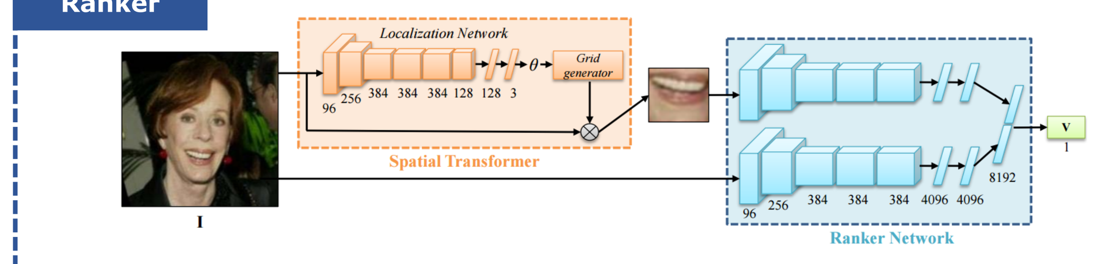
Architecture: Given a random seed Q, which is sampled from a multivariate Gaussian, the control module outputs a *pair of control parameters* Y (attributes) and Z (latent) that are used as inputs to the generator module.

Generator



Attribute2Image [Yan et al. '16]: Attribute-conditioned image generator

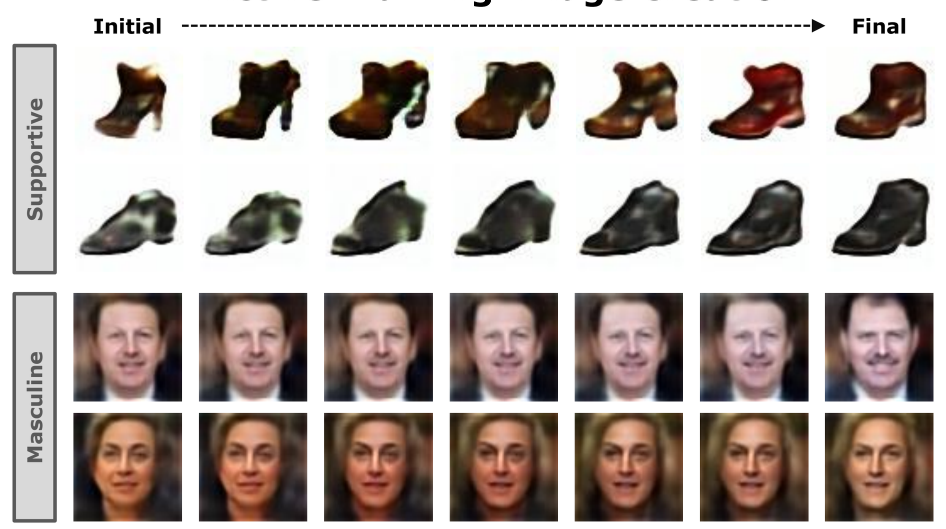
Ranker



DeepSTN [Singh & Lee '16]: Deep learning-to-rank model using spatial transformer

Image Synthesis & Progression

Active Training Image Creation

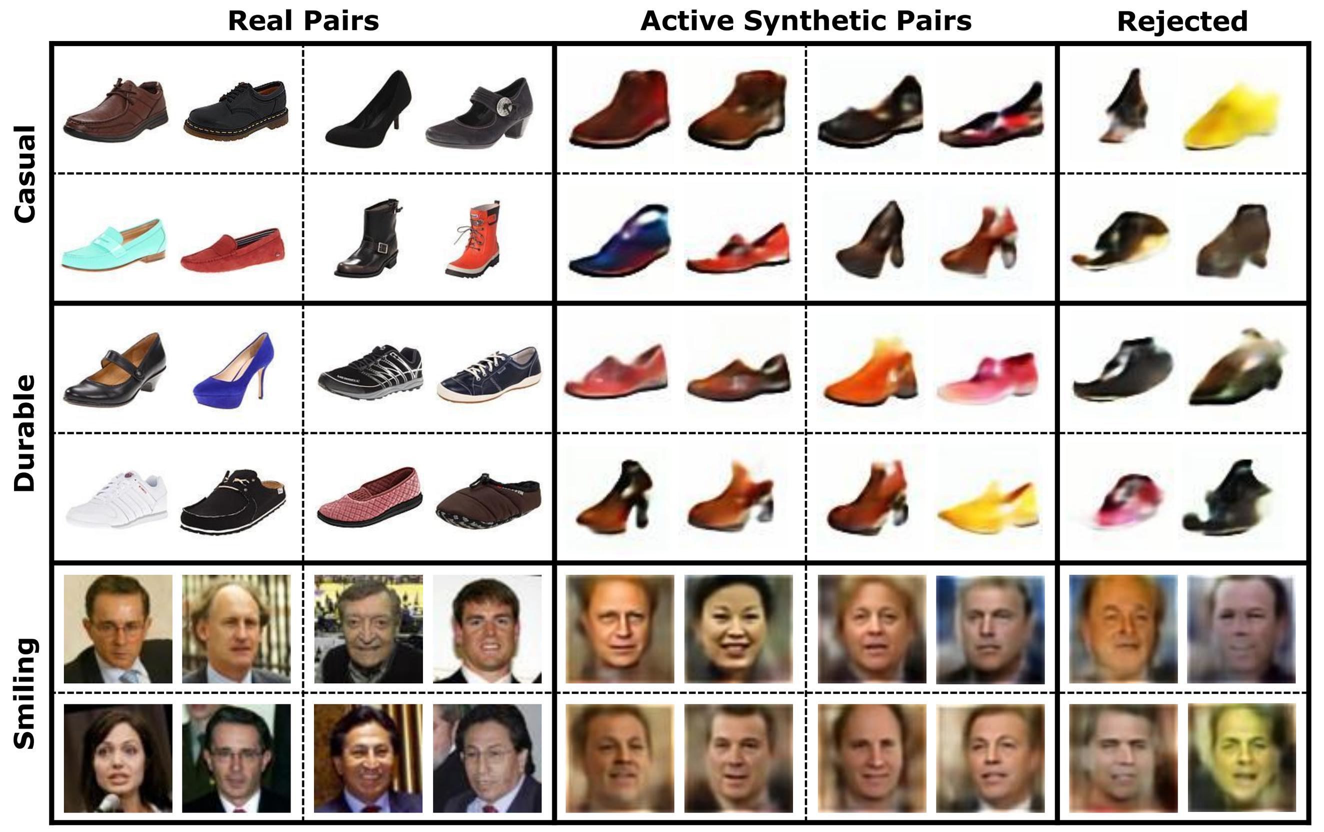


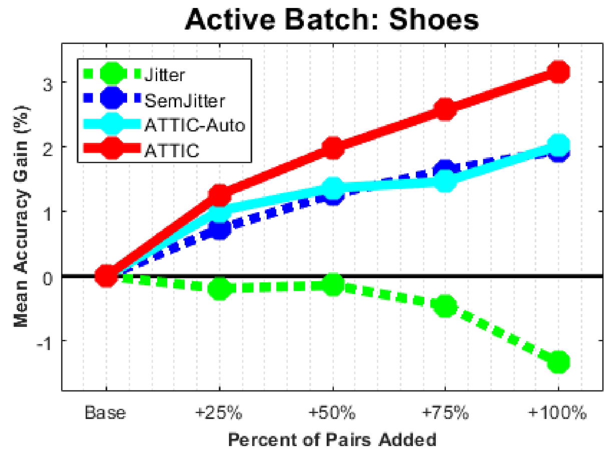
[Yu & Grauman '17]

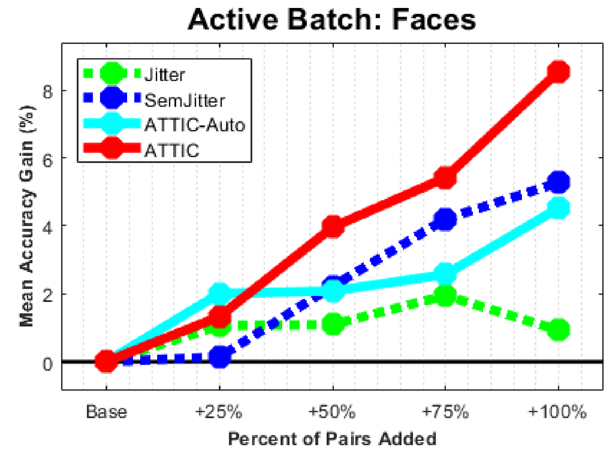
Semantic Jitter (SemJitter)

Observation: Whereas prior approach (top) samples synthetic pairs offline in a *passive* manner by adjusting one attribute at a time, our approach (left) *actively* modifies each pair as a single unit by adjusting multiple attributes simultaneously.

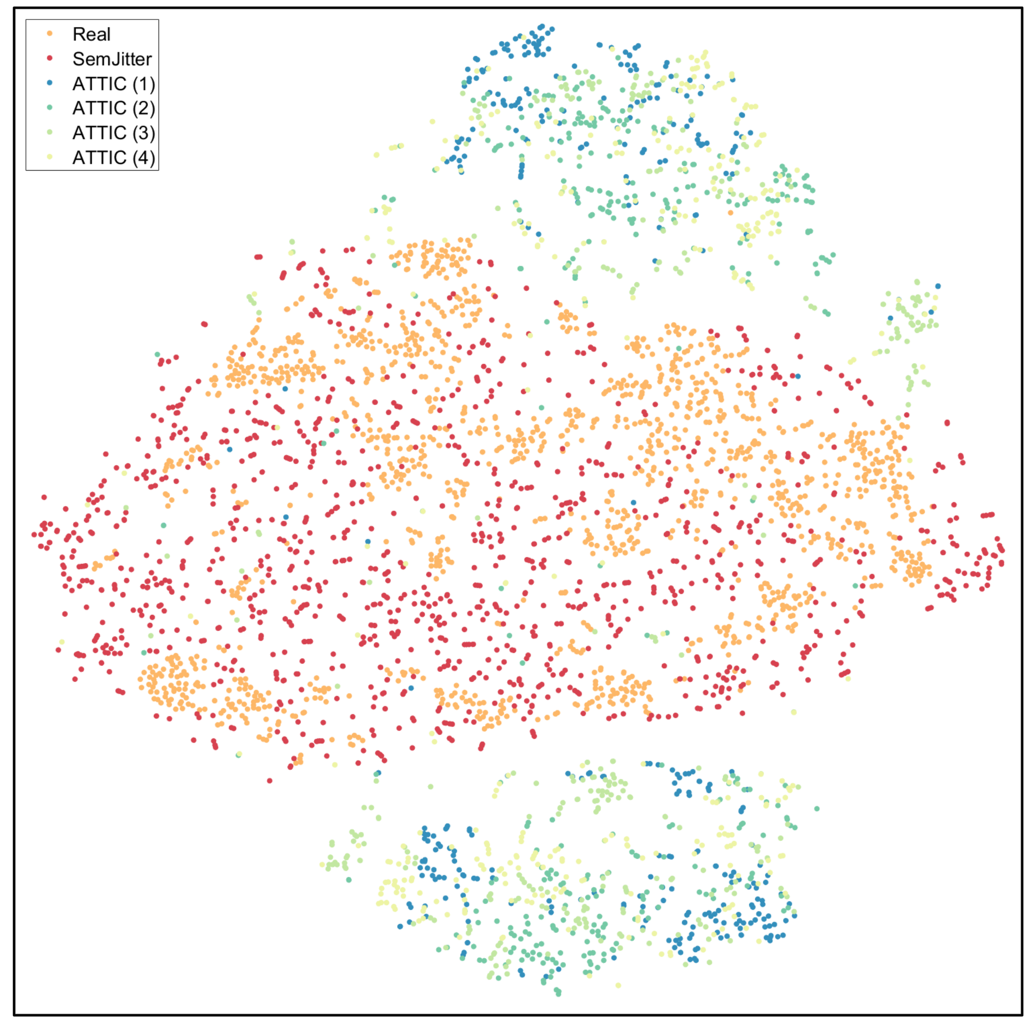
Experimental Results

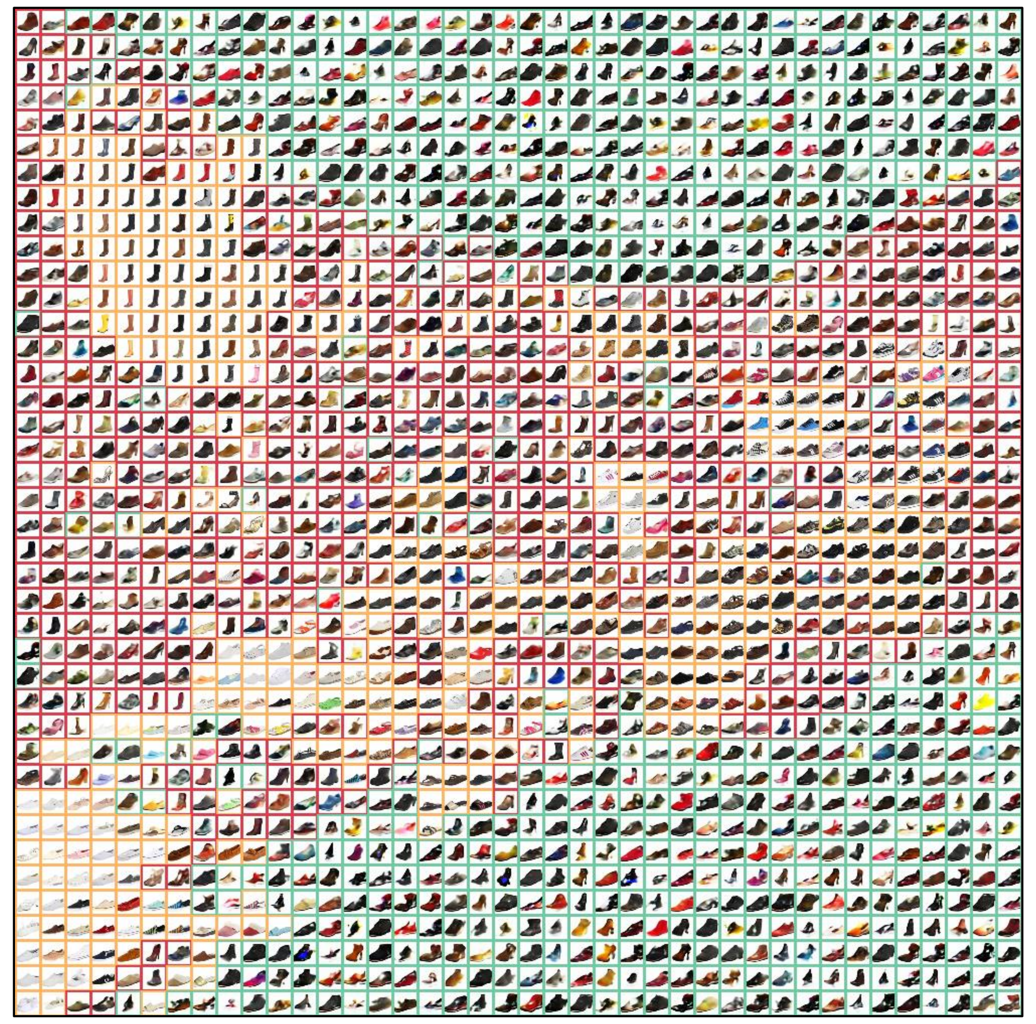






	Real	Jitter	SemJitter	ATTIC	SemJitter (Auto)	ATTIC (Auto)
Zappos50K (shoes)	86.74	85.60	86.87	87.62	87.59	**89.07**
LFW/LFW-10 (faces)	82.30	82.99	84.81	**84.99**	83.73	84.59





t-SNE Embedding: Densification of the training space through interpolation and extrapolation beyond the real training images.