Recognizing Image Style: Extended Abstract

Sergey Karayev[1]
Matthew Trentacoste[2]
Helen Han[1]
Aseem Agarwala[2]
Trevor Darrell[1]
Aaron Hertzmann[2]
Holger Winnemoeller[2]

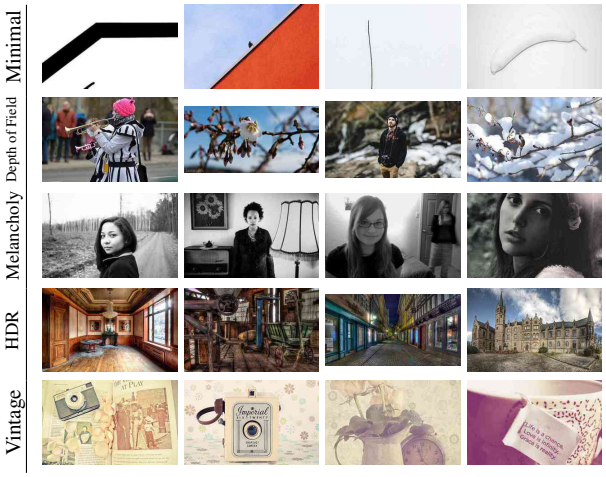
[1] University of California, Berkeley
[2] Adobe

Deliberately-created images convey meaning, and *visual style* is often a significant component of image meaning. For example, a political candidate portrait made in the lush colors of a Renoir painting tells a different story than if it were in the harsh, dark tones of a horror movie. While understanding style is crucial to image understanding, very little research in computer vision has explored visual style.

We present two novel datasets of image style, describe an approach to predicting style of images, and perform a thorough evaluation of different image features for these tasks. We find that features learned in a multi-layer network generally perform best – even when trained with object class (not style) labels. Our approach shows excellent classification performance on both datasets, and we use the learned classifiers to extend traditional tag-based image search to consider stylistic constraints.

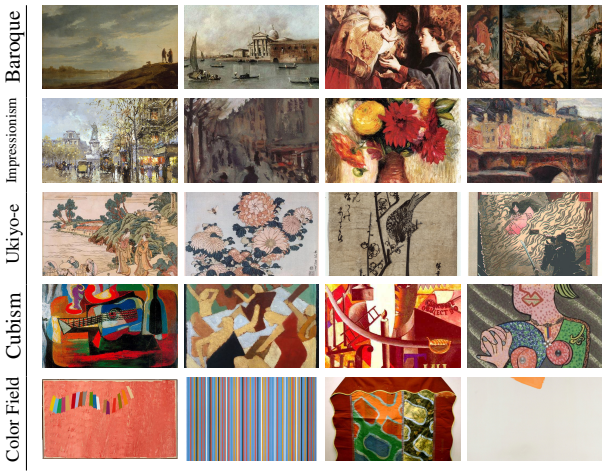
Flickr Style Using curated Flickr Groups, we gather 80K photographs annotated with 20 style labels, ranging from photographic techniques ("Macro," "HDR"), composition styles ("Minimal," "Geometric"), moods ("Serene," "Melancholy"), genres ("Vintage," "Romantic," "Horror"), to types of scenes ("Hazy," "Sunny").

Top five predictions on the test set for a selection of styles:



Wikipaintings Using community-annotated data, we gather 85K paintings annotated with 25 style/genre labels.

Top five predictions on the test set for a selection of styles:



Features and Learning We test the following features: **L*a*b color** histogram, **GIST** descriptor, Graph-based **visual saliency**, Meta-class binary (**MC-bit**) object features, and deep convolutional neural networks (CNN), using the Caffe implementation of Krizhevsky's ImageNet architeture (referred to as the **DeCAF** feature, with subscript denoting network layer). Notably, the last two of these are features designed and trained for object recognition.

As we hypothesize that style features may be content dependent, we also train **Content** classifiers using the CNN features and an aggregated version of the PASCAL VOC dataset, and use them in second-stage fusion with other features.

Evaluation Mean APs on three datasets for the considered single-channel features and their second-stage combination. Only the clearly superior features are evaluated on the Flickr and Wikipaintings datasets.

	Fusion x Content	$DeCAF_6$	MC-bit	L*a*b* Hist	GIST	random
AVA Style	**0.581**	0.579	0.539	0.288	0.220	0.132
Flickr	**0.368**	0.336	0.328	-	-	0.052
Wikipaintings	**0.473**	0.356	0.441	-	-	0.043

We compare our predictors to human observers, using Amazon Mechanical Turk experiments, and find that our classifiers predict Group membership at essentially the same level of accuracy as Turkers. We also test on the AVA aesthetic prediction task, and show that using the "deep" object recognition features improves over state-of-the-art results.

Applications Example of filtering image search results by style. Our Flickr Style classifiers are applied to images found on Pinterest. The images are searched by the text contents of their captions, then filtered by the response of the style classifiers. Here we show top five results for the query "Dress."

Code & Data All data, trained predictors, and code are available at `http://sergeykarayev.com/recognizing-image-style/`.