Visualizing and Describing Fine-grained Categories as Textures

Tsung-Yu Lin, Mikayla Timm, Chenyun Wu, Subhransu Maji
University of Massachusetts Amherst

Understanding the remarkable performance of texture-based classifiers on fine-grained recognition. Describing texture attributes of fine-grained categories using natural language for interpretability.

Visualizing Maximal Textures

Texture-based classifier (Bilinear CNN [1]):
• VGG-16 Network (layer 2_2, 3_3, 4_3, 5_3)
• Covariance matrix + Normalization + Linear layer
• Softmax for predicting class probabilities

Maximal textures: Maximize target class probability subject to a "natural" image prior [2]:

$$
\min_{x} \sum_{i=1}^{n} L(C_i, \hat{C}_i) + \gamma TV(x)
$$

Maximal images capture discriminative texture properties.

Describing Maximal Textures

Dataset: For each image in the Describable Texture Dataset [3], we collected attribute phrases (e.g. "black and white dots.")

Prediction: We trained a deep network to predict attribute phrases on our dataset. The top 20 attribute phrases for each maximal image are shown as a word cloud.

For each category in a dataset, we show training examples (top), texture images (middle), and texture attributes (bottom).


Motivation

Caltech-UCSD Birds
Oxford Flowers
FGVC Butterflies and Moths
FGVC Fungi
FGVC Flowers

Visit us at: fgvc.org
At CVPR 2019 • Long Beach, CA
June 17, 2019
Visual Fine-Grained Categorization
6th Workshop on Fine-Grained Categorization...