LEARNED FORENSIC SOURCE SIMILARITY FOR UNKNOWN CAMERA MODELS

Owen Mayer, Matthew C. Stamm
Drexel University, Philadelphia, PA, USA

Overview

- Existing forensic camera model identification approaches assume a closed set of camera models
  - Not feasible to scale to real-world applications
- Instead, ask: Are two image patches captured by the **same or different** camera model?
  - Open set camera model comparison
  - Image splicing detection and localization
- Approach: learn a forensic **similarity measure** in two successive phases

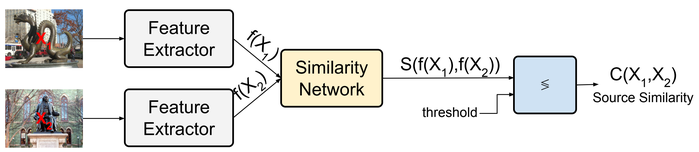
  Learning Phase A: CNN based feature extractor for camera model identification

  Learning Phase B: similarity network to compare pairs of features
  - Output a score indicating whether two input image patches were captured by the same or different camera model
- Evaluate on unknown camera models

System

For two image patches $X_1,\ X_2 \in \mathbb{X}$:

$$C(X_1, X_2) = \begin{cases} 0 & \text{if } X_1, X_2 \text{ from different camera models} \\ 1 & \text{if } X_1, X_2 \text{ from the same camera model} \end{cases}$$



- Feature extractor: $f : \mathbb{X} \to \mathbb{R}^N$
- Similarity network: $S : \mathbb{R}^N \times \mathbb{R}^N \to [0, 1]$
  - 1 indicates similar camera models, 0 dissimilar

Overall system:

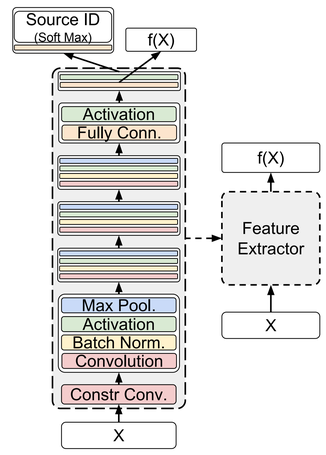
$$C(X_1, X_2) = \begin{cases} 0 & \text{if } S(f(X_1), f(X_2)) \leq \eta \\ 1 & \text{if } S(f(X_1), f(X_2)) > \eta \end{cases}$$

where η is the decision threshold

Learning Phase A - Feature Extractor

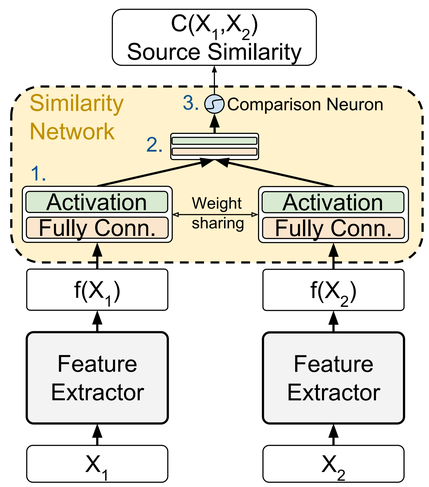
Learn a low-dimensional feature extractor that encodes high-level camera model information

- Employ MISLnet convolutional neural network architecture
  - constrained convolutional layer suppresses image content
  - 256 x 256 patches, green channel
- 5 convolutional layers
- 2 fully connected layers
  - 200 neurons each layer
- Train on 2M patches from 40 camera models
- Features are extracted from second fully connected layer



Learning Phase B - Similarity Network

Learn a mapping of feature pairs to a source similarity score, indicating different or same camera model

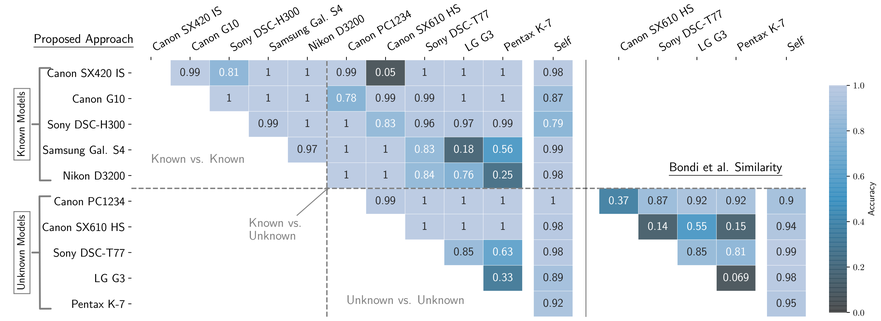


1. Two fully connected layers in *Siamese* configuration
   - maps each input feature vector to new feature space
2. Fully connected layer maps concatenated Siamese features
3. Comparison neuron (sigmoid)
   - activation value corresponds to similarity of camera models
- Feature is extractor fixed during training
- Trained on 20 camera models not used in Learning Phase A

Experimental Results

Source differentiation

Evaluated accuracy on patches from 5 unknown (not used in training) and 5 known camera models, using 100,000 unique patches randomly paired

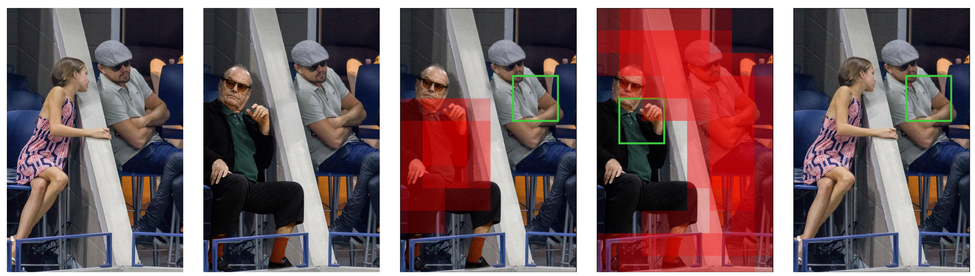


Classification accuracy by camera model

- Known vs. known 95.8% accuracy
- Known vs. unknown 84.0% accuracy
- Unknown vs unknown 90.6% accuracy
- Confusion between like makes (e.g. Canon), and like types (e.g. cellphones)

Splicing localization

Spliced images are often a composite of content from multiple cameras



Original Spliced Host Reference Foreign Ref. Host Ref., Original

Spliced image downloaded from reddit.com/r/photoshopbattles

Spliced Host Reference

Spliced image created using an AR sticker with a Pixel 2

om82@drexel.edu