LEARNED FORENSIC SOURCE SIMILARITY FOR UNKNOWN CAMERA MODELS Owen Mayer, Matthew C. Stamm

Overview

- Existing forensic camera model identification proaches assume a closed set of camera mode - 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 , $X_2 \in \mathbb{X}$:
$\mathcal{C}(X_1,X_2) = egin{cases} 0 & ext{if } X_1,X_2 ext{ from different camera mod} \ 1 & ext{if } X_1,X_2 ext{ from the same camera mod} \end{cases}$
Feature Feature Feature Feature Feature Feature Feature Feature Feature Feature Feature Feature Feature Feature Similarity Network S(f(X_1),f(X_2)) threshold Source S
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:
$ \left(\bigcap_{i \in S} \left(f(X_1) - f(X_2) \right) \right) < n $

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

misl.ece.drexel.edu

Learning Phase A - Feature Extractor

า	а	p-
e	S	

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

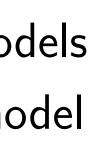
Source ID

<u>(Soft Max)</u>

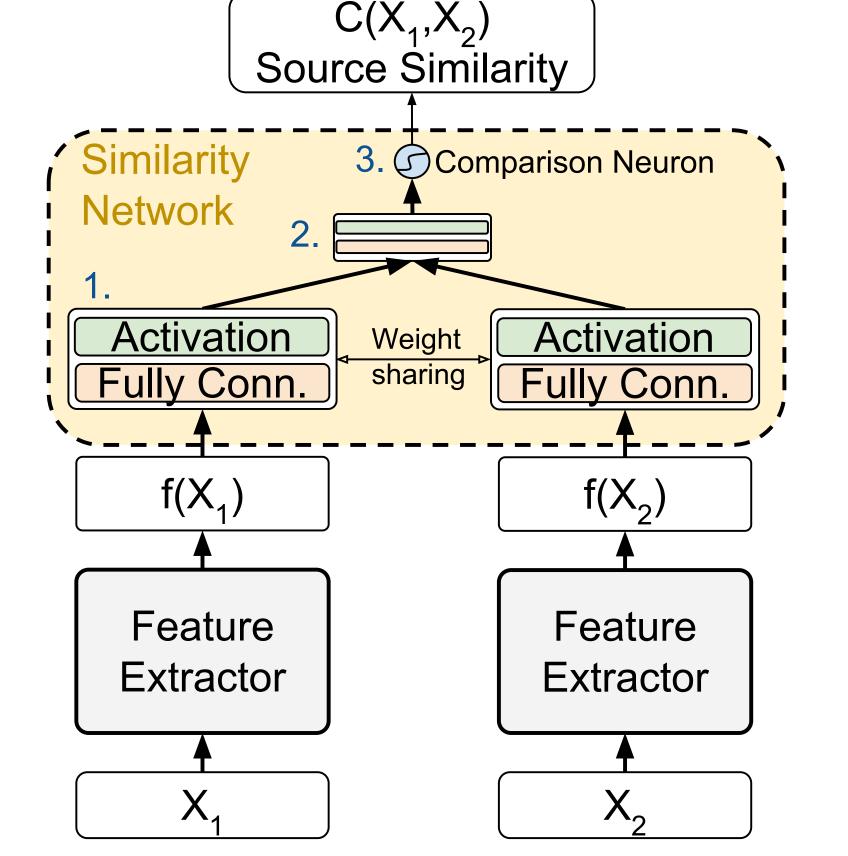
- Employ MISLnet convolutional neural network architecture - constrained convolutional layer suppresses image content - 256 x 256 patches, green channel
- ► 5 convolutional layers
- ► 2 fully connected layers - 200 neurons each layer
- ► Train on 2M patches from 40 camera models
- Features are extracted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



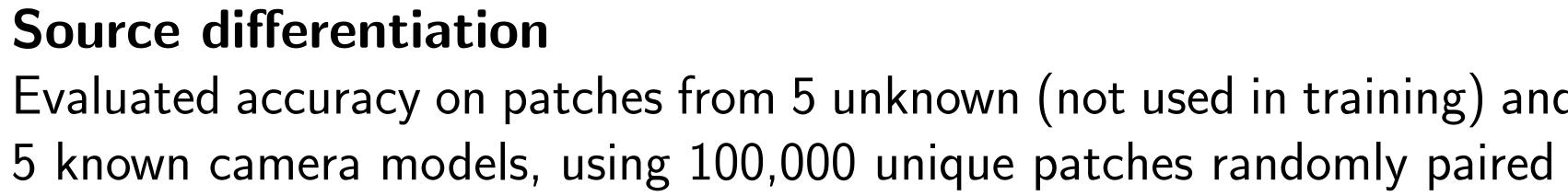
(X1,X2) Similarity

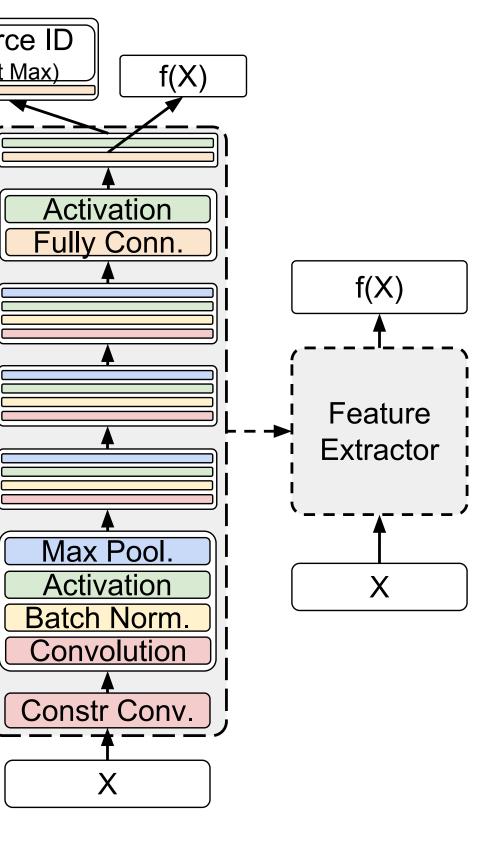


- 1. Two fully connected layers in *Siamese* configuration maps each input feature vector to new feature space
- 2. Fully connected layer maps concatenated Siamese features
- 3. Comparison neuron (sigmoid)
- activation value corresponds to similarity of camera models
- Feature is extractor fixed during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

Drexel University, Philadelphia, PA, USA

Experimental Results





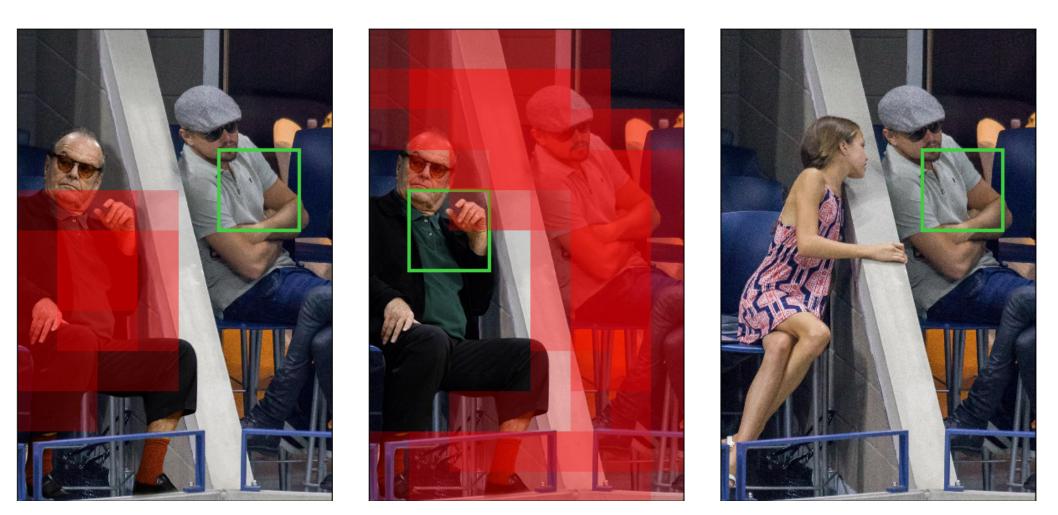
Known camera models, using 100,000 unique pateries randoning parte																		
P	roposed Approach	Canon SX420	ion G10	NDSC-H?	300 Nike	.54 on D3200 Can	on PC123	ion SX610	y DSC-TT	G ³ Pent	ax K-7	5elf	Car	ion SX61	NHS LG	7 5 ³ Penté	self	
Г	Canon SX420 IS -	0.99	0.81	1	1	0.99	0.05	1	1	1	0.9	8		·	·	·	·	1.0
dels	Canon G10 -		1	1	1	0.78	0.99	0.99	1	1	0.8	37						- 1.0
Known Models	Sony DSC-H300 -			0.99	1	1	0.83	0.96	0.97	0.99	0.7	[′] 9						- 0.8
Knov	Samsung Gal. S4 -	Known vs.		0.97	1	1	0.83	0.18	0.56	0.9	9							
Nikon D3200 -					1	1	0.84	0.76	0.25	0.9	8		- 0.6					
[Canon PC1234 -						0.99	1	1	1	1		0.37	0.87	0.92	0.92	0.9	Accuracy
lodels	Canon SX610 HS -	Known vs. Unknown						1	1	1	0.9	8		0.14	0.55	0.15	0.94	- 0.4
	Sony DSC-T77 -								0.85	0.63	0.9	8			0.85	0.81	0.99	
Unknown Models	LG G3 -									0.33	0.8	39	0.069 0.98					- 0.2
Pentax K-7 -				Unk	nown v	s. Unkn	own		0.9)2					0.95	- 0.0		

- ► 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



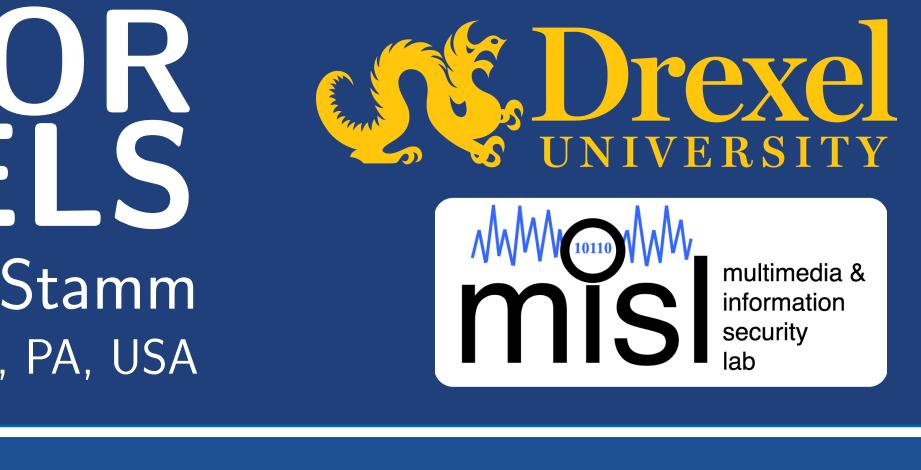




Original Spliced



Host Reference Spliced Spliced image created using an AR sticker with a Pixel 2



Evaluated accuracy on patches from 5 unknown (not used in training) and

Classification accuracy by camera model

Spliced images are often a composite of content from multiple cameras

Foreign Ref. Host Ref., Original Host Reference Spliced image downloaded from reddit.com/r/photoshopbattles