Seeing Behind the Camera: Identifying the Authorship of a Photograph

Christopher Thomas   Adriana Kovashka

Department of Computer Science
University of Pittsburgh

{chris, kovashka}@cs.pitt.edu

Abstract

We introduce the novel problem of identifying the photographer behind a photograph. To explore the feasibility of current computer vision techniques to address this problem, we created a new dataset of over 180,000 images taken by 41 well-known photographers. Using this dataset, we examined the effectiveness of a variety of features (low and high-level, including CNN features) at identifying the photographer. We also trained a new deep convolutional neural network for this task. Our results show that high-level features greatly outperform low-level features. We provide qualitative results using these learned models that give insight into our method's ability to distinguish between photographers, and allow us to draw interesting conclusions about what specific photographers shoot. We also demonstrate two applications of our method.

1. Introduction

"Motif Number 1", a simple red fishing shack on the river, is considered the most frequently painted building in America. Despite its simplicity, artists' renderings of it vary wildly from minimalistic paintings of the building focusing on the sunset behind it to more abstract portrayals of its reflection in the water. This example demonstrates the great creative license artists have in their trade, resulting in each artist producing works of art reflective of their personal style. Though the differences may be more subtle, even artists practicing within the same movement will produce distinct works, owing to different brush strokes, choice of focus and objects portrayed, use of color, portrayal of space, and other features emblematic of the individual artist. While predicting authorship in paintings and classifying painterly style are challenging problems, there have been attempts in computer vision to automate these tasks [32, 21, 19, 33, 2, 9, 5].

While researchers have made progress towards matching the human ability to categorize paintings by style and authorship [32, 5, 2], no attempts have been made to rec-

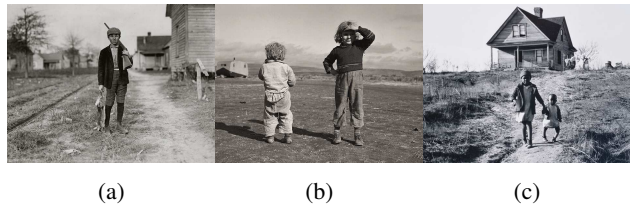(a)   (b)   (c)

Figure 1: Three sample photographs from our dataset taken by Hine, Lange, and Wolcott, respectively. Our top-performing feature is able to correctly determine the author of all three photographs, despite the very similar content and appearance of the photos.

ognize the authorship of *photographs*. This is surprising because the average person is exposed to many more photographs daily than to paintings.

Consider again the situation posed in the first paragraph, in which multiple artists are about to depict the same scene. However this time instead of painters, imagine that the artists are photographers. In this case, the stylistic differences previously discussed are not immediately apparent. The stylistic cues (such as brush stroke) available for identifying a particular artist are greatly reduced in the photographic domain due to the lessened authorial control in that medium (we do not consider photomontaged or edited images in this study). This makes the problem of identifying the author of a photograph significantly more challenging than that of identifying the author of a painting.

Fig. 1 shows photographs taken by Lewis Hine, Dorothea Lange, and Marion Wolcott, three iconic American photographers.[1] All three images depict child poverty and there are no obvious differences in style, yet our method is able to correctly predict the author of each.

The ability to accurately extract stylistic and authorship information from artwork computationally enables a wide array of useful applications in the age of massive online image databases. For example, a user who wants to retrieve

---

[1]Both Lange and Wolcott worked for the Farm Security Administration (FSA) documenting the hardship of the Great Depression, while Hine worked to address a number of labor rights issues.

more work from a given photographer, but does not know his/her name, can speed up the process by querying with a sample photo and using "Search by artist" functionality that first recognizes the artist. Automatic photographer identification can be used to detect unlawful appropriation of others' photographic work, *e.g.* in online portfolios, and could be applied in resolution of intellectual property disputes. It can also be employed to analyze relations between photographers and discover "schools of thought" among them. The latter can be used in attributing historical photographs with missing author information. Finally, understanding a photographer's style might enable the creation of novel photographs in the spirit of a known author.

This paper makes several important contributions: 1) we propose the problem of photographer identification, which no existing work has explored; 2) due to the lack of a relevant dataset for this problem, we create a large and diverse dataset which tags each image with its photographer (and possibly other metadata); 3) we investigate a large number of pre-existing and novel visual features and their performance in a comparative experiment in addition to human baselines obtained from a small study; 4) we provide numerous qualitative examples and visualizations to illustrate: the features tested, successes and failures of the method, and interesting inferences that can be drawn from the learned models; 5) we apply our method to discover schools of thought between the authors in our dataset; and 6) we show preliminary results on generating novel images that *look like* a given photographer's work.[2]

The remainder of this paper is structured as follows. Section 2 presents other research relevant to this problem and delineates how this paper differs from existing work. Section 3 describes the dataset we have assembled for this project. Section 4 explains all of the features tested and how they were learned, if applicable. Section 5 contains our quantitative evaluation of the different features and an analysis of the results. Section 6 provides qualitative examples, as well as two applications of our method. Section 7 concludes the paper.

2. Related Work

The task of automatically determining the author of a particular work of art has always been of interest to art historians whose job it is to identify and authenticate newly discovered works of art. The problem has been studied by vision researchers, who attempted to identify Vincent van Gogh forgeries, and to identify distinguishing features of painters [31, 14, 19, 10]. While the early application of art analysis was for detecting forgeries, more recent research has studied how to categorize paintings by school (*e.g.*,

"Impressionism" vs "Secession") [32, 21, 19, 33, 2, 5, 7]. [32] explored a variety of features and metric learning approaches for computing the similarity between paintings and styles. Features based on visual appearance and image transformations have found some success in distinguishing more conspicuous painter and style differences in [7, 33, 21], all of which explored low level-image features on simple datasets. Recent research has suggested that when coupled with object detection features, the inclusion of low-level features can yield state-of-the-art performance [5]. [2] used the Classeme [34] descriptor as their semantic feature representation. While it is not obvious that the object detections captured by Classemes would distinguish painting styles, Classemes outperformed all of the low-level features. This indicates that the objects appearing in a painting are also a useful predictor of style.

Our work also considers authorship identification, but the change of domain from painting to photography poses novel challenges that demand a different solution than that which was applied for painter identification. The distinguishing features of painter styles (paint type, smooth or hard brush, etc.) are inapplicable to the photography domain. Because the photographer lacks the imaginative canvas of the painter, variations in photographic style are much more subtle. Complicating matters further, many of the photographers in our dataset are from roughly the same time period, some even working for the same government agencies with the same stated job purpose. Thus, photographs taken by the subjects tend to be very similar in appearance and content, making distinguishing them particularly challenging, even for humans.

There has been work in computer vision that studies aesthetics in photography [27, 28, 11]. Some work also studies style in architecture [12, 23], vehicles [24], or yearbook phootgraphs [15]. However, all of these differ from our goal of *identifying authorship* in photography. Most related to our work is the study of visual style in photographs, conducted by [20]. Karayev *et al.* conducted a broad study on both paintings and photographs. The 20 style classes and 25 art genres considered in their study are coarse (HDR, Noir, Minimal, Long Exposure, etc.) and much easier to distinguish than the photographs in our dataset, many of which are of the same types of content and have very similar visual appearance. While [20] studied style in the context of photographs and paintings, we explore the novel problem of *photographer identification*. We find it unusual that this problem has remained unexplored for so long, given that photographs are more abundant than paintings, and there has been work in computer vision to analyze paintings. Given the lower level of authorial control that the photographer possesses compared to the painter, we believe that the photographer classification task is more challenging, in that it often requires attention to subtler cues than brush stroke,

---

[2]Automatically creating a novel Rembrandt *painting* [1] gained media attention in April 2016, five months after we submitted our work.

Adams	245	Brumfield	1138	Capa	2389	Bresson	4693	Cunningham	406	Curtis	1069	Delano	14484
Duryea	152	Erwitt	5173	Fenton	262	Gall	656	Genthe	4140	Glinn	4529	Gottscho	4009
Grabill	189	Griffiths	2000	Halsman	1310	Hartmann	2784	Highsmith	28475	Hine	5116	Horydczak	14317
Hurley	126	Jackson	881	Johnston	6962	Kandell	311	Korab	764	Lange	3913	List	2278
McCurry	6705	Meiselas	3051	Mydans	2461	O'Sullivan	573	Parr	20635	Prokudin-Gorsky	2605	Rodger	1204
Rothstein	12517	Seymour	1543	Stock	3416	Sweet	909	Van Vechten	1385	Wolcott	12173		

Table 1: Listing of all photographers and the number of photos by each in our dataset.

for example. Besides our experimental analysis of this new problem, we also contribute the first large dataset of well-known photographers and their work.

In Sec. 6.3, we propose a method for generating a new photograph in the style of an author. This problem is distinct from style transfer [4, 8, 3] which adjusts the tone or color of a photograph. Using [3] on our generated photographs did not produce a visible improvement in their quality.

3. Dataset

A significant contribution of this paper is our photographer dataset.[3] It consists of 41 well known photographers and contains 181,948 images of varying resolutions. We searched Google for "famous photographers" and used the list while also choosing authors with large, curated collections available online. Table 1 contains a listing of each photographer and their associated number of images in our dataset. The timescale of the photos spans from the early days of photography to the present day. As such, some photos have been developed from film and some are digital. Many of the images were harvested using a web spider with permission from the Library of Congress's photo archives and the National Library of Australia's digital collection's website. The rest were harvested from the Magnum Photography online catalog, or from independent photographers' online collections. Each photo in the dataset is annotated with the ID of the author, the URL from which it was obtained, and possibly other meta-data, including: the title of the photo, a summary of the photo, and the subject of the photo (if known). The title, summary, and subject of the photograph were provided by either the curators of the collection or by the photographer. Unlike other datasets obtained through web image search which may contain some incorrectly labeled images, our dataset has been painstakingly assembled, authenticated, and described by the works' curators. This rigorous process ensures that the dataset and its associated annotations are of the highest quality.

4. Features

Identification of the correct photographer is a complex problem and relies on multiple factors. Thus, we explore a broad space of features (both low and high-level). The term "low-level" means that each dimension of the feature vector

---

[3]It can be downloaded at `http://www.cs.pitt.edu/~chris/photographer`.

has no inherent "meaning." High-level features have articulatable semantic meaning (i.e. the presence of an object in the image). We also train a deep convolutional neural network from scratch in order to learn custom features specific to this problem domain.

Low-Level Features

- **L*a*b* Color Histogram:** To capture color differences among the photographers, we use a 30-dimensional binning of the L*a*b* color space. Color has been shown useful for dating historical photographs [30].

- **GIST:** GIST [29] features have been shown to perform well at scene classification and have been tested by many of the prior studies in style and artist identification [20, 32]. All images are resized to 256 by 256 pixels prior to having their GIST features extracted.

- **SURF:** Speeded-up Robust Features (SURF) [6] is a classic local feature used to find patterns in images and has been used as a baseline for artist and style identification [5, 7, 2]. We use k-means clustering to obtain a vocabulary of 500 visual words and apply a standard bag-of-words approach using normalized histograms.

High-Level Features

- **Object Bank:** The Object Bank [25] descriptor captures the location of numerous object detector responses. We believe that the spatial relationships between objects may carry some semantic meaning useful for our task.

- **Deep Convolutional Networks:**

  - **CaffeNet:** This pre-trained CNN [18] is a clone of the winner of the ILSVRC2012 challenge [22]. The network was trained on approximately 1.3M images to classify images into 1000 different object categories.

  - **Hybrid-CNN:** This network has recently achieved state-of-the-art performance on scene recognition benchmarks [38]. It was trained to recognize 1183 scene and object categories on roughly 3.6M images.

  - **PhotographerNET:** We trained a CNN with the same architecture as the previous networks to identify the author of photographs from our dataset. The network was trained for 500,000 iterations on 4 Nvidia K80 GPUs on our training set and validated on a set disjoint from our training and test sets.

Low			High													
				CaffeNet				**Hybrid-CNN**				**PhotographerNET**				
Color	GIST	SURF-BOW	Object Bank	Pool5	FC6	FC7	FC8	Pool5	FC6	FC7	FC8	Pool5	FC6	FC7	FC8	TOP
0.31	0.33	0.37	0.59	*0.73*	0.7	0.69	0.6	**0.74**	0.73	0.71	0.61	0.25	0.25	*0.63*	0.47	0.14

Table 2: Our experimental results. The F-measure of each feature is reported. The best feature overall is in **bold**, and the best one per CNN in *italics*. Note that high-level features greatly outperform low-level ones. Chance performance is 0.024.

To disambiguate layer names, we prefix them with a C, H, or P depending on whether the feature came from CaffeNet, Hybrid-CNN, or PhotographerNET, respectively. For all networks, we extract features from the Pool5, FC6, FC7 and FC8 layers, and show the result of using those features during SVM training in Table 2. The score in the TOP column for PhotographerNET is produced by classifying each test image as the author who corresponds to the dimension with the maximum response value in PhotographerNET's output (FC8).

5. Experimental Evaluation

To tested the effectiveness of the aforementioned features on the photographer classification task, using our new photographer dataset. We randomly divided our dataset into a training set (90%) and test set (10%). Because a validation set is useful when training a CNN to determine when learning has peaked, we created a validation set by randomly sampling 10% of the images from the training set and excluding them from the training set for our CNN only. The training of our PhotographerNET was terminated when performance started dropping on the validation set.

For every feature in Table 2 (except TOP which assigns the max output in FC8 as the photographer label) we train a one-vs-all multiclass SVM using the framework provided by [13]. All SVMs use linear kernels.

Table 2 presents the results of our experiments. We report the F-measure for each of the features tested. We observe that the deep features significantly outperform all low-level standard vision features, concordant with the findings of [20, 5, 32]. Additionally, we observe that Hybrid-CNN features outperform CaffeNet by a small margin on all features tested. This suggests that while objects are clearly useful for photographer identification given the impressive performance of CaffeNet, the added scene information of Hybrid-CNN provides useful cues beyond those available in the purely object-oriented model. We observe that Pool5 is the best feature within both CaffeNet and Hybrid-CNN. Since Pool5 roughly corresponds to parts of objects [37, 36, 17], we can conclude that seeing the *parts* of objects, not the *full* objects, is most discriminative for identifying photographers. This is intuitive because an artistic photograph contains many objects, so some of them may not be fully visible.

The Object Bank feature achieves nearly the same performance as C-FC8 and H-FC8, the network layers with explicit semantic meaning. All three of these features encapsulate object information, though Object Bank detects significantly fewer classes (177) than Hybrid-CNN (978) or CaffeNet (1000). Despite detecting fewer categories, Object Bank encodes more fine-grained spatial information about where the objects detected were located in the image, compared to H-FC8 and C-FC8. This finer-grained information could be giving it a slight advantage over these CNN object detectors, despite its fewer categories.

One surprising result from our experiment is that PhotographerNET does not surpass either CaffeNet or Hybrid-CNN, which were trained for object and scene detection on different datasets.[4] PhotographerNET's top-performing feature (FC7) outperforms the deepest (FC8) layers in both CaffeNet and Hybrid-CNN, which correspond to object and scene classification, respectively. However, P-FC7 performs worse than their shallower layers, especially H-Pool5. Layers of the network shallower than P-FC7, such as P-FC6 and P-Pool5, demonstrate a sharp decrease in performance (a trend opposite to what we see for CaffeNet and Hybrid-CNN), suggesting that PhotographerNET has learned different and less predictive intermediate feature extractors for these layers than CaffeNet or Hybrid-CNN. Attributing a photograph to the author with highest P-FC8 response (TOP) is even weaker because unlike the P-FC8 method, it does not make use of an SVM. It may be that the task PhotographerNET is trying to learn is too high-level and challenging. Because PhotographerNET is learning a task even more high-level than object classification and we observe that the full-object-representation is not very useful for this task, one can conclude that for photographer identification, there is a mismatch between the high-level nature of the task, and the level of representation that is useful.

In Fig. 2, we provide a visualization that might explain the relative performance of our top-performing PhotographerNET feature (P-FC7) and the best feature overall (H-Pool5). We compute the t-distributed stochastic neighborhood embeddings [35] for P-FC7 and H-Pool5. We use the embeddings to project each feature into 2-D space. We then plot the embedded features by representing them with their corresponding photographs.

We observe that H-Pool5 divides the image space in semantically meaningful ways. For example, we see that pho-

---

[4]We also tried fine-tuning the last three layers of CaffeNet and Hybrid-CNN with our photographer data, but we did not obtain an increase in performance.

tos containing people are grouped mainly at the top right, while buildings and outdoor scenes are at the bottom. We notice H-Pool5's groupings are agnostic to color or border differences. In contrast, PhotographerNET's P-FC7 divides the image space along the diagonal into black and white vs. color regions. It is hard to identify semantic groups based on the image's content. However, we can see that images that "look alike" by having similar borders or similar colors are closer to each other in the projection. This indicates that PhotographerNET learned to use lower-level features to perform photographer classification, whereas Hybrid-CNN learned higher-level semantic features for object/scene recognition. One possible explanation for this is that because the photos within each class (photographer) of our dataset are so visually diverse, the network is unable to learn semantic features for objects which do not occur frequently enough. In contrast, networks trained explicitly for object recognition only see images of that object in each class, enabling them to more easily learn object representations. Interestingly, these semantic features learned on a different problem outperform the features learned on our photographer identification problem.

To establish a human baseline for the task of photographer identification, we performed two small pilot experiments. We created a website where participants could view 50 randomly chosen images training images for each photographer. The participants were asked to review these and were allowed to take notes. Next, they were asked to classify 30 photos chosen at random from a special balanced test set. Participants were allowed to keep open the page containing the images for each photographer during the test phase of the experiment. In our first experiment, one participant studied and classified images for all 41 photographers and obtained an F1-score of 0.47. In a second study, a different participant performed the same task but was only asked to study and classify the ten photographers with the most data, and obtained an F1-score of 0.67. Our top-performing feature's performance in Table 2 (on all 41 photographers) surpasses both human F1-scores even on the smaller task of ten photographers, demonstrating the difficulty of the photographer identification problem on our challenging dataset.

Finally, to demonstrate the difficulty of the photographer classification problem and to explore the types of errors different features tend to make, we present several examples of misclassifications in Fig. 3. Test images are shown on the left. Using the SVM weights to weigh image descriptors, we find the training image (1) from the incorrectly predicted class (shown in the middle) and (2) from the correct class (shown in the right), with minimum distance to the test image. The first row (Fig. 3a-3c) depicts confusion using SURF features. All three rooms have visually similar decor and furniture, offering some explanation to Fig.

(a) P-FC7 t-SNE embeddings.

(b) H-Pool5 t-SNE embeddings.

Figure 2: t-SNE embeddings for two deep features. We observe that PhotographerNET relies more heavily on lower-level cues (like color) than higher-level semantic details.

3a's misclassification as a Gottscho image. The second row (Fig. 3d-3f) shows a misclassification by CaffeNet. Even though all three scenes contain people at work, CaffeNet lacks the ability to differentiate between the scene types (indoor vs. outdoor and place of business vs. house). In contrast, Hybrid-CNN was explicitly trained to differenti-

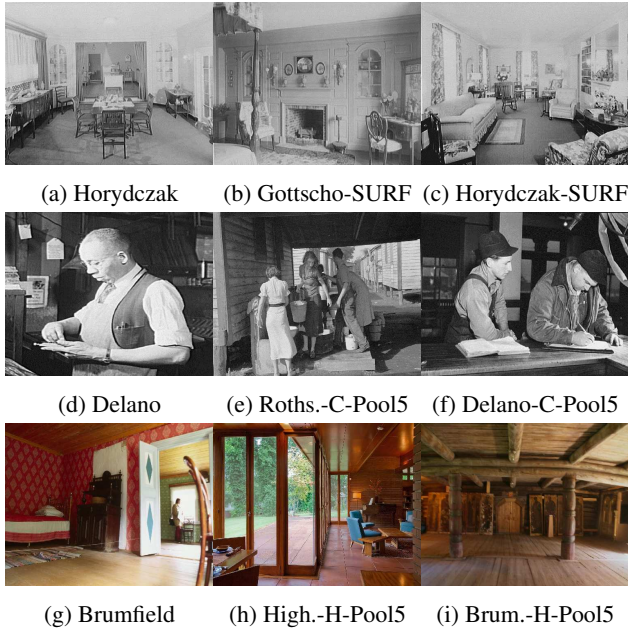| (a) Horydczak | (b) Gottscho-SURF | (c) Horydczak-SURF |
| (d) Delano | (e) Roths.-C-Pool5 | (f) Delano-C-Pool5 |
| (g) Brumfield | (h) High.-H-Pool5 | (i) Brum.-H-Pool5 |

Figure 3: Confused images. The first column shows the test image, the second shows the closest image in the predicted class, and the third shows the closest image from the correct class. Can you tell which one doesn't belong?

ate these types of scenes. The final row shows the type of misclassification made by our top-performing feature, H-Pool5. Hybrid-CNN has confused the indoor scene in Fig. 3g as a Highsmith. However, we can see that Highsmith took a similar indoor scene containing similar home furnishings (Fig. 3h). These examples illustrate a few of the many confounding factors which make photographer identification challenging.

6. Qualitative Results

The experimental results presented in the previous section indicate that classifiers can exploit semantic information in photographs to differentiate between photographers at a much higher fidelity than low-level features. At this point, the question becomes not *if* computer vision techniques can perform photographer classification relatively reliably but *how* they are doing it. What did the classifiers learn? In this section, we present qualitative results which attempt to answer this question and enable us to draw interesting insights about the photographers and their subjects.

6.1. Photographers and objects

Our first set of qualitative experiments explores the relationship of each photographer to the objects which they photograph and which differentiate them. Each dimension of the 1000-dimensional C-FC8 vector produced by CaffeNet represents a probability that its associated ImageNet synset is the class portrayed by the image. While C-FC8

does not achieve the highest F-measure, it has a clear semantic mapping to ImageNet synsets and thus can be more easily used to reason about what the classifiers have learned. Because the C-FC8 vector is high-dimensional, we "collapse" the vector for purposes of human consideration. To do this, we map each ImageNet synset to its associated WordNet synset and then move up the WordNet hierarchy until the first of a number of manually chosen synsets[5] are encountered, which becomes the dimension's new label. This reduces C-FC8 to 54 coarse categories by averaging all dimensions with the same coarse label. In Fig. 4, we show the average response values for these 54 coarse object categories for each photographer. Green indicates positive values and red indicates negative values. Darker shades of each color are more extreme.

We apply the same technique to collapse the learned SVM weights. During training, each one-vs-all linear SVM learns a weight for each of the 1000 C-FC8 feature dimensions. Large positive or negative values indicate a feature that is highly predictive. Unlike the previous technique which simply shows the average object distribution per photographer, using the learned weights allows us to see what categories specifically *distinguish* a photographer from others. We show the result in Fig. 5.

Finally, while information about the 54 *types* of objects photographed by each author is useful, finer-grained detail is also available. We list the top 10 individual categories with highest H-FC8 weights (which captures both objects and scenes). To do this, we extract and average the H-FC8 vector for all images in the dataset for each photographer. We list the top 10 most represented categories for a select group of photographers in Table 3, and include example photographs by each photographer.

We make the following observations about the photographers' style from Figs. 4 and 5 and Table 3. From Fig. 4, we conclude that Brumfield shoots significantly fewer people than most photographers. Instead, Brumfield shoots many "buildings" and "housing." Peering deeper, Brumfield's top ten categories in Table 3 reveal that he frequently shot architecture (such as mosques and stupas). In fact, Brumfield is an architectural photographer, particularly of Russian architecture. In contrast, Van Vechten has high response values for categories such as "clothing", "covering", "headdress" and "person". Van Vechten's photographs are almost exclusively portraits of people, so we observe a positive SVM weight for "person" in Fig. 5.

Comparing Figs. 4 and 5, we see that there is not a clear correlation between object frequency and the object's SVM weight. For instance, the "weapon" category is frequently

---

[5]These synsets were manually chosen to form a natural human-like grouping of the 1000 object categories. Because the manually chosen synsets are on multiple levels of the WordNet hierarchy, synsets are assigned to their deepest parent.
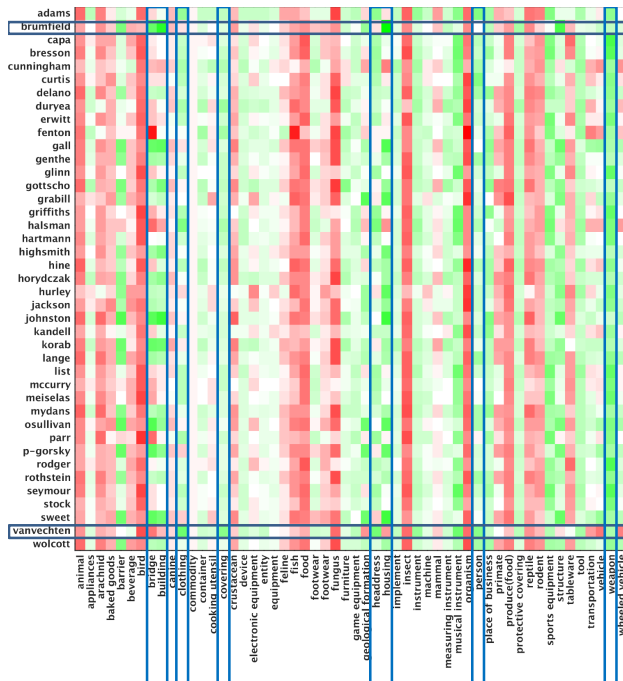
Figure 4: Average C-FC8 collapsed by WordNet. Please zoom in or view the supplementary file for a larger image.
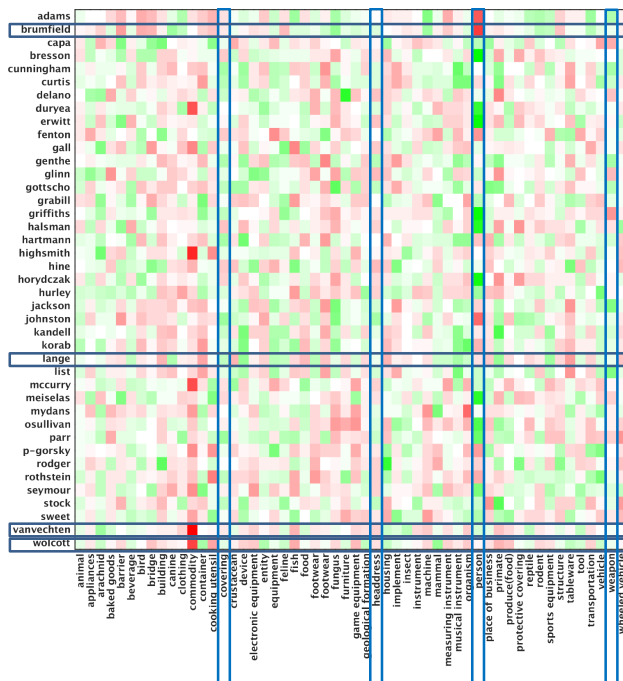

Figure 5: C-FC8 SVM weights collapsed by WordNet. Please zoom in or view supplementary for a larger image.

represented given Fig. 4, yet is only predictive of a few photographers (Fig. 5). The "person" category in Fig. 5 has high magnitude weights for many photographers, indicating its utility as a class predictor. Note that the set of objects distinctive for a photographer does not fully depend

on the photographer's environment. For example, Lange and Wolcott both worked for the FSA, yet there are notable differences between their SVM weights in Fig. 5.

6.2. Schools of thought

Taking the idea of photographic style one step further, we wanted to see if meaningful genres or "schools of thought" of photographic style could be inferred from our results. We know that twelve of the photographers in our dataset were members of the Magnum Photos cooperative. We cluster the H-Pool5 features for all 41 photographers into a dendrogram, using agglomerative clustering, and discover that nine of those twelve cluster together tightly, with only one non-Magnum photographer in their cluster. We find that three of the four founders of Magnum form their own even tighter cluster. Further, five photographers in our dataset that were employed by the FSA are grouped in our dendrogram, and two portrait photographers (Van Vechten and Curtis) appear in their own cluster. See the supplementary file for the figure. These results indicate that our techniques are not only useful for describing individual photographers but can also be used to situate photographers in broader "schools of thought."

6.3. New photograph generation

Our experimental results demonstrated that object and scene information is useful for distinguishing between photographers. Based on these results, we wanted to see whether we could take our photographer models yet another step further by generating new photographs imitating photographers' styles. Our goal was to create "pastiches" assembled by cropping objects out of each photographer's data and pasting them in new scenes obtained from Flickr. We first learned a probability distribution over the 205-scene types detected by Hybrid-CNN for each photographer. We then learned a distribution of objects and their most likely spatial location for each photographer, conditioned on the scene type. To do this, we trained a Fast-RCNN [16] object detector on 25 object categories which frequently occurred across all photographers in our dataset using data we obtained from ImageNet. We then sampled from our joint probability distributions to choose which scene to use and which objects should appear in it and where. We randomly selected a detection (in that photographer's data) for each object probabilistically selected to appear, then cropped out the detection and segmented the cropped region using [26]. We inserted the segment into the pastiche according to that photographer's spatial model for that object.

We show six pastiches generated using this approach in Fig. 6. The top row shows generated images for six photographers, and the bottom shows real images from the corresponding photographer that resemble the generated ones.

Adams	hospital room	hospital	office	mil. uniform	bow tie	lab coat	music studio	art studio	barbershop	art gallery
Brumfield	dome	mosque	bell cote	castle	picket fence	stupa	tile roof	vault	pedestal	obelisk
Delano	hospital	construction site	railroad track	slum	stretcher	barbershop	mil. uniform	train station	television	crutch
Hine	mil. uniform	pickelhaube	prison	museum	slum	barbershop	milk can	rifle	accordion	crutch
Kandell	flute	marimba	stretcher	assault rifle	oboe	rifle	panpipe	cornet	mil. uniform	sax
Lange	shed	railroad track	construction site	slum	yard	cemetery	hospital	schoolhouse	train railway	train station
Van Vechten	bow tie	suit	sweatshirt	harmonica	neck brace	mil. uniform	cloak	trench coat	oboe	gasmask

Adams   Brumfield   Delano   Hine   Kandell   Lange   Van Vechten

Table 3: Top ten objects and scenes for select photographers, and sample images.

(a) Delano   (b) Erwitt   (c) Highsmith   (d) Hine   (e) Horydczak   (f) Rothstein

Figure 6: Generated images for six photographers (top row) and real photographs by these authors (bottom row). Although results are preliminary, we observe interesting similarities between the synthetic and real work.

For example, Delano takes portraits of individuals in uniforms and of "common people," Erwitt photographs people in street scenes without their knowledge or participation, and Rothstein photographs people congregating. Highsmith captures large banner ads and Americana, Hine children working in poor conditions, and Horydczak buildings and architecture. While these are preliminary results, we see similarities between the synthetic and authentic photos.

7. Conclusion

In this paper, we have proposed the novel problem of photograph authorship attribution. To facilitate research on this problem, we created a large dataset of 181,948 images by renowned photographers. In addition to tagging each photo with the photographer, the dataset also provides rich metadata which could be useful for future research in computer vision on a variety of tasks.

Our experiments reveal that high-level features perform significantly better overall than low-level features or humans. While our trained CNN, PhotographerNET, performs reasonably well, early proto-object and scene-detection features perform significantly better. The inclusion of scene information provides moderate gains over the purely object-driven approach explored by [20, 32]. We also provide an approach for performing qualitative analysis on the photographers by determining which objects respond strongly to each photographer in the feature values and learned classifier weights. Using these techniques, we were able to draw interesting conclusions about the photographers we studied as well as broader "schools of thought." We also showed initial results for a method that creates new photographs in the spirit of a given author.

In the future, we will develop further applications of our approach, *e.g.* teaching humans to better distinguish between the photographers' styles. We will also continue our work on using our models to generate novel photographs of known photographers' styles.

References

[1] The Next Rembrandt. https://www.nextrembrandt.com/. Accessed April 9, 2016.

[2] R. S. Arora. *Towards automated classification of fine-art painting style: A comparative study*. PhD thesis, Rutgers University-Graduate School-New Brunswick, 2012.

[3] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand. Fast local laplacian filters: Theory and applications. *ACM Transactions on Graphics (TOG)*, 33(5):167, 2014.

[4] S. Bae, S. Paris, and F. Durand. Two-scale tone management for photographic look. *ACM Transactions on Graphics (TOG)*, 25(3):637–645, 2006.

[5] Y. Bar, N. Levy, and L. Wolf. Classification of artistic styles using binarized features derived from a deep neural network. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 71–84. Springer, 2014.

[6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, 2008.

[7] A. Blessing and K. Wen. Using machine learning for identification of art paintings. Technical report, Stanford University, 2010.

[8] V. Bychkovsky, S. Paris, E. Chan, and F. Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 97–104. IEEE, 2011.

[9] G. Carneiro, N. P. da Silva, A. Del Bue, and J. P. Costeira. Artistic image classification: an analysis on the printart database. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 143–157. Springer, 2012.

[10] B. Cornelis, A. Dooms, I. Daubechies, and P. Schelkens. Report on digital image processing for art historians. In *SAMPTA'09*, 2009.

[11] S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1657–1664. IEEE, 2011.

[12] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. Efros. What makes paris look like paris? *ACM Transactions on Graphics*, 31(4), 2012.

[13] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

[14] H. Farid. Image forgery detection. *Signal Processing Magazine, IEEE*, 26(2):16–25, 2009.

[15] S. Ginosar, K. Rakelly, S. Sachs, B. Yin, and A. Efros. A century of portraits: A visual historical record of american high school yearbooks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pages 1–7, 2015.

[16] R. Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

[17] X. Huang, C. Shen, X. Boix, and Q. Zhao. Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 262–270, 2015.

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.

[19] C. R. Johnson Jr, E. Hendriks, I. J. Berezhnoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang. Image processing for artist identification. *Signal Processing Magazine, IEEE*, 25(4):37–48, 2008.

[20] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller. Recognizing image style. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2014.

[21] D. Keren. Recognizing image "style" and activities in video using local features and naive bayes. *Pattern Recognition Letters*, 24(16):2913–2922, 2003.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105, 2012.

[23] S. Lee, N. Maisonneuve, D. Crandall, A. A. Efros, and J. Sivic. Linking past to present: Discovering style in two centuries of architecture. In *IEEE International Conference on Computational Photography*, 2015.

[24] Y. J. Lee, A. Efros, and M. Hebert. Style-aware mid-level representation for discovering visual connections in space and time. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1857–1864. IEEE, 2013.

[25] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1378–1386, 2010.

[26] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille. The secrets of salient object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 280–287, 2014.

[27] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1784–1791. IEEE, 2011.

[28] N. Murray, L. Marchesotti, and F. Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2408–2415. IEEE, 2012.

[29] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision (IJCV)*, 42(3):145–175, 2001.

[30] F. Palermo, J. Hays, and A. A. Efros. Dating historical color images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 499–512. Springer, 2012.

[31] G. Polatkan, S. Jafarpour, A. Brasoveanu, S. Hughes, and I. Daubechies. Detection of forgery in paintings using supervised learning. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2921–2924. IEEE, 2009.

[32] B. Saleh and A. Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*, 2015.

[33] L. Shamir, T. Macura, N. Orlov, D. M. Eckley, and I. G. Goldberg. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception (TAP)*, 7(2):8, 2010.

[34] L. Torresani, M. Szummer, and A. Fitzgibbon. Efficient object category recognition using classemes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 776–789. Springer, 2010.

[35] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[36] D. Wei, B. Zhou, A. Torralba, and W. Freeman. Understanding intra-class knowledge inside cnn. *arXiv preprint arXiv:1507.02379*, 2015.

[37] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 818–833. Springer, 2014.

[38] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems (NIPS)*, pages 487–495, 2014.