Seeing Behind the Camera: Identifying the Authorship of a Photograph (Supplementary Material)

Christopher Thomas Adriana Kovashka
Department of Computer Science
University of Pittsburgh
{chris, kovashka}@cs.pitt.edu

1 Introduction

This document includes supplemental results to those found in the main text. In this document, we present the dendrogram discussed in our "Schools of thought" section from the main text. We used this dendrogram to draw the conclusions about clusters of photographers that we describe in the main text. We also include larger and more readable versions of Figures 4-5 from the main text, which capture the coarse object categories that photographers tend to shoot, and how the one-vs-all SVMs weigh each category.

We also include a few results that we did not have room for in the main text. First, we include a figure which shows the best feature for distinguishing each pair of photographers. From this figure, we can see that even features which do not perform well overall can be useful for distinguishing certain photographers. Using this result, we include examples of misclassifications made by the best performing feature for select pairs of photographers. This illustrates how challenging the problem is since even the best feature for that particular photographer pair makes mistakes. We then include additional examples of synthetic photographs generated by our algorithm, to complement the results shown in Section 6.3 of the main text. Finally, we show additional t-SNE visualizations at a higher resolution than those shown in our main text. These visualizations illustrate how the different features we tested group photographs and give us insight into how they are doing so.

2 Schools of Thought

In Section 6.2, we describe an application of our approach to discover the "schools of thought" among photographers. To do this, we performed agglomerative clustering on feature vectors for each photographer. These were obtained by averaging the training image feature vectors for each photograph per photographer. The resulting clusters of photographers are interesting. Figure 1 shows the dendrogram we describe in the main text. For completeness, we also include the description of the figure that appeared in the main text, but we also refer to clusters in the figure by the color with which they are marked.

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 *(see the purple cluster)*, 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 *(see the pink cluster)*. Further, five photographers in our dataset that were employed by the FSA are grouped in our dendrogram *(see the green cluster)*, and the two portrait photographers (Van Vechten and Curtis) appear in their own cluster *(see the blue cluster)*. 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."

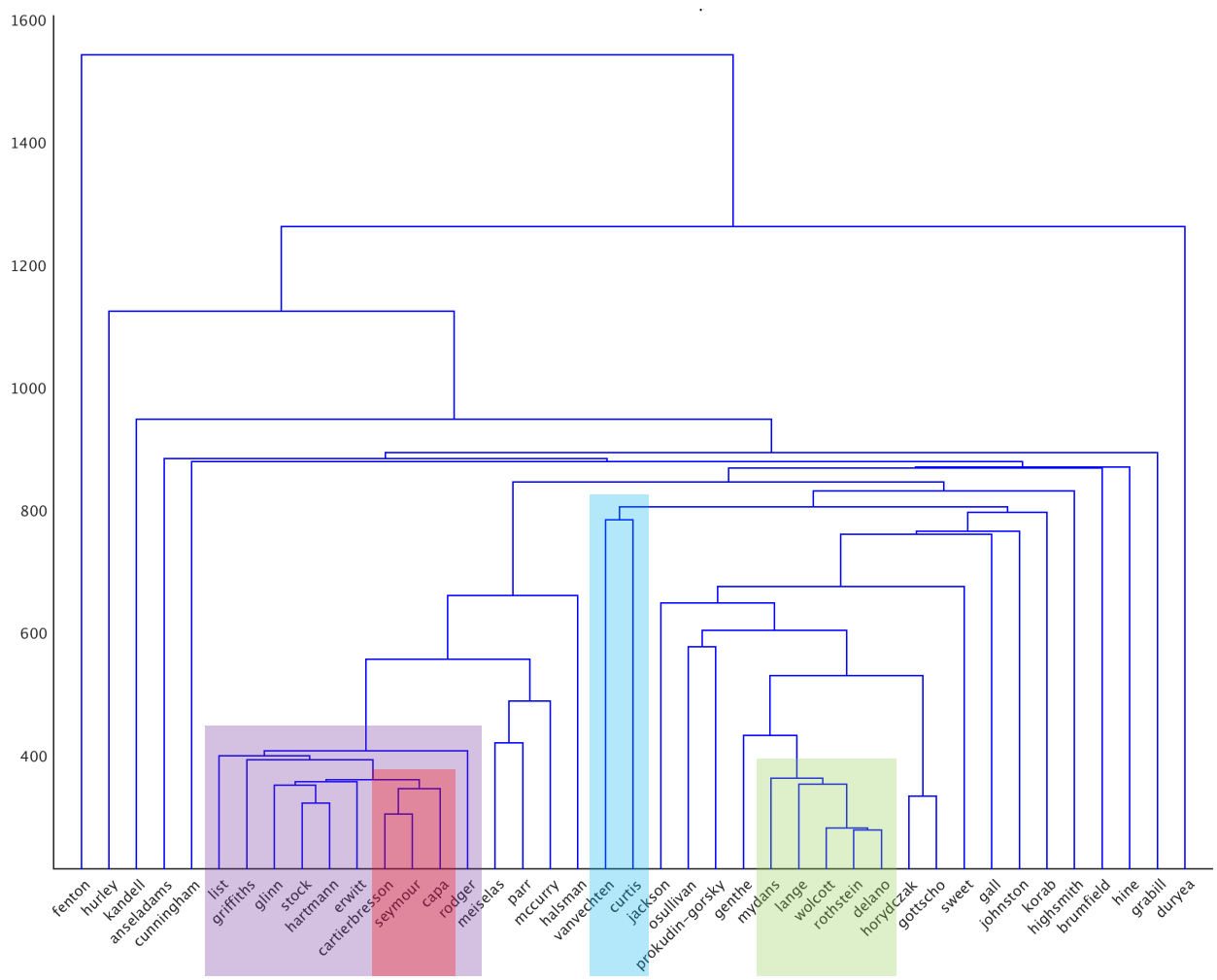


Figure 1: "Schools of Thought" created using H-Pool5

3 Collapsed C-FC8 Objects (Figure 4 from main text)

This is a larger version of Figure 4 from our paper. It was produced by collapsing the FC8 vector from each photographer into 54 coarse object categories from WordNet, using the procedure described in the main text. We then averaged the collapsed feature vectors over the training set to produce one averaged vector per photographer. We visualize the averaged vectors below. Bright green values indicate stronger positive responses while bright red responses indicate stronger negative responses. In other words, bright green categories tend to occur frequently in each photographer's train set, whereas negative categories very rarely appear. Please refer to our observations in the main text (Section 6.1) for a discussion of the figure.

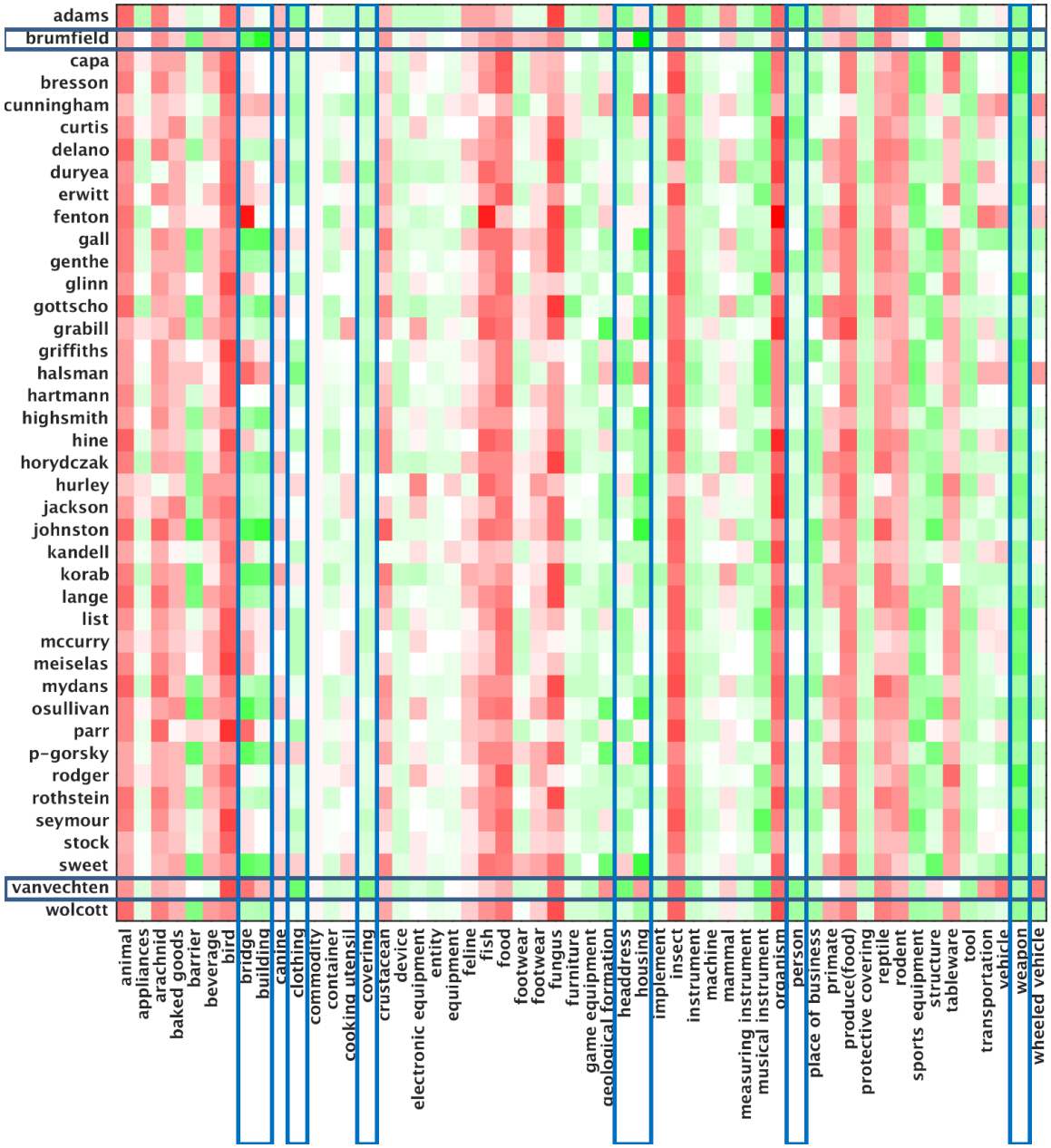


Figure 2: Average collapsed object responses of C-FC8 for each photographer

4 Collapsed C-FC8 SVM Weights (Figure 5 from main text)

This is a larger version of Figure 5 from our paper. It was produced by training one-vs-all linear SVMs on the collapsed FC8 vectors described in the previous section. We visualize the learned weights for each photographer here. Bright green values indicate stronger positive weights while bright red responses indicate stronger negative weights. More intermediate colors are not as predictive of the class as are brighter colors. Please refer to our observations in the main text (Section 6.1) for a discussion of the figure.

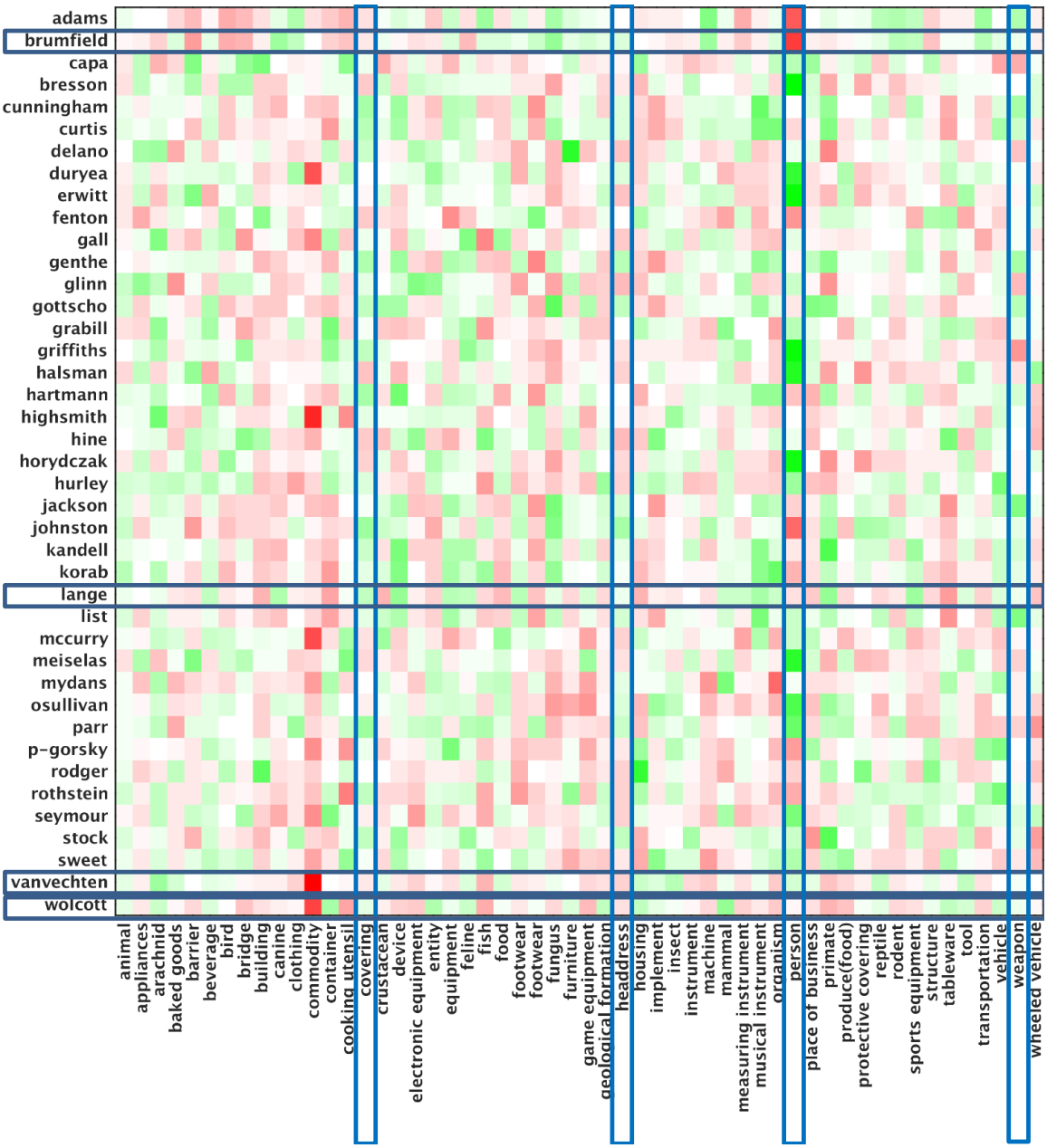


Figure 3: Collapsed SVM weights of C-FC8 for each photographer

5 Confusion Matrix for Best Performing Feature Overall (H-Pool5)

In this section, we show the confusion matrix for our best performing feature overall (H-Pool5). The bottom-right hand corner is the F-1 Score (0.74). The confusion between many photographers makes sense. For example, Delano, Rothstein, and Wolcott all worked for the FSA. We can see from the matrix that they are frequently confused. Interestingly, many of the photographers within the Magnum Photos Cooperative are also frequently confused, such as Capa, Erwitt, Glinn and Parr.

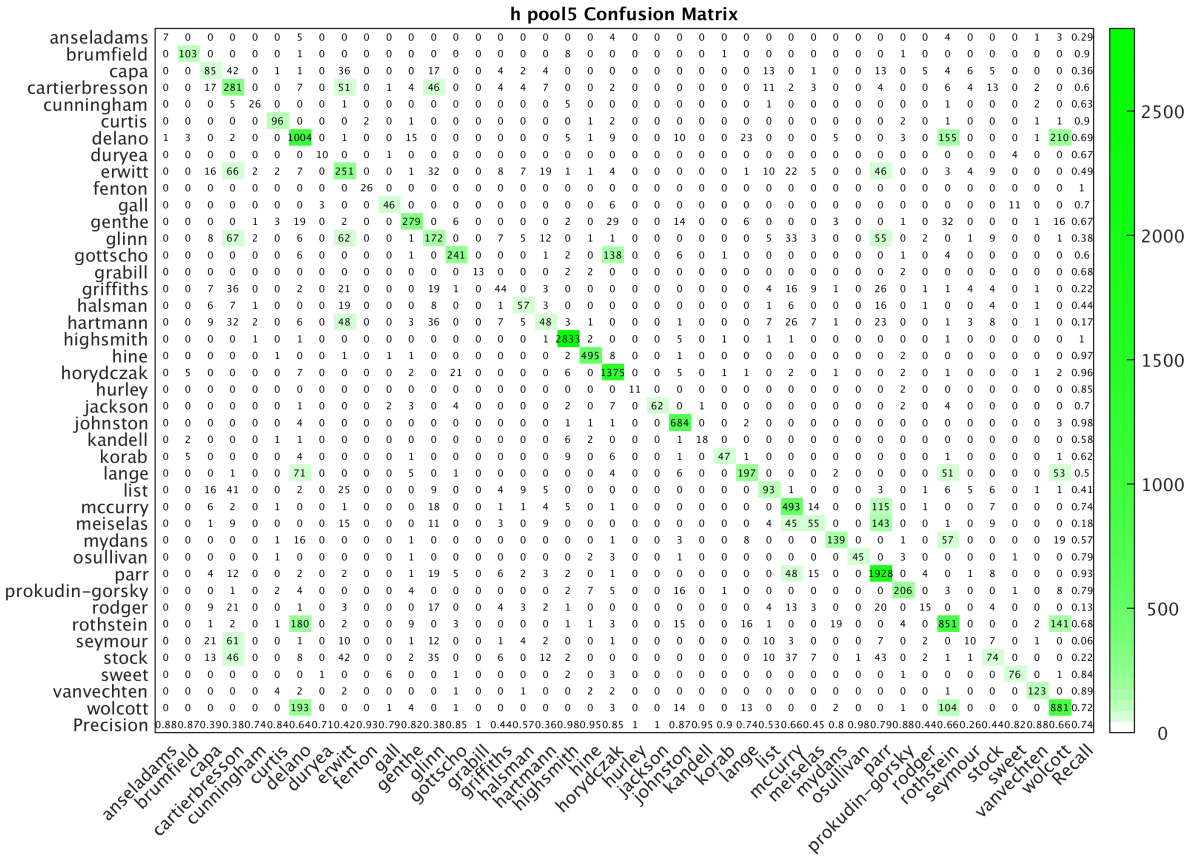


Figure 4: H-Pool5 confusion matrix.

6 Best Feature for Each Photographer-Pair

This figure shows the best feature for distinguishing between each pair of photographers. Our results show that even though H-Pool5 performs the best overall, other features are useful for distinguishing between certain photographers. For example, photographers may shoot similar objects and scenes, but their photographs may be substantially different in color. Our color feature may be able to distinguish this pair of photographers even when our higher-level features fail.

We abbreviate the feature names as follows. For CaffeNet features, we abbreviate C-FC8 as C8, C-FC7 as C7, C-FC6 as C6, and C-Pool5 as C5. For Hybrid-CNN features we use an H instead of a C and for PhotographerNET we use a P. We abbreviate GIST as GT, SURF as SF, ObjectBank as OB, and Color as CR. A * indicates that there are multiple features which achieve the same performance for discriminating between those photographers. In the case of a tie, we show one of the features randomly. From our table, we observe that C_FC8, C_FC7, and H_FC8 never appear as the best feature for distinguishing any pair of photographers. These high level features represent objects and scenes (in the case of FC8) or proto-objects (in the case of C_FC7) and are always outperformed either by deeper levels of their own network or a different feature.

	anseladams	brumfield	capa	cartierbresson	cunningham	curtis	delano	duryea	erwitt	fenton	gall	genthe	glinn	gottscho	grabill	griffiths	halsman	hartmann	highsmith	hine	horydczak	hurley	jackson	johnston	kandell	korab	lange	list	mccurry	meiselas	mydans	osullivan	parr	prokudin-gorsky	rodger	rothstein	seymour	stock	sweet	vanvechten	wolcott
anseladams	h5*	p5*	p5*	p6*	p7*	ob*	p5*	ob*	p7*	p5*	p6*	p5*	p6*	p5*	p6*	p5*	p6*	p8*	cr*	p5*	p5*	p5*	p6*	p6*	p5*	p6*	p5*	p5*	p6*	h7*	p5*	p5*	p8*	p8*	p5*	p5*	p5*	p6*	p5*	p5*	p6*
brumfield	p5*	h5*	p5*	p7*	p7*	p5*	cr*	p6*	p6*	p6*	p6*	p8*	p8*	p5*	p8*	p6*	p8*	p5*	p6*	p5*	p5*	p6*	p6*	p6*	p5*	p7*	p5*	p5*	gt*	p5*	p6*	h7*	cr*	p5*	p8*	p5*	p6*	p8*	p6*	p6*	p6*
capa	p5*	p6*	h5*	p5*	p6*	p5*	p5*	p7*	p6*	p5*	p7*	p6*	cr*	p7*	p5*	p7*	p8*	gt*	p6*	p5*	h7*	p5*	p7*	p6*	p5*	p7*	p5*	p5*	p5*	p7*	p5*	p5*	cr	p6*	p7*	p5*	p5*	cr*	p6*	p5*	p5*
cartierbresson	cr*	p7*	p6*	h6*	p5*	p7*	p5*	p6*	p5*	p5*	p8*	p5*	p5*	p5*	p5*	p6*	p6*	cr*	p7*	p6*	p6*	ob*	p7*	p6*	p7*	p5*	p5*	p5*	cr*	p8*	p5*	ob*	h5*	p8*	p5*	p5*	p6*	p5*	p5*	p6*	p5*
cunningham	p5*	p5*	p5*	p6*	h6	p6*	p6*	p7*	p8*	p5*	p6*	p5*	p6*	p6*	p5*	p5*	p5*	p5*	p6*	p5*	p6*	p5*	p7*	p5*	p8*	p5*	p5*	p5*	p6*	p7*	p5*	p5*	cr*	p8*	p5*	ob*	p5*	p5*	p6*	cr*	p5*
curtis	p8*	p5*	p5*	p7*	p6*	h5	c6*	p8*	p8*	p5*	p5*	p6*	ob*	p5*	p7*	p5*	p6*	p8*	p5*	p5*	p5	p5*	p5*	p6*	p5*	p5*	p6*	p7*	p6*	p7*	p6*	p5*	h5*	cr*	p5*	p5*	p5*	cr*	p5*	p6*	p5*
delano	p5*	p5*	p5*	cr*	p5*	p6*	h6	p6*	p5*	p5*	p5*	p8*	p6*	p5*	cr*	p8*	p5*	p7*	p6*	p5*	p6*	p6*	p6*	p5*	p6*	p5*	p5*	p8*	p8*	cr*	p7*	p5*	gt*	c5*	p5*	p6*	p5	p5*	p5*	h5*	p5*	p5
duryea	p6*	p5*	p5*	p6*	p6*	p5*	p5*	c5	p5*	p6*	p5*	p5*	p5*	p5*	p5*	p7*	p8*	p6*	p5*	sf*	p6*	p5*	p6*	p5*	p7*	p6*	h5*	p5*	p6*	p7*	p6*	p8*	p6*	ob*	p5*	p6*	p6*	p5*	p6*	p6*	p5*
erwitt	p5*	p5*	p5*	p5*	p7*	p5*	p6*	p5*	h6	cr*	p5*	p5*	p5*	cr*	p5*	p8*	p5*	p6*	p5*	p8*	p5	p7*	p5*	c6*	cr*	p5*	p5*	p5*	p5*	cr*	p7*	p6*	cr	p5*	p6*	p5*	p5*	p6*	p5*	p5*	p6*
fenton	p5*	p5*	p5*	p5*	p7*	p5*	p5*	cr*	p5*	p5*	h6*	p6*	p5*	p8*	p5*	p5*	p6*	ob*	p6*	p5*	p6*	p5*	p5*	p5*	p5*	p7*	p5*	p5*	p5*	p6*	cr*	p6*	p5*	p6*	p5*	p5*	p5*	p6*	cr*	p5*	p7*	p5*
gall	p6*	p6*	p6*	cr*	p6*	p5*	p5*	p5*	cr*	p5*	c5	p5*	p5*	p8*	p7*	p5*	p6*	p7*	h6*	p6*	p6*	p7*	p7*	p5*	p8*	p5*	p5*	p5*	p5*	p8*	p5*	p5*	h5*	p5*	p6*	p5*	cr*	p5*	p8*	p6*	p5*
genthe	p5*	p6*	p6*	p6*	p5*	p5*	p5*	p5*	p6*	p6*	p5*	h5	sf*	p5*	p7*	p5*	p8*	p5*	p6*	p5*	p5	p6*	p8*	p5	p7*	p5*	p6*	p7*	p6*	p7*	p5*	p6*	h5*	p8*	p6*	p5	p8*	p5*	cr*	p5*	p5
glinn	p6*	p6*	p5*	p6*	p5*	p5*	p5*	p5*	p5*	p5*	p7*	h7	p5*	p5*	p6*	p5*	ob*	p5*	p5*	p5*	cr*	p5*	p5*	p6*	p5*	p7*	p5*	p5*	p8*	sf*	p8*	cr	p6*	p5*	p5*	gt*	p8*	p5*	p5*	p5*	
gottscho	p5*	p5*	p5*	p5*	p6*	p8*	p5*	p5*	p6*	p5*	p5*	p6*	p5*	h5	p7*	p6*	p6*	p5*	p5*	p6*	p5	p5*	p7*	p5*	p5*	p5*	p7*	p8*	p5*	p7*	p5*	p5*	h5*	p7*	p7*	p6*	p5*	p6*	p6*	p6*	p6*
grabill	p5*	p5*	p8*	p7*	p5*	p5*	p7*	p5*	p7*	p6*	p6*	p5*	p5*	p6*	c5	p6*	p7*	p5*	p5*	cr*	p8*	p7*	p5*	p6*	p6*	p7*	p7*	p6*	p7*	p7*	p5*	p8*	p6*	p5*	p6*	p6*	p5*	p7*	p7*	p5*	p5*
griffiths	p6*	p5*	ob*	gt*	p5*	p5*	p5*	p5*	p5*	p5*	p7*	p5*	p6*	p6*	p5*	h5	p8*	p5*	p5*	p5*	p5*	p6*	p7*	p5*	p5*	p6*	p5*	p5*	p5*	sf*	ob*	p8*	cr	p6*	p5*	p5*	p6*	p5*	p5*	p6*	p5*
halsman	p5*	p5*	p5*	p5*	p6*	p6*	sf*	p7*	p6*	p5*	p7*	p5*	p5*	p7*	p5*	p8*	h6	p5*	p5*	p6*	p5*	p5*	p8*	p6*	p7*	p5*	p5*	p8*	cr*	p8*	p6*	sf*	h6	p5*	p5*	p5*	cr*	p5*	p5*	cr*	p5*
hartmann	p5*	p5*	p6*	p5*	sf*	p5*	p6*	p5*	p6*	p6*	p7*	ob*	p5*	p7*	p6*	p5*	p5*	h6	p5*	p5*	p5*	p8*	cr*	p5*	p5*	p6*	p5*	p5*	p5	sf*	p6*	p7*	cr	p8*	p6*	p6*	p5*	p5*	p5*	p8*	p5*
highsmith	p5*	p5*	p8*	p5*	p6*	p5*	p5*	p7*	p8*	p5*	p5*	p5*	p6*	p5*	p7*	cr*	p5*	ob*	h5	p6*	h5*	p6*	p5*	p5*	p8*	p5*	p5*	p8*	p5*	p8*	p5*	p5*	h5*	p8*	p5*	p6*	cr*	p5*	p5*	p5*	ob*
hine	p5*	p5*	p6*	p6*	p5*	p5*	p6*	p6*	p6*	p5*	p7*	p5*	p5*	p8*	p5*	p5*	p6*	p5*	p5*	h5	p5*	p5*	p7*	p5*	p7*	p8*	p5*	p7*	p5*	p8*	p5*	p6*	h5*	p5*	p5*	p5*	ob*	p5*	p5*	p7*	p5*
horydczak	p5*	p8*	p5*	ob*	p6*	p8*	p5*	p7*	p5*	p5*	p5*	p6*	p5*	p8*	p5*	p5*	p5*	p5*	p6*	p5*	h5	p5*	p5*	cr*	p5*	p5*	p6*	p5*	p5*	p7*	p6*	cr*	h5*	cr*	p5*	p6*	p6*	p5*	p6*	p6*	p5*
hurley	p6*	p7*	sf*	p7*	p5*	p5*	p6*	p5*	p5*	p7*	p5*	p6*	p7*	p5*	p5*	p6*	p6*	sf*	p5*	p7*	p5*	h5	p5*	ob*	p7*	ob*	p5*	p5*	p5*	gt*	p7*	p7*	h7*	p8*	p5*	p6*	p7*	p7*	p5*	p6*	p7*
jackson	p5*	p6*	cr*	p5*	p5*	p6*	p5*	p7*	p5*	p5*	p5*	p5*	p5*	p6*	p5*	p5*	p5*	p6*	p5*	p5*	p5*	p5*	h5	p6*	p5*	p5*	sf*	ob*	p5*	p8*	p6*	p5*	ob*	p8*	p5*	p6*	p5*	p5*	p7*	p5*	p5*
johnston	p7*	p6*	p5*	p6*	p8*	p6*	p6*	p7*	cr*	p6*	p5*	p5*	p5*	p5*	p7*	p5*	p6*	p5*	h7*	gt*	p5*	p5*	p6*	h5	p7*	p5*	p5*	p5*	p6*	p7*	p5*	p5*	p6*	p6*	cr*	p6*	p7*	p5*	p8*	p6*	p6*
kandell	p6*	p5*	p7*	p5*	sf*	p6*	p5*	p8*	p6*	ob*	p8*	p5*	p8*	p5*	p7*	p6*	p6*	p6*	p6*	p5*	p6*	p6*	p5*	p6*	h5	p5*	p5*	p5*	cr*	p8*	p6*	p6*	h6*	p6*	p5*	p5*	p5*	p5*	cr*	p6*	
korab	p7*	sf*	h6*	p6*	p5*	p5*	p6*	p5*	p5*	p7*	p6*	p6*	p5*	p8*	p6*	p7*	p5*	p5*	p6*	p5*	p5*	p5*	p7*	p5*	p5*	h5	p5*	ob*	p5*	p7*	p5*	p5*	ob*	p6*	p6*	p5*	p5*	p5*	p5*	p8*	p6*
lange	p5*	p5*	p7*	p7*	p6*	p6*	p5*	p5*	c5	p5*	p5*	p6*	p8*	p5*	p8*	p6*	p5*	p5*	p5*	p6*	p6*	p5*	p5*	p5*	p5*	ob*	p6*	c5	p6*	p7*	sf*	p5*	p6*	p6*	p6*	p6*	p5	h5*	p8*	cr*	p5*	p5
list	h6*	p6*	cr*	p6*	cr*	p5*	p5*	p6*	p5*	p5*	ob*	p5*	p5*	p5*	p5*	p7*	p5*	p5*	p5*	p5*	p5*	p7*	p5*	p5*	p5*	p7*	p6*	p7	p6*	p7*	p5*	p5*	h6*	p5*	p6*	p5*	p5*	p5*	p7*	p5*	ob*
mccurry	p6*	p5*	p5*	p5*	p6*	p6*	h5*	p5*	p5*	p6*	p5*	p6*	p5*	p6*	p5*	p5*	p5*	p5*	p6*	p5*	p7*	p7*	p7*	p6*	p5*	p6*	p5*	h6	p8*	p5*	p7*	h6*	p5*	p6*	p5*	p5*	gt*	p5*	p6*	p6*	
meiselas	p6*	h5*	p5*	cr*	ob*	p6*	p5*	p5*	cr*	p5*	p7*	p7*	p6*	p5*	p5*	p5*	p5*	p5*	p5*	p5*	c5*	h5*	p5*	p6*	p5*	p5*	p5*	p5*	cr*	p6	p6*	p5*	cr	p5*	p5*	p6*	p6*	p6*	p8*	p6*	
mydans	p6*	p8*	p6*	p7*	p7*	p6*	p6*	p5*	ob*	p6*	p6*	p6*	p6*	p5*	p6*	cr*	p6*	p5*	p6*	p5*	c6*	p7*	p6*	p5*	ob*	p7*	p5*	p5*	p6*	ob*	h5	p5*	p7*	p5*	p8*	p5	p8*	p5*	p5*	p6*	p5
osullivan	p7*	p6*	p5*	p8*	p5*	p5*	p6*	sf*	p6*	p5*	cr*	p5*	p6*	p5*	p5*	p6*	p5*	p5*	cr*	p5*	p5	p5*	p6*	p5*	p5*	cr*	p7*	ob*	p6*	p7*	cr*	h5	p6*	cr*	p6*	p6*	p5*	p7*	p5*	p5*	p5*
parr	p6*	p5*	p5*	p6*	p6*	p7*	c5*	p5*	p6*	p5*	p5*	p5*	cr*	cr*	p5*	cr*	cr*	ob*	p5*	p6*	p6*	p6*	p5*	p6*	p5*	p5*	p5*	p7*	h7*	cr*	cr*	p5*	p6*	h6	p5*	p6*	ob*	p5*	p6*	p5*	p7*	p5*
prokudin-gorsky	p6*	p5*	p6*	p5*	p6*	p5*	p6*	p5*	p5*	p6*	cr*	cr*	p6*	p5*	p6*	p6*	p5*	p6*	p6*	p5*	p5*	p7*	p5*	p7*	p8*	p5*	p7*	p7*	cr*	p5*	p6*	h7*	ob	p5*	p5*	p5*	p5*	p6*	p7*	p5*	
rodger	p5*	p5*	p5*	p5*	p5*	p6*	p6*	p7*	p5*	p5*	p5*	p5*	p6*	p6*	p5*	p5*	p5*	p6*	p5*	p6*	p5*	p7*	p6*	p5*	p7*	p6*	p6*	p5*	cr*	p8*	p6*	p5*	cr	p8*	h7	p5*	p5*	p6*	p5*	p5*	p5*
rothstein	p5*	p7*	p5*	p5*	p5*	p6*	p6*	gt*	cr*	p5*	p6*	p5*	p6*	p5*	p5*	p5*	p5*	p7*	p5*	p7*	p5	p5*	p8*	p5	p5*	p5*	p5*	p5*	p5*	cr*	ob*	p7*	h5*	p5*	p5*	h5	p7*	p5*	p6*	p6*	p5
seymour	p8*	p6*	cr*	cr*	p6*	cr*	p5*	p6*	p6*	p6*	p5*	p6*	p6*	p5*	p5*	p6*	p5*	p5*	h6*	p7*	h7*	p6*	p6*	p5*	p5*	p8*	p6*	p8*	p5*	p7*	p5*	p6*	cr	p5*	p6*	p5*	h5	p5*	p5*	p5*	p5*
stock	ob*	p8*	p6*	cr*	p5*	p7*	p5*	p6*	p5*	p5*	p8*	p5*	p6*	p6*	p5*	p5*	cr*	p5*	p5*	p5*	p6*	p5*	p6*	p5*	p5*	p8*	p5*	cr*	cr*	p8*	p5*	p5*	c6	p5*	p6*	p5*	p5*	h6	p6*	ob*	p6*
sweet	p6*	p7*	p5*	p5*	p5*	p5*	p8*	gt*	p5*	p7*	p6*	p5*	p6*	p7*	p5*	p5*	p5*	p6*	h7*	sf*	p5*	p7*	p6*	gt*	p7*	p5*	p5*	p8*	p5*	p8*	p7*	p6*	gt*	cr*	p6*	p6*	p5*	cr*	h6*	p5*	p5*
vanvechten	p6*	p6*	cr*	p5*	p5*	gt*	h6*	p6*	p5*	p5*	p6*	p5*	p8*	p7*	p7*	p5*	p6*	p6*	p6*	gt*	p5	p5*	p8*	cr*	p6*	p8*	p5*	sf*	ob*	p7*	p6*	p5*	h5*	p8*	p5*	p6*	p5*	p7*	p6*	h6*	ob*
wolcott	p7*	p5*	p5*	p5*	cr*	p7*	p5*	p5*	p5*	p5*	p8*	p6*	p8*	p8*	p5*	p8*	ob*	p7*	p6*	p5*	p5	p5*	p6*	p5*	p5*	p7*	p5*	p5*	p6*	p7*	p5*	p6*	p7*	p8*	p5*	p5	c5*	p5*	p6*	p5*	h5

Figure 5: Best feature at distinguishing each pair of photographers

7 Misclassifications With Best Feature

Even the best feature for each pair of photographers makes some mistakes because of the difficulty of the problem of photographer authorship attribution. We illustrate some of those misclassifications here. The image on the left of each set of three is the test image. The image in the middle is the photograph from the class which the SVM misclassified it as, which is closest to the test image (according to the best feature, which is indicated). The image on the right is the closest photograph from the correct class according to the best feature. The similarity of the images demonstrates how challenging this problem is. Note that the feature used in each example as the best for each pair of photographers may not appear in the table above for that pair of photographers because of ties (denoted by * in Figure 5).

Bresson | Parr-HPool5 | Bresson-HPool5

Bresson | Parr-HPool5 | Bresson-HPool5

Halsman | Parr-HFC6 | Halsman-HFC6

Halsman

Parr-HFC6

Halsman-HFC6

McCurry

Parr-HFC6

McCurry-HFC6

Stock

Parr-CFC6

Stock-CFC6

Delano

Lange-GIST

Delano-GIST

Cunningham

Johnston-PFC7

Cunningham-PFC7

8 New Photograph Generation

In this section, we present additional generation results generated using the procedure described in Section 6.3 of the main text. Here, we review this procedure again, in some more depth than the space in the main text allowed.

We learn probability distributions for each photographer over the 205 scene types from H-CNN. We downloaded new scene images to serve as backgrounds for our generated photographs from Flickr. We then chose 25 object types that were well represented across all photographers and trained a Fast-RCNN object detector to detect them using data from ImageNet. We ran this detector on our dataset and learned probability distributions for each object type, conditioned on the scene type (as determined from Hybrid-CNN). We also learned spatial distributions for each object type class for each photographer. These distributions allow us to choose scenes, objects, and their locations in a manner similar to our photographers. To create a new image for a photographer, we first sample from the scene distribution to choose a scene background type. After the background is chosen, we choose up to 5 objects to appear in that scene from the photographer's object distribution for that scene type. The actual objects come from other photographs by the same photographer (using our Fast-RCNN detections). We perform salient object segmentation to perform background subtraction. We then place each object in the scene by using the object's learned spatial distribution for each photographer to probabilistically select a location for the object. We indicate the target photographer underneath each pastiche.

Wolcott

Delano

Hine

Highsmith

Hine

Horydczak

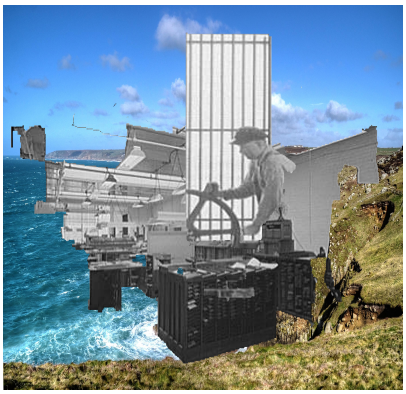
Parr

Wolcott

Hine

Hine

Horydczak



Horydczak

Delano

Erwitt

Highsmith

Erwitt

Hine

Highsmith

Erwitt Hine Delano

McCurry Wolcott Johnston

Hine Horydczak McCurry

Figure 11: Generated images

9 t-SNE Visualizations

We show results of projecting features from all three deep networks tested: CaffeNet, Hybrid-CNN, and PhotographerNET. We choose the best performing feature from each network to project here. We project the high-dimensional features to 2-D and then plot the photographs associated with each feature in the position of its 2-D projected feature. We make several interesting observations from these projections. Hybrid-CNN and CaffeNet do not appear to rely on lower-level image statistics and instead focus on image semantics, while PhotographerNET relies heavily on lower-level details like color. Additionally, while CaffeNet groups photographs mainly by objects, Hybrid-CNN groups by both objects and scene type. More details are below.

Figure 12: CaffeNet Pool5 t-SNE Result. Notice that photographs with similar *semantic* content cluster together. For example, the bottom right contains people and the top left contains buildings. Color and black and white images are mixed throughout.

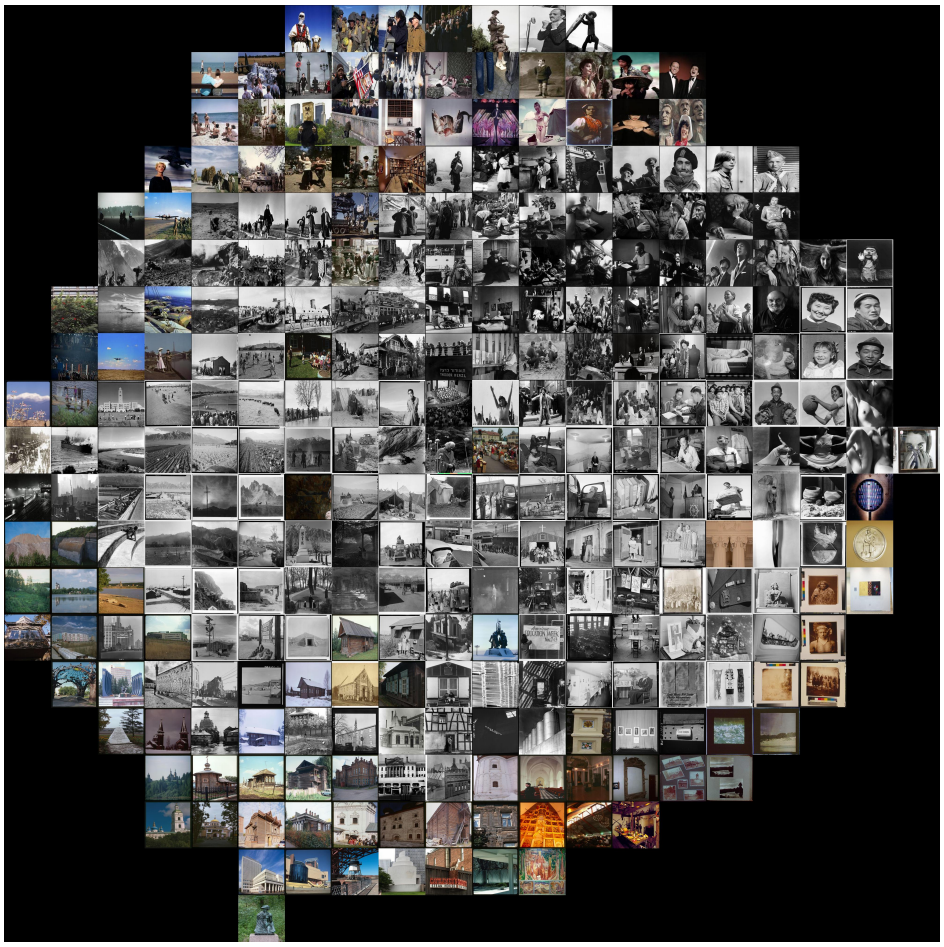


Figure 13: Hybrid-CNN Pool5 t-SNE Result. The result is similar to CaffeNet in that photographs with similar *semantics* are closer together in the projection. However, in addition to objects, Hybrid-CNN also groups photos on the type of scene. For example, outdoor photographs are closer together than indoor.

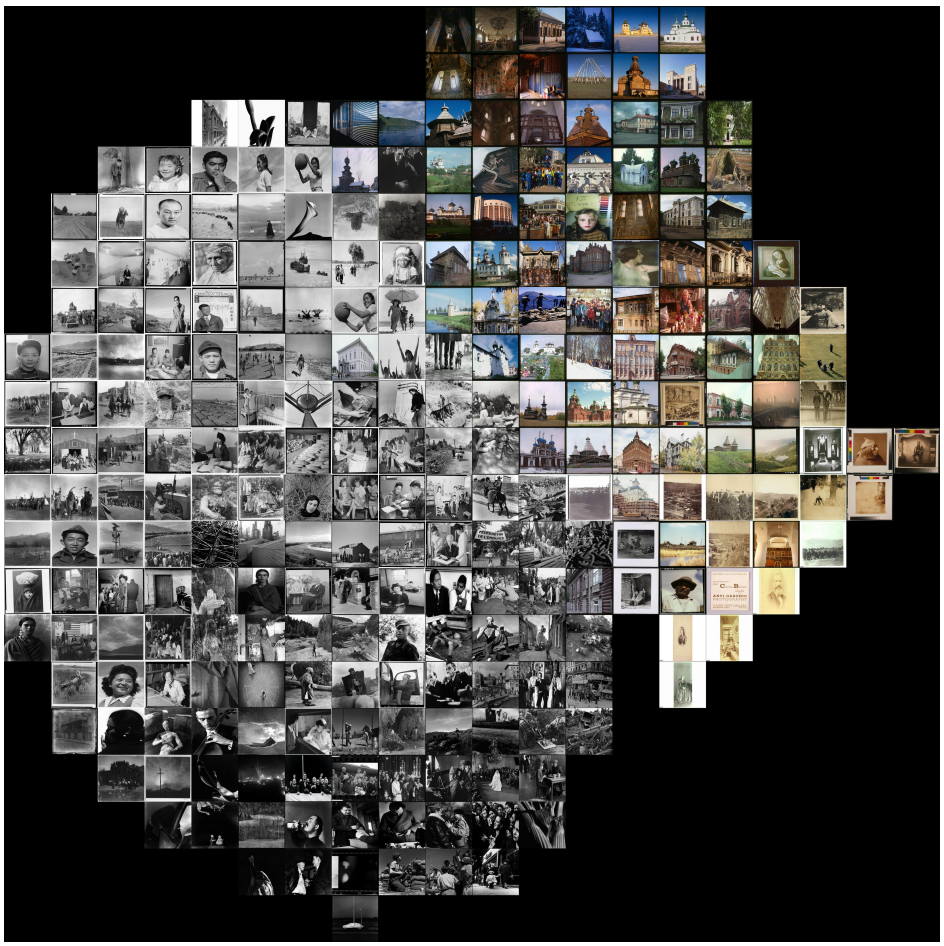


Figure 14: PhotographerNET FC7 t-SNE Result. We observe that PhotographerNET divides the image space by low-level details rather than semantics. We observe that black and white images form their own cluster on the left while color images appear at the top right. Images with similar colors or borders occur close together. For example, the top right contains images which are mostly blue. This indicates that PhotographerNET relies more heavily on lower-level details than the other networks tested.