

Recognizing Image Style: Supplementary Materials

Sergey Karayev¹

Matthew Trentacoste²

Helen Han¹

Aseem Agarwala²

Trevor Darrell¹

Aaron Hertzmann²

Holger Winnemoeller²

¹ University of California, Berkeley

² Adobe

List of Figures

1	Top five most confident predictions on the Flickr Style test set: styles 1-8. . .	3
2	Top five most confident predictions on the Flickr Style test set: styles 9-14. . .	4
3	Top five most confident predictions on the Flickr Style test set: styles 15-20. . .	5
4	Distribution of image style, genre, and date in the Wikipaintings dataset. . .	7
5	Confusion matrix of our best classifier on the AVA Style dataset.	12
6	Confusion matrix of our best classifier on the Flickr dataset.	13
7	Confusion matrix of our best classifier on the Wikipaintings dataset.	14

List of Tables

1	Exact Flickr group names, and their sizes.	2
2	All per-class APs on all evaluated features on the Flickr dataset.	6
3	All per-class APs on all evaluated features on the Wikipaintings dataset. . .	8
4	All per-class APs on all evaluated features on the AVA Style dataset.	9
5	Comparison of Flickr Style per-class accuracies for our method and Mech Turkers.	10
6	Per-class accuracies on the Wikipaintings dataset, using the MC-bit feature. . .	11

1 MTurk Study Details

Test images were grouped into 10 images per Human Interface Task (HIT). Each task asks the Turker to evaluate the style (e.g., “Is this image VINTAGE?”) for each image. For each style, we provided a short blurb describing the style in words, and provided 12-15 hand-chosen positive and negative examples for each Flickr Group. Each HIT included 2 sentinels:

Table 1: Exact Flickr group names, and their sizes.

Style	Group names [num images]
Bokeh	Bokeh Photography (1/day) [187K]
Bright	Colour Mania [100K]
Depth of Field	Depth of Field [116K], Finest DoF [54K]
Detailed	Details aller Art - Details of all kind [22K], Detail pictures [5K]
Ethereal	Ethereal World [21K]
Geometric Composition	Geometric Beauty [168K]
Hazy	Misty hazy smokey [14K]
HDR	HDR ADDICTED [374K]
Horror	Horror [16K]
Long Exposure	Long Exposure [619K]
Macro	Closer and Closer Macro Photography [990K]
Melancholy	melancholy [106K]
Minimal	Less Is More... [44K]
Noir	Film Noir Mood [7K]
Romantic	Romantic Images [20K]
Serene	Šerene [68K]
Pastel	pastel and dreamy [120K], Pastel Soft tone [7K]
Sunny	Sun, sun and more sun [23K]
Texture	Texture [103K]
Vintage	Vintage Feelings [4K], Vintage & Retro [61K]

images which were very clearly positives and similar to the examples. HITs were rejected when Turkers got both sentinels wrong. Turkers were paid 0.10 per HIT, and were allowed to perform multiple hits. Manual inspection of the results indicate that the Turkers understood the task and were performing effectively. A few Turkers sent unsolicited feedback indicating that they were really enjoying the HITs (“some of the photos are beautiful”) and wanted to perform them as effectively as possible.

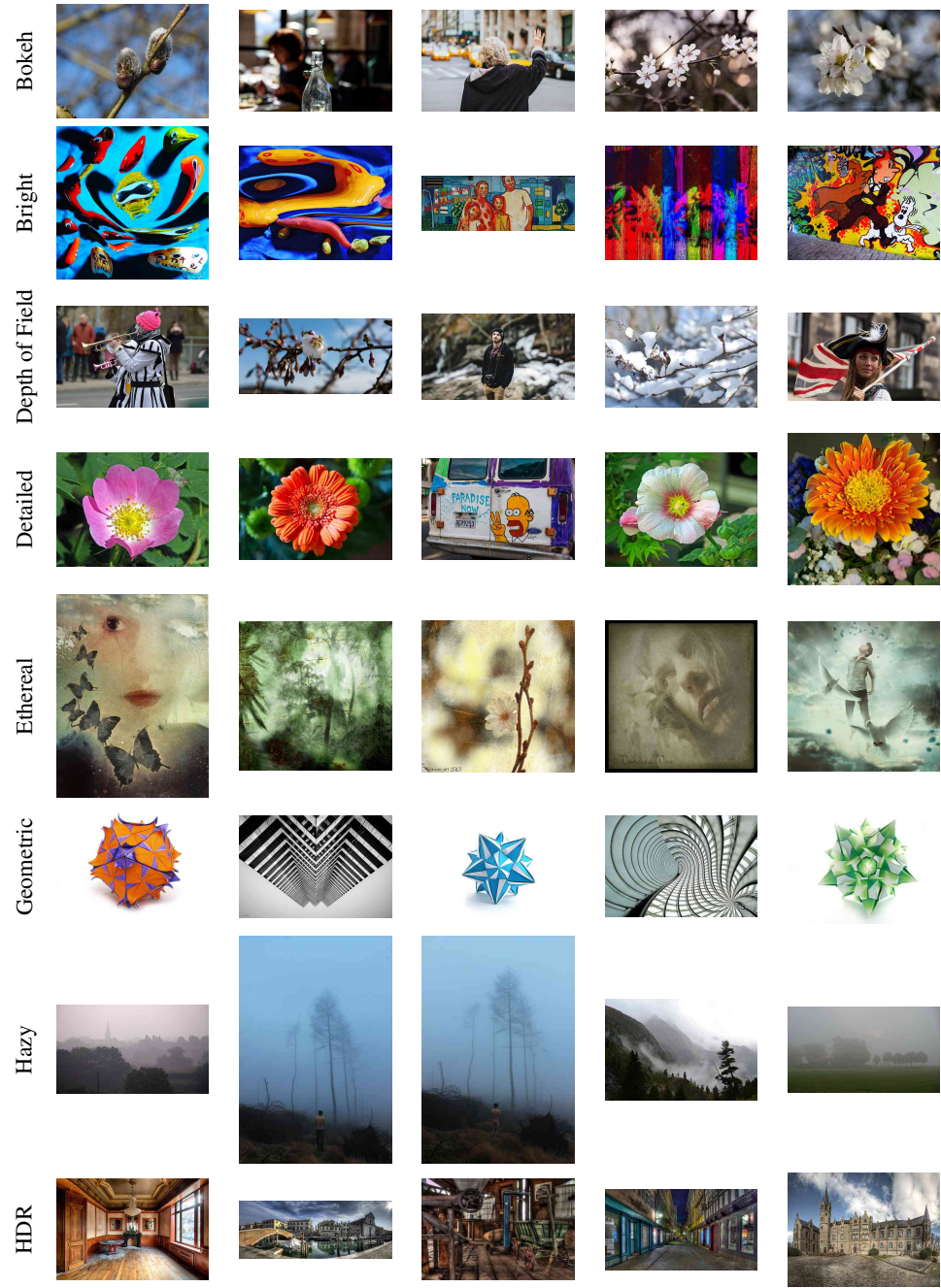


Figure 1: Top five most confident predictions on the Flickr Style test set: styles 1-8.

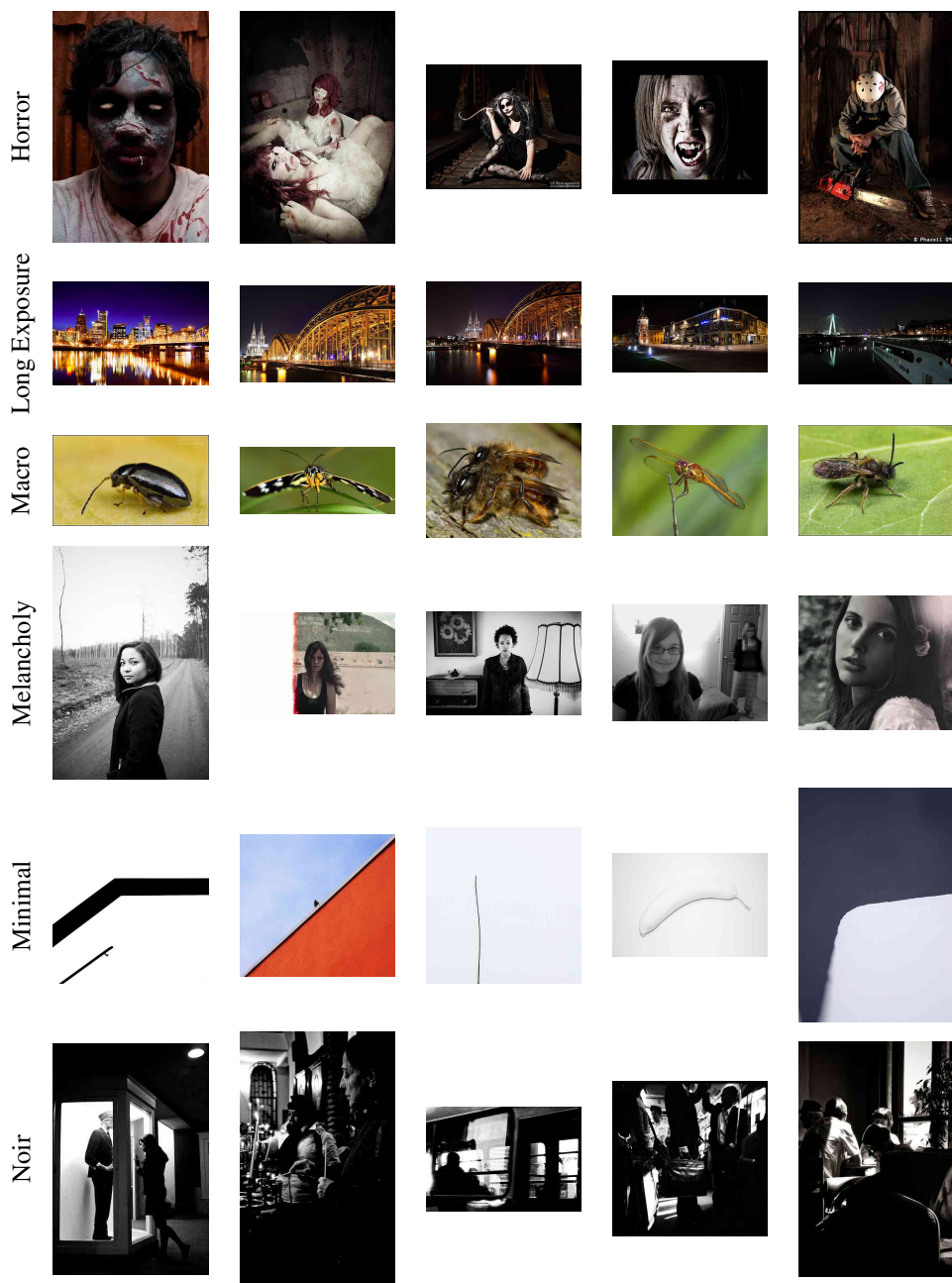


Figure 2: Top five most confident predictions on the Flickr Style test set: styles 9-14.

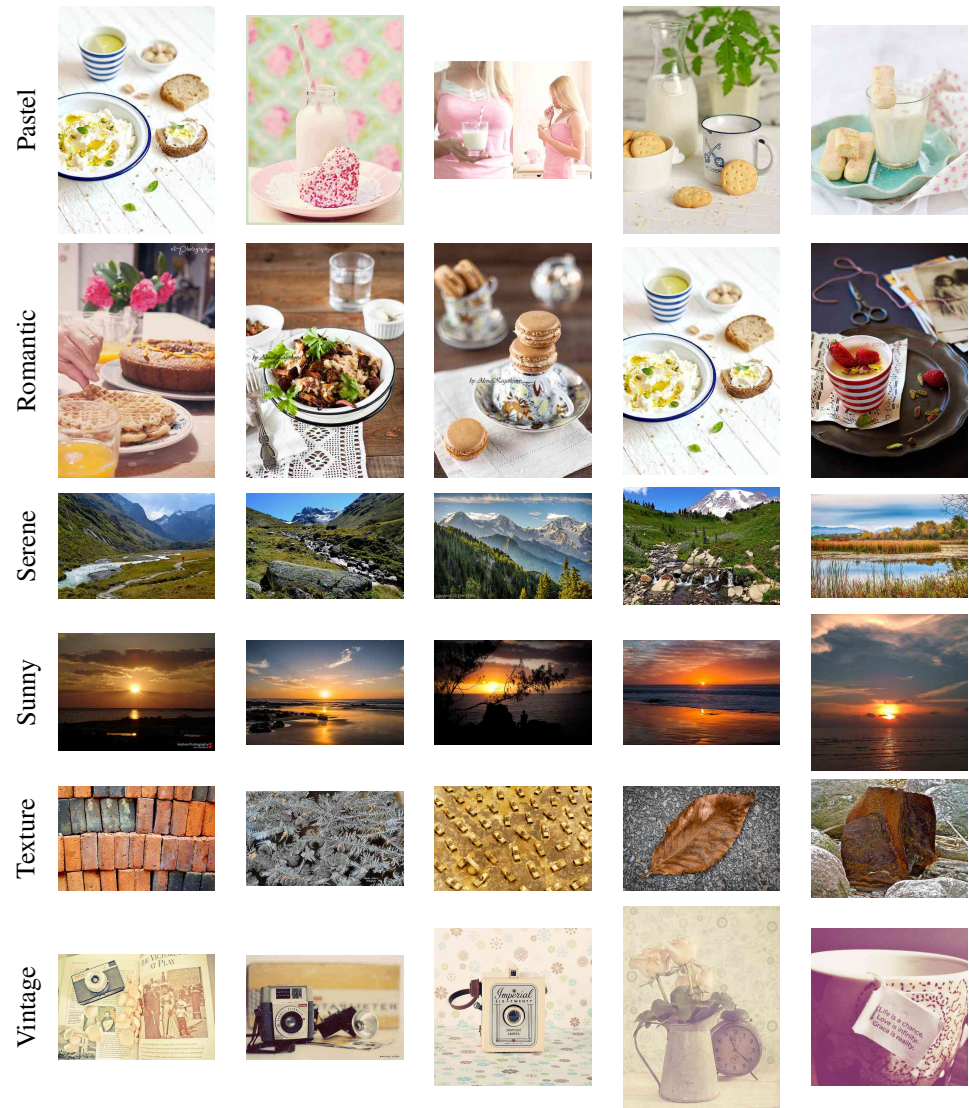


Figure 3: Top five most confident predictions on the Flickr Style test set: styles 15-20.

Table 2: All per-class APs on all evaluated features on the Flickr dataset.

	Fusion x Content	Caffe FC6	MC binary
Bokeh	0.288	0.253	0.248
Bright	0.251	0.236	0.183
Depth_of_Field	0.169	0.152	0.148
Detailed	0.337	0.277	0.278
Ethereal	0.408	0.393	0.335
Geometric_Composition	0.411	0.355	0.360
HDR	0.487	0.406	0.475
Hazy	0.493	0.451	0.447
Horror	0.400	0.396	0.295
Long_Exposure	0.515	0.457	0.463
Macro	0.617	0.582	0.530
Melancholy	0.168	0.147	0.136
Minimal	0.512	0.444	0.481
Noir	0.494	0.481	0.408
Pastel	0.258	0.245	0.211
Romantic	0.227	0.204	0.185
Serene	0.281	0.257	0.239
Sunny	0.500	0.481	0.453
Texture	0.265	0.227	0.229
Vintage	0.282	0.273	0.222
mean	0.368	0.336	0.316

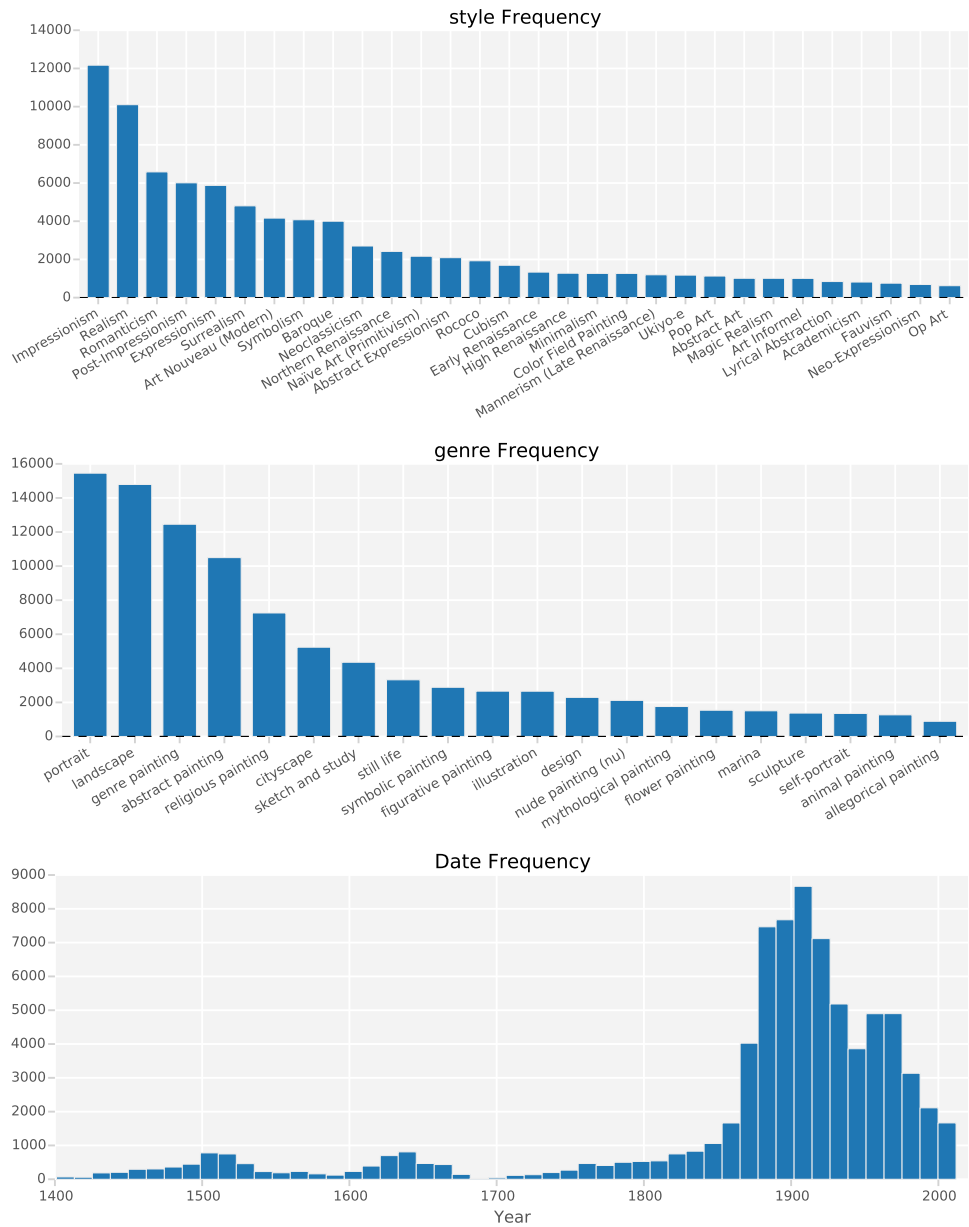


Figure 4: Distribution of image style, genre, and date in the Wikipaintings dataset.

Table 3: All per-class APs on all evaluated features on the Wikipaintings dataset.

	Fusion x Content	MC-bit	DeCAF ₆
Abstract_Art	0.341	0.314	0.258
Abstract_Expressionism	0.351	0.340	0.243
Art_Informel	0.221	0.217	0.187
Art_Nouveau_(Modern)	0.421	0.402	0.197
Baroque	0.436	0.386	0.313
Color_Field_Painting	0.773	0.739	0.689
Cubism	0.495	0.488	0.400
Early_Renaissance	0.578	0.559	0.453
Expressionism	0.235	0.230	0.186
High_Renaissance	0.401	0.345	0.288
Impressionism	0.586	0.528	0.411
Magic_Realism	0.521	0.465	0.428
Mannerism_(Late_Renaissance)	0.505	0.439	0.356
Minimalism	0.660	0.614	0.604
Nave_Art_(Primitivism)	0.395	0.425	0.225
Neoclassicism	0.601	0.537	0.399
Northern_Renaissance	0.560	0.478	0.433
Pop_Art	0.441	0.398	0.281
Post-Impressionism	0.348	0.348	0.292
Realism	0.408	0.309	0.266
Rococo	0.616	0.548	0.467
Romanticism	0.392	0.389	0.343
Surrealism	0.262	0.247	0.134
Symbolism	0.390	0.390	0.260
Ukiyo-e	0.895	0.894	0.788
mean	0.473	0.441	0.356

Table 4: All per-class APs on all evaluated features on the AVA Style dataset.

	Fusion	DeCAF ₆	MC-bit	Murray	L*a*b*	GIST	Saliency
Complementary_Colors	0.469	0.548	0.329	0.440	0.294	0.223	0.111
Duotones	0.676	0.737	0.612	0.510	0.582	0.255	0.233
HDR	0.669	0.594	0.624	0.640	0.194	0.124	0.101
Image_Grain	0.647	0.545	0.744	0.740	0.213	0.104	0.104
Light_On_White	0.908	0.915	0.802	0.730	0.867	0.704	0.172
Long_Exposure	0.453	0.431	0.420	0.430	0.232	0.159	0.147
Macro	0.478	0.427	0.413	0.500	0.230	0.269	0.161
Motion_Blur	0.478	0.467	0.458	0.400	0.117	0.114	0.122
Negative_Image	0.595	0.619	0.499	0.690	0.268	0.189	0.123
Rule_of_Thirds	0.352	0.353	0.236	0.300	0.188	0.167	0.228
Shallow_DOF	0.624	0.659	0.637	0.480	0.332	0.276	0.223
Silhouettes	0.791	0.801	0.801	0.720	0.261	0.263	0.130
Soft_Focus	0.312	0.354	0.290	0.390	0.127	0.126	0.114
Vanishing_Point	0.684	0.658	0.685	0.570	0.123	0.107	0.161
mean	0.581	0.579	0.539	0.539	0.288	0.220	0.152

Table 5: Comparison of Flickr Style per-class accuracies for our method and Mech Turkers. We first give the full results table, then show the significant deviations between human and machine performance, and between using Flickr and MTurk ground truth.

	MTurk acc., Flickr g.t.	Our acc., Flickr g.t.	Our acc., MTurk g.t.
Bright	69.10	73.38	73.63
Depth of Field	68.92	68.50	81.05
Detailed	65.47	75.25	68.44
Ethereal	76.92	80.62	77.95
Geometric Composition	81.52	77.75	80.31
HDR	71.84	82.00	76.96
Hazy	83.49	80.75	81.64
Horror	89.85	84.25	81.64
Long Exposure	73.12	84.19	76.79
Macro	92.25	86.56	88.39
Melancholy	67.77	70.88	71.25
Minimal	79.71	83.75	78.57
Noir	81.35	85.25	85.88
Pastel	66.94	74.56	75.47
Romantic	60.91	68.00	66.25
Serene	69.49	70.44	76.80
Sunny	84.48	84.56	79.94
Vintage	68.77	75.50	67.80
Mean	75.11	78.12	77.15

	Our acc., Flickr g.t.	Our acc., MTurk g.t.	% change from Flickr to MTurk g.t.
Vintage	75.50	67.80	-10.19
Detailed	75.25	68.44	-9.05
Long Exposure	84.19	76.79	-8.79
Minimal	83.75	78.57	-6.18
HDR	82.00	76.96	-6.15
Sunny	84.56	79.94	-5.46
Serene	70.44	76.80	9.03
Depth of Field	68.50	81.05	18.32

	Our acc., Flickr g.t.	MTurk acc., Flickr g.t.	Acc. difference
Horror	84.25	90.42	-6.17
Macro	86.56	91.71	-5.15
Romantic	68.00	61.04	6.96
Pastel	74.56	66.87	7.69
HDR	82.00	72.79	9.21
Long Exposure	84.19	73.83	10.35
Detailed	75.25	63.30	11.95

Table 6: Per-class accuracies on the Wikipaintings dataset, using the MC-bit feature.

Style	Accuracy	Style	Accuracy
Symbolism	71.24	Impressionism	82.15
Expressionism	72.03	Northern Renaissance	82.32
Art Nouveau (Modern)	72.77	High Renaissance	82.90
Nave Art (Primitivism)	72.95	Mannerism (Late Renaissance)	83.04
Surrealism	74.44	Pop Art	83.33
Post-Impressionism	74.51	Early Renaissance	84.69
Romanticism	75.86	Abstract Art	85.10
Realism	75.88	Cubism	86.85
Magic Realism	78.54	Rococo	87.33
Neoclassicism	80.18	Ukiyo-e	93.18
Abstract Expressionism	81.25	Minimalism	94.21
Baroque	81.45	Color Field Painting	95.58
Art Informel	82.09		

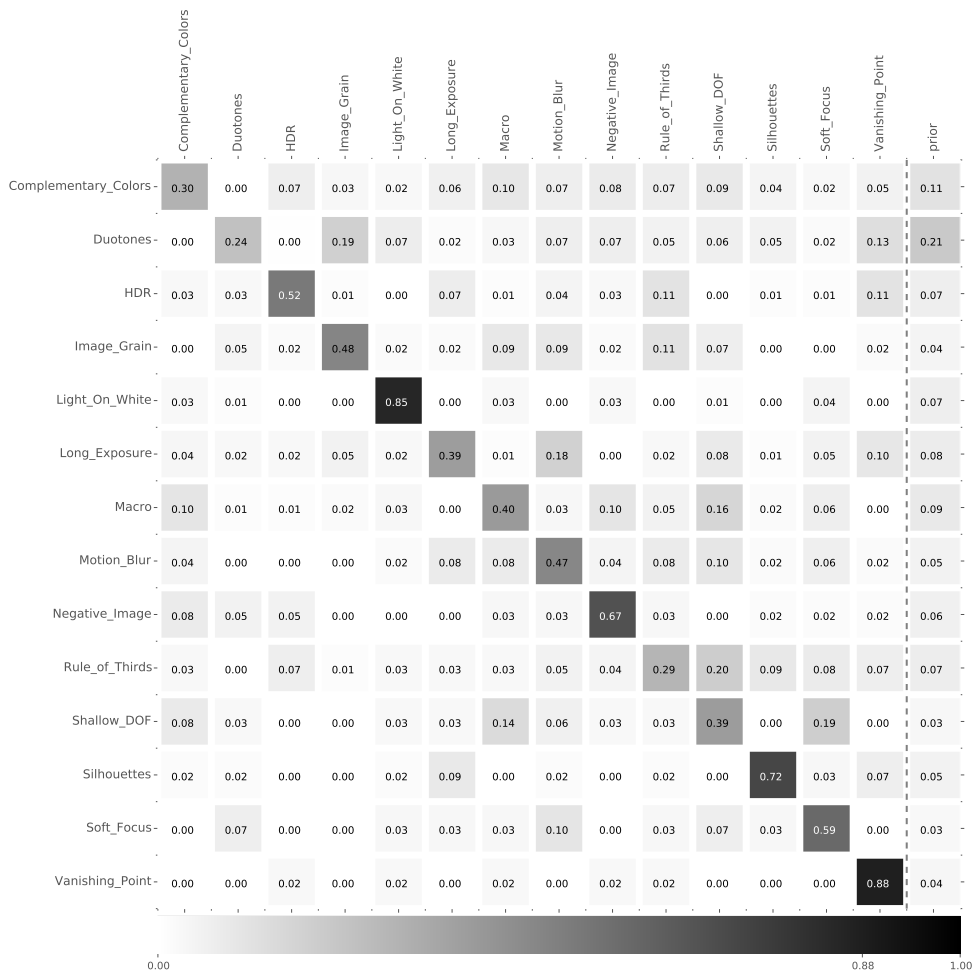


Figure 5: Confusion matrix of our best classifier (Late-fusion x Content) on the AVA Style dataset. The right-most “prior” column reflects the distribution of ground-truth labels in the test set. The confusions are mostly understandable: “Soft Focus” vs. “Depth of Field” for example.

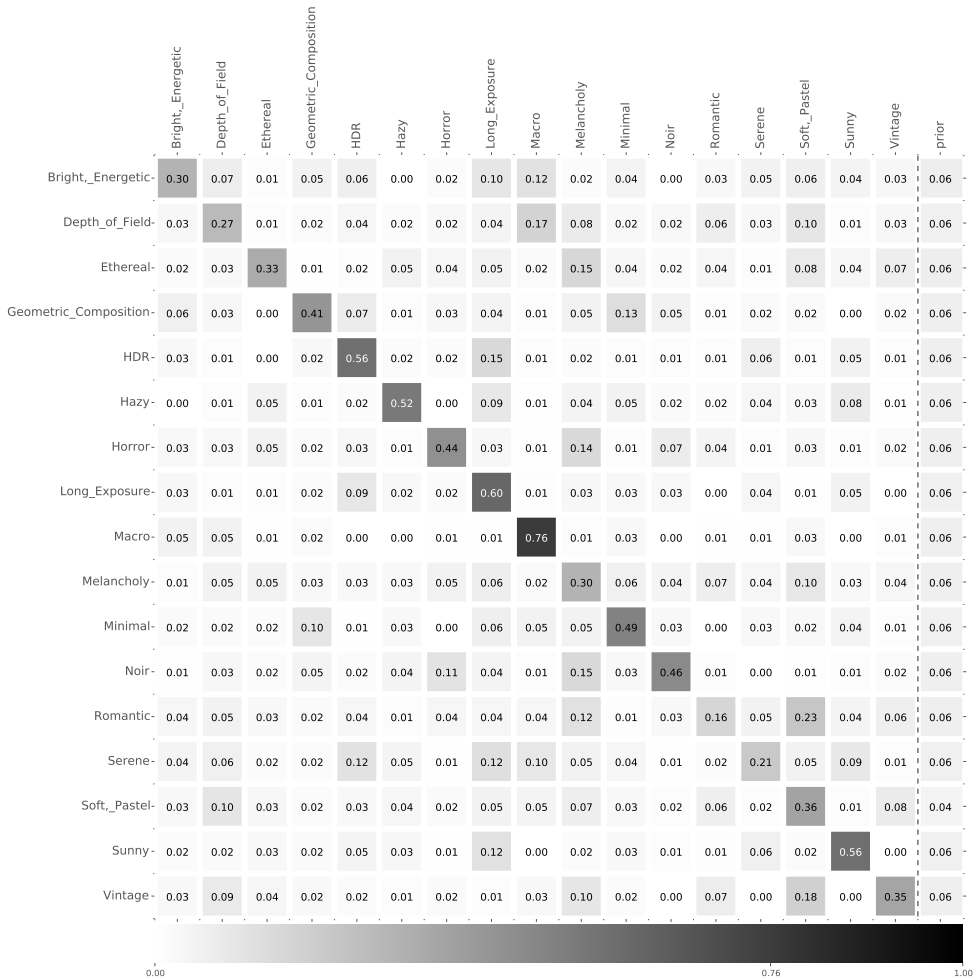


Figure 6: Confusion matrix of our best classifier (Late-fusion × Content) on the Flickr dataset.

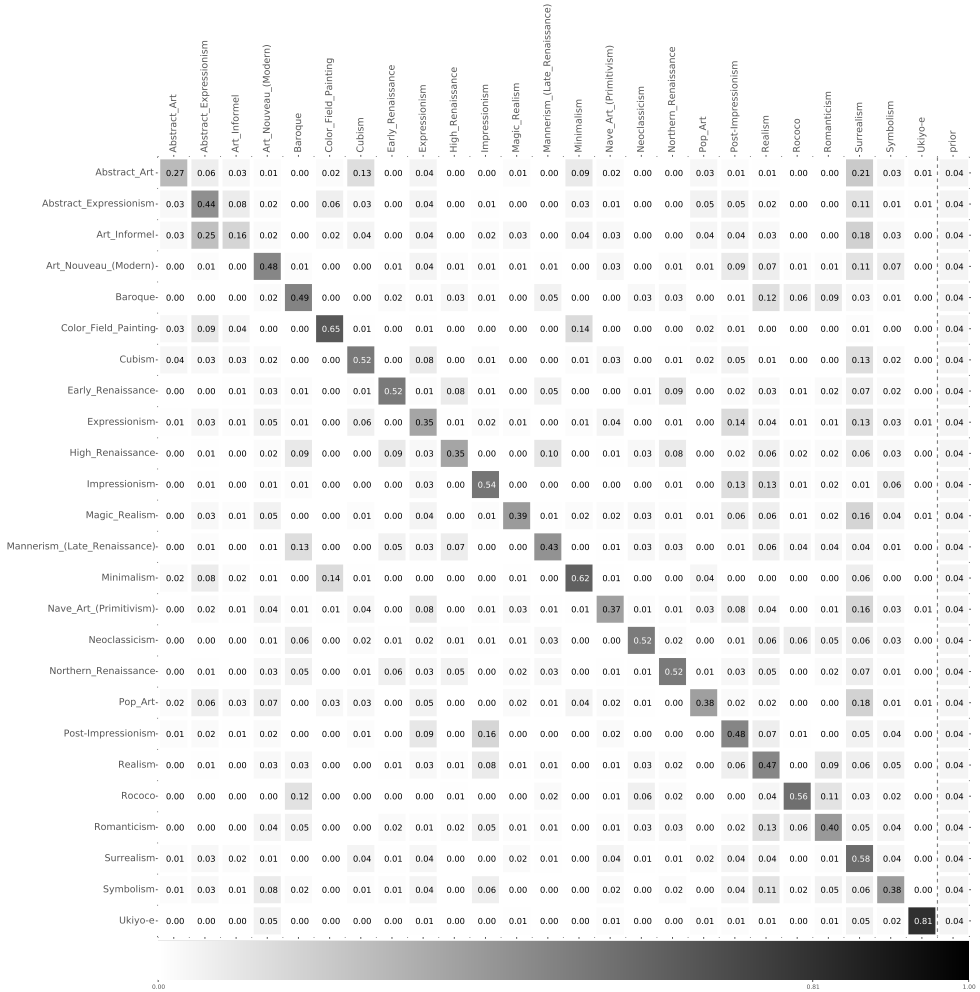


Figure 7: Confusion matrix of our best classifier (Late-fusion \times Content) on the Wikipaintings dataset.