Trinity College Dublin

The University of Dublin

V-SENSE

A Geometry-Sensitive Representation for Photographic Style Classification

Koustav Ghosal, Mukta Prasad, Aljosa Smolic

IMVIP 2018, Belfast

Aesthetic vs Physical Properties

Physical Properties	Building, Windows, Logo	Man, Woman, Street	Glass, Lights, Room
			
Aesthetic Properties	Good Symmetry, Soft Colours	Black and White, High Contrast	Good Bokeh, Complementary Colours

Aesthetic properties are less quantifiable, subjective properties and hence harder to be modelled.

Objective

V-SENSE

Automatic Photographic Style Prediction

Submit

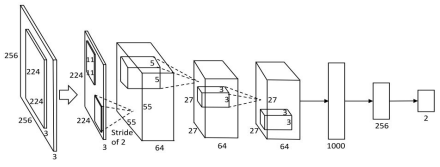
Style	Probability
Complementary_Colors	0.998
Macro	0.963
Rule_of_Thirds	0.893
Shallow_DOF	0.728
Silhouettes	0.711
Negative_Image	0.593
Soft_Focus	0.576
HDR	0.436
Motion_Blur	0.358
Long_Exposure	0.262
Light_On_White	0.210
Vanishing_Point	0.062
Image_Grain	0.055
Duotones	0.008

https://github.com/V-Sense/A-Geometry-Sensitive-Approach-for-Photographic-Style-Classification.git

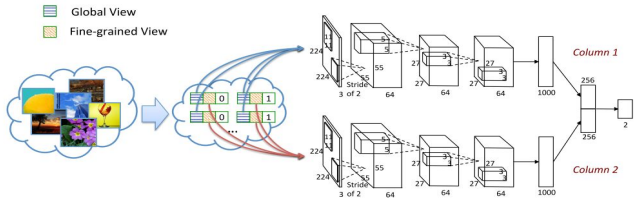
Traditional Approach

State of the art : Train a CNN on a photographic dataset.

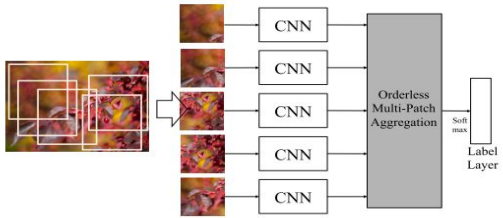
RAPID 1 Column [1]



RAPID 2 Columns [1]



Multipatch [2]



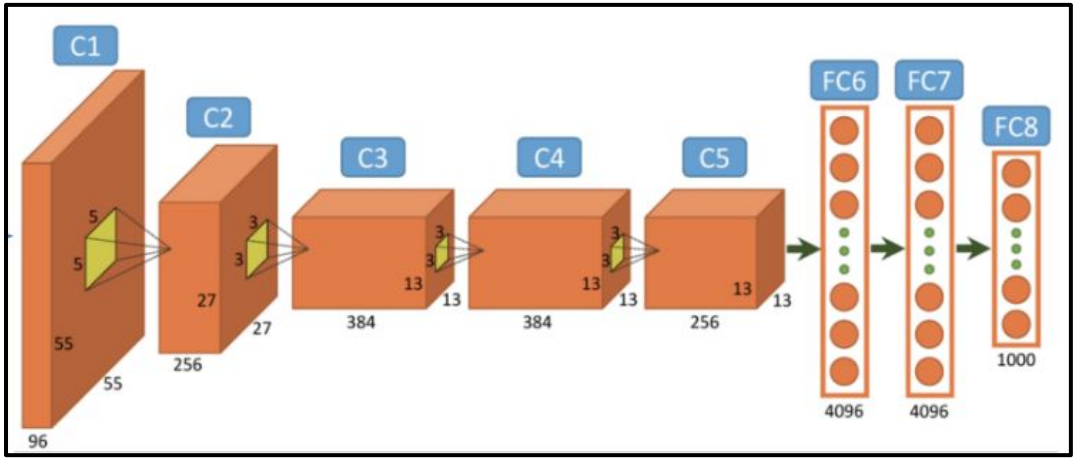
Local

Local + Global

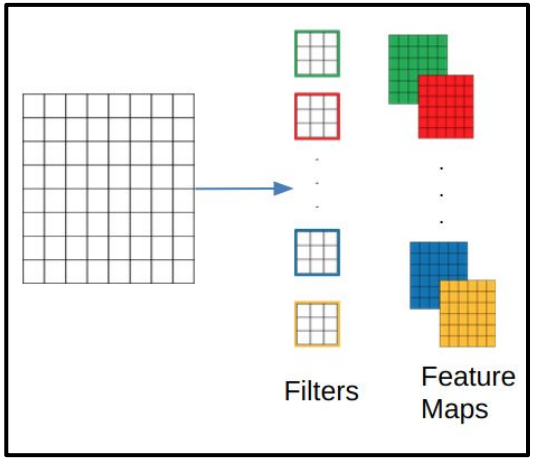
Local + Global + Randomness

1. Lu, X., Lin, Z., Jin, H., Yang, J., & Wang, J. Z. (2014, November). **Rapid: Rating pictorial aesthetics using deep learning**. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 457-466). ACM.
2. Lu, X., Lin, Z., Shen, X., Mech, R., & Wang, J. Z. (2015). **Deep multi-patch aggregation network for image style, aesthetics, and quality estimation**. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 990-998).

Standard Convolutional Filters



A standard CNN

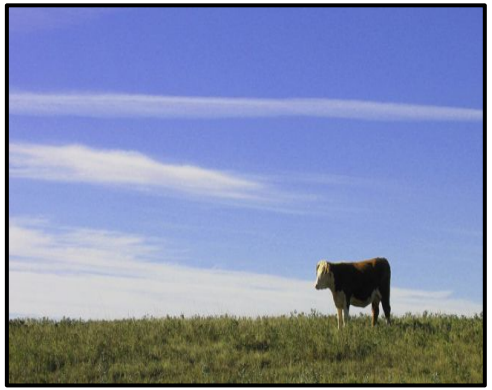


Parameter sharing

'translation invariant',
'appearance cognizant

Image courtesy : **https://www.saagie.com/blog/**

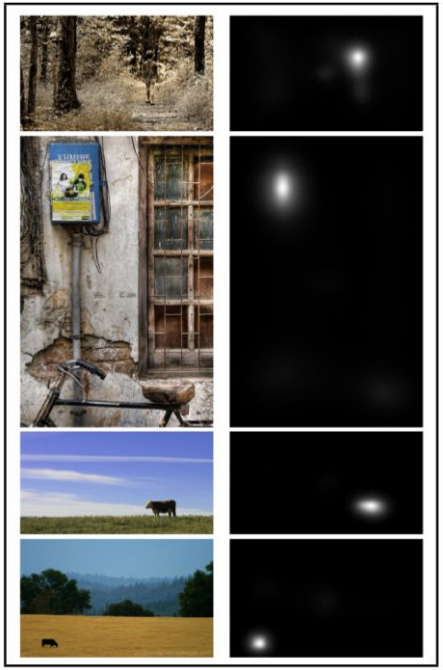
Problems

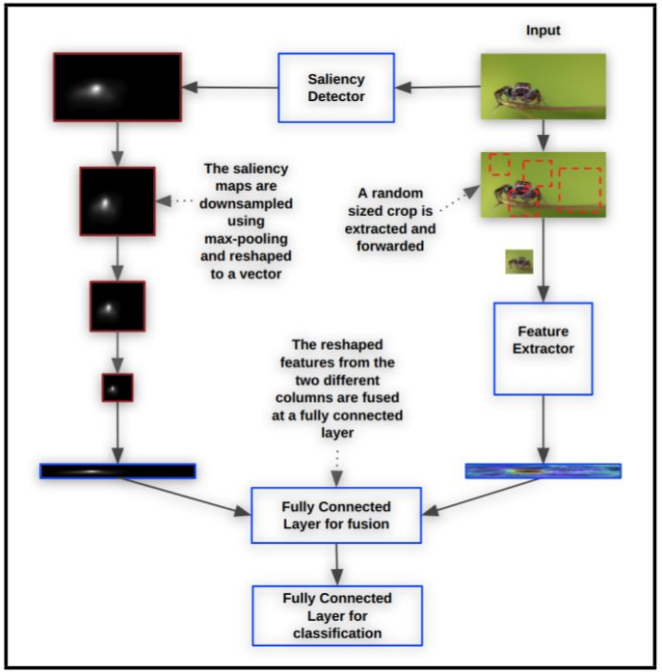
- Diverse Aspect Ratio
- Position of the main subject is crucial for framing

Image courtesy : Flickr

Approach



Saliency Maps



Network

Sal-RGB = Saliency Maps [Cornia *et al.*] + RGB-based features.

'location cognizant', 'appearance invariant'

Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. 2016. **Predicting human eye fixations via an LSTM-based saliency attentive model**. arXiv preprint arXiv:1611.09571 (2016)

Datasets

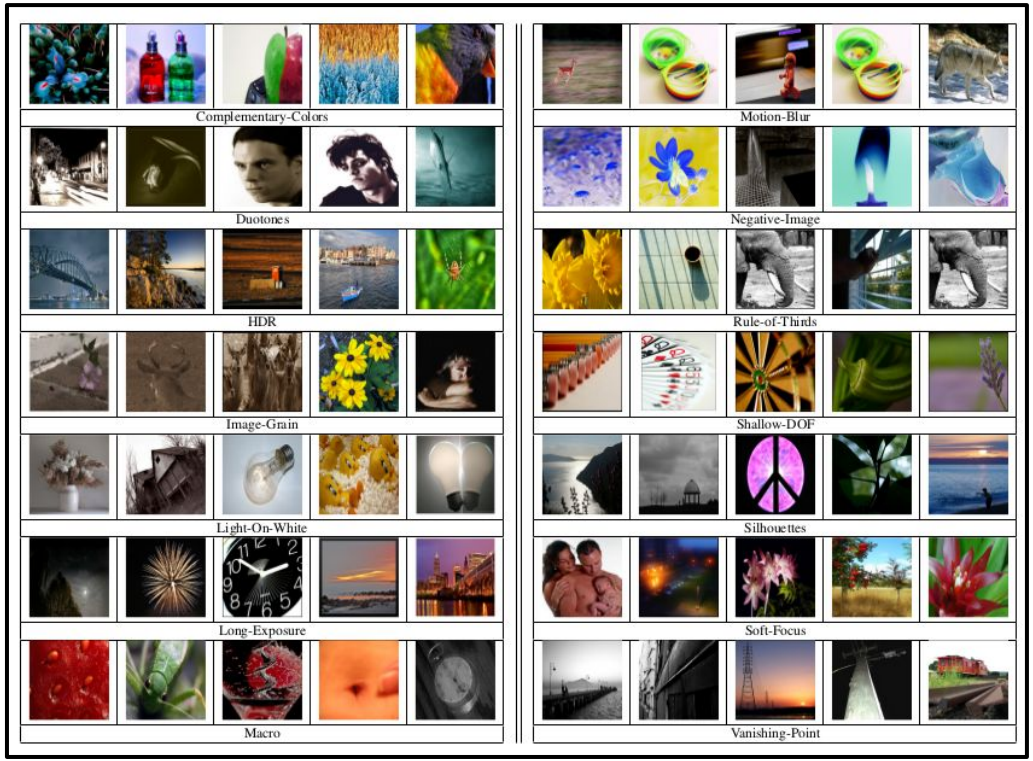
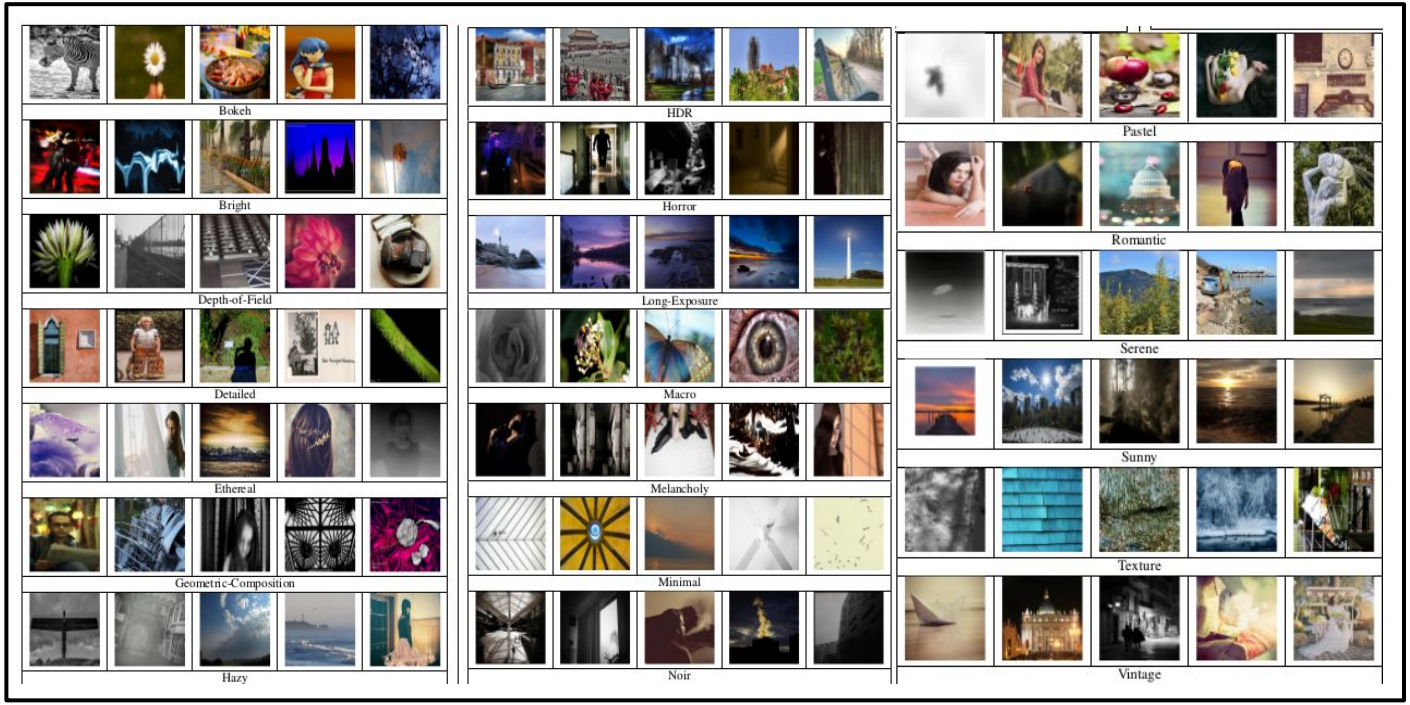


Image courtesy : **AVA Style dataset**

AVA Style

- 14 categories
- ~14000 Images
- Multi-labelled test data
- Highly imbalanced training data



Flickr Style

- 20 categories
- ~80000 images
- Complex Classes

Image courtesy : **Flickr Style dataset**

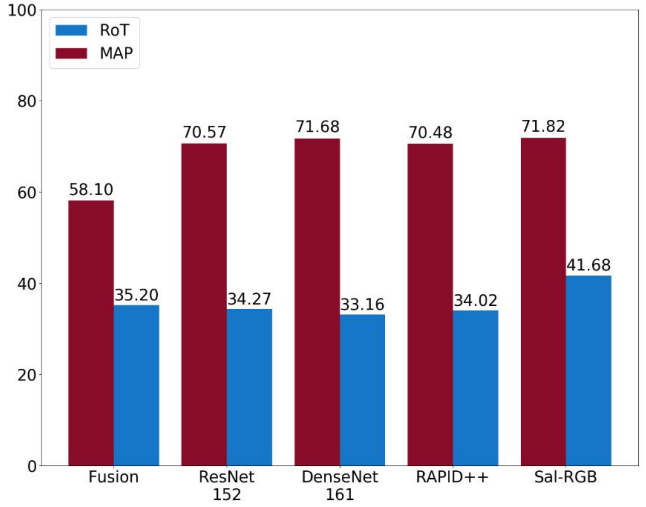
Results : Mean Average Precision (MAP)

Algorithm	Network	Augmentation	AVA	Flickr Style
State of the art	Fusion [Karayev et al., 2014]	centre crop	58.10	36.80
	RAPID [Lu et al., 2014]	random crop, warp	56.81	-
	Multi-Patch [Lu et al., 2015]	random crop	64.07	-
Our baselines	DenseNet161 [Huang et al., 2016]	random crop	71.68	43.83
	ResNet152 [He et al., 2016]	random crop	70.57	43.65
	RAPID++	random crop, warp	70.48	41.93
Our method	**Sal-RGB**	random crop	71.82	43.45

- Lu, X., Lin, Z., Jin, H., Yang, J., & Wang, J. Z. (2014, November). **Rapid: Rating pictorial aesthetics using deep learning**. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 457-466). ACM.
- Lu, X., Lin, Z., Shen, X., Mech, R., & Wang, J. Z. (2015). **Deep multi-patch aggregation network for image style, aesthetics, and quality estimation**. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 990-998).
- Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., & Winnemoeller, H. (2013). **Recognizing image style**. *arXiv preprint arXiv:1311.3715*.
- K. He, X. Zhang, S. Ren, and J. Sun, "**Deep residual learning for image recognition**," arXiv:1512.03385, 2015
- G. Huang, Z. Liu, and K. Q. Weinberger. **Densely connected convolutional networks**. arXiv preprint arXiv:1608.06993, 2016

Results : Per Class Precision on AVA

Styles	Fusion (SoA) [Karayev et al. 2014]	Densenet161 [Huang et al. 2016a]	ResNet152 [He et al. 2016]	RAPID++	**Sal-RGB**
Complementary_Colors	0.469	62.33	62.15	61.49	61.41
Duotones	0.676	86.58	84.82	84.77	87.58
HDR	0.669	74.95	70.08	71.51	72.86
Image_Grain	0.647	81.55	79.48	83.15	82.20
Light_On_White	0.908	84.69	83.41	85.64	82.99
Long_Exposure	0.453	64.16	65.38	63.94	61.94
Macro	0.478	64.89	65.52	64.90	66.58
Motion_Blur	0.478	63.93	62.12	61.21	61.98
Negative_Image	0.595	87.40	86.11	82.01	87.71
Rule_of_Thirds	**0.352**	**33.16**	**34.27**	**34.02**	**41.68**
Shallow_DOF	0.624	82.08	82.42	82.95	82.39
Silhouettes	0.791	93.73	92.49	91.14	93.05
Soft_Focus	0.312	49.89	44.91	44.57	46.41
Vanishing_Point	0.684	74.16	74.80	75.45	76.76

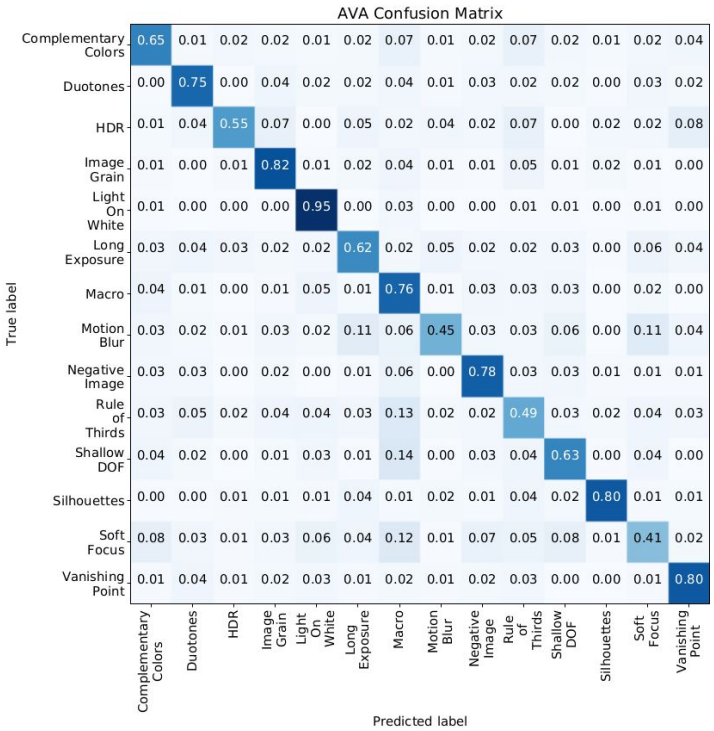


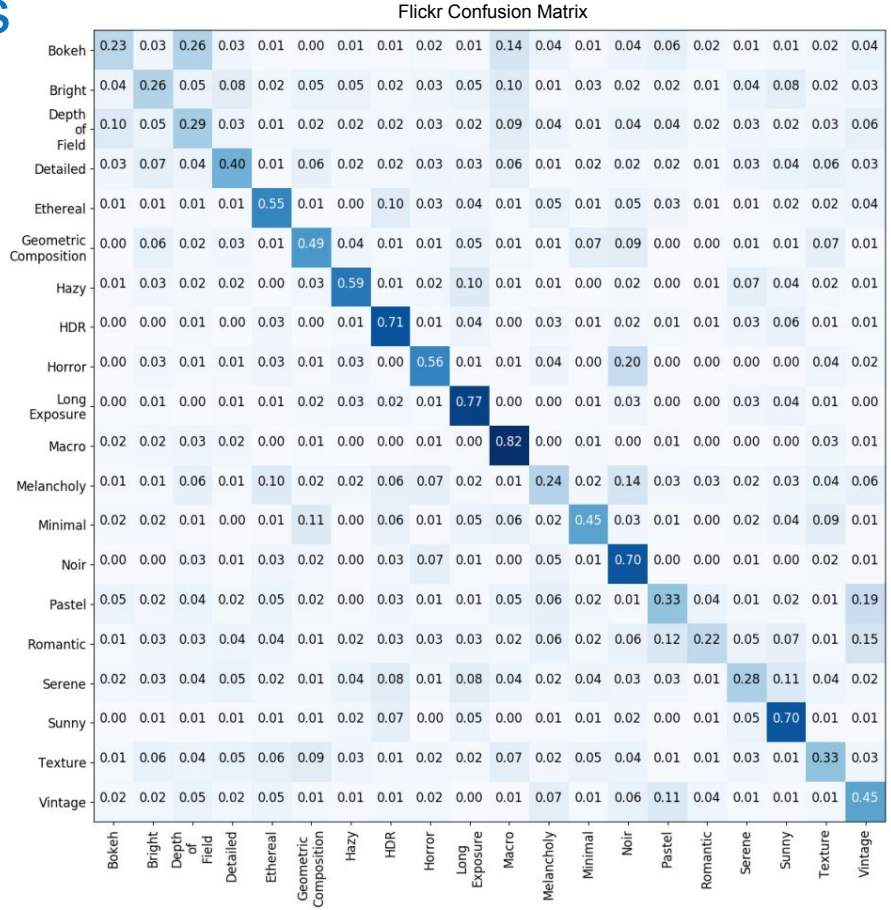
Rule of Thirds perform significantly better

Results : Per Class Precision on Flickr

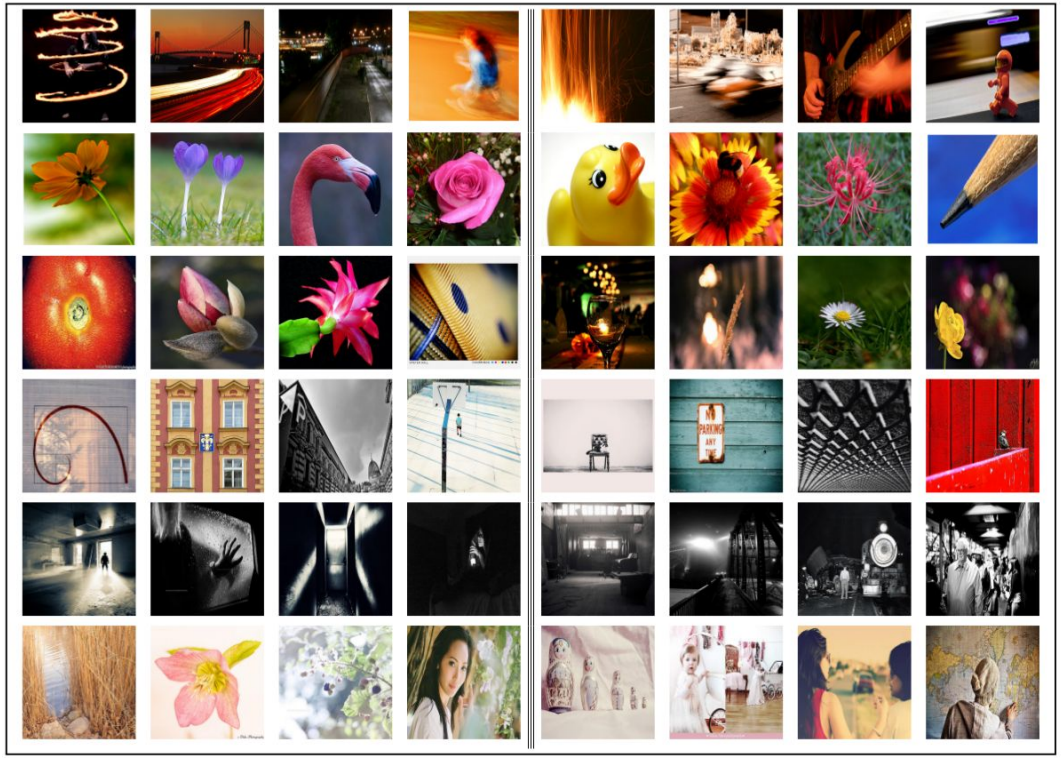
Styles	Fusion (SoA) [Karayev et al. 2014]	Densenet161 [Huang et al. 2016a]	ResNet152 [He et al. 2016]	RAPID++	**Sal-RGB**
Bokeh	28.80	30.24	31.34	29.39	29.78
Bright	25.10	22.97	23.12	22.69	23.33
Depth_of_Field	16.90	18.24	17.28	16.19	17.91
Detailed	33.70	37.96	38.27	38.50	38.09
Ethereal	40.80	50.31	50.88	48.15	50.03
Geometric_Composition	41.10	47.56	47.57	45.47	47.83
Hazy	48.70	61.59	60.01	57.68	60.92
HDR	49.30	65.44	65.24	61.03	64.92
Horror	40.00	64.24	64.17	58.40	64.16
Long_Exposure	51.50	65.36	64.76	61.40	63.62
Macro	61.70	67.44	70.26	69.60	68.18
Melancholy	16.80	19.82	20.33	18.50	19.71
Minimal	51.20	45.78	46.22	46.18	45.34
Noir	49.40	58.40	57.27	54.69	57.86
Pastel	25.80	34.15	34.05	30.71	34.17
Romantic	22.70	30.13	25.15	25.76	28.62
Serene	28.10	30.41	30.04	30.04	29.80
Sunny	50.00	59.99	60.56	58.57	58.58
Texture	26.50	28.98	30.52	29.72	29.65
Vintage	28.20	37.60	36.02	35.97	36.55

Results : Common Confusions





Results : Common Confusions



Long Exposure	Motion Blur
Shallow DOF	Macro
Shallow DOF	Bokeh
Geometric Composition	Minimal
Horror	Noir
Pastel	Vintage

Limitations

- Limited to a small number of attributes
- Geometric understanding is still quite low
- Attributes are overlapping which results in a high false positive rate.

Future Work

- Increase the number of style attributes
- Extend the system to videos and 360 images

Code and Data

https://github.com/V-Sense/A-Geometry-Sensitive-Approach-for-Photographic-Style-Classification.git

Contact

ghosalk@tcd.ie

Trinity College Dublin

The University of Dublin

V-SENSE

Many Thanks!