A PRELIMINARY BENCHMARK OF FOUR SALIENCY ALGORITHMS ON COMIC ART

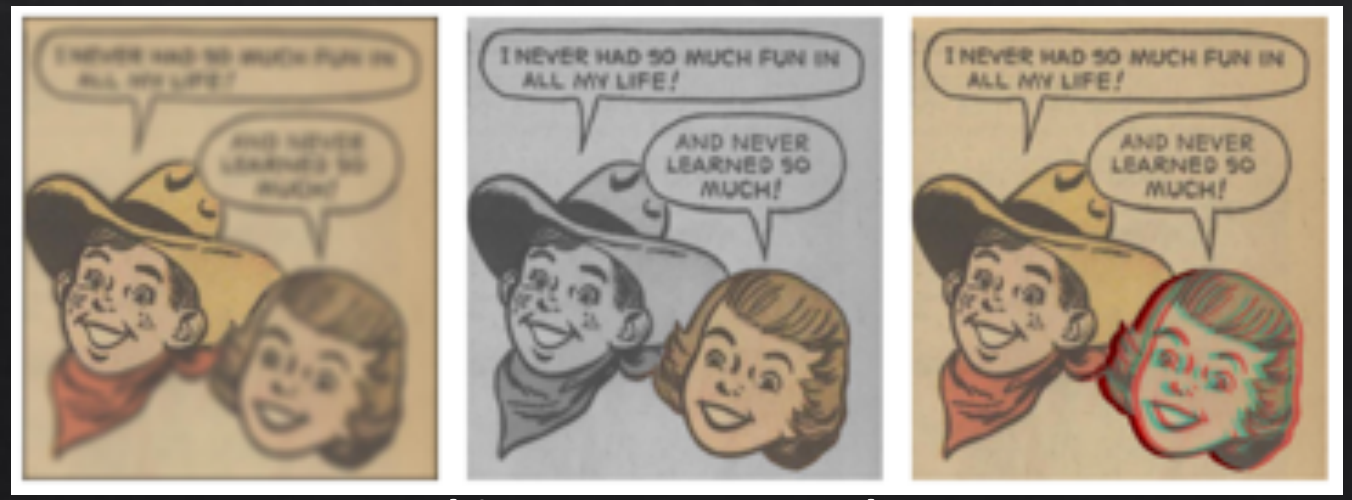
Khimya Khetarpal [1] & Eakta Jain [2]

University of Florida

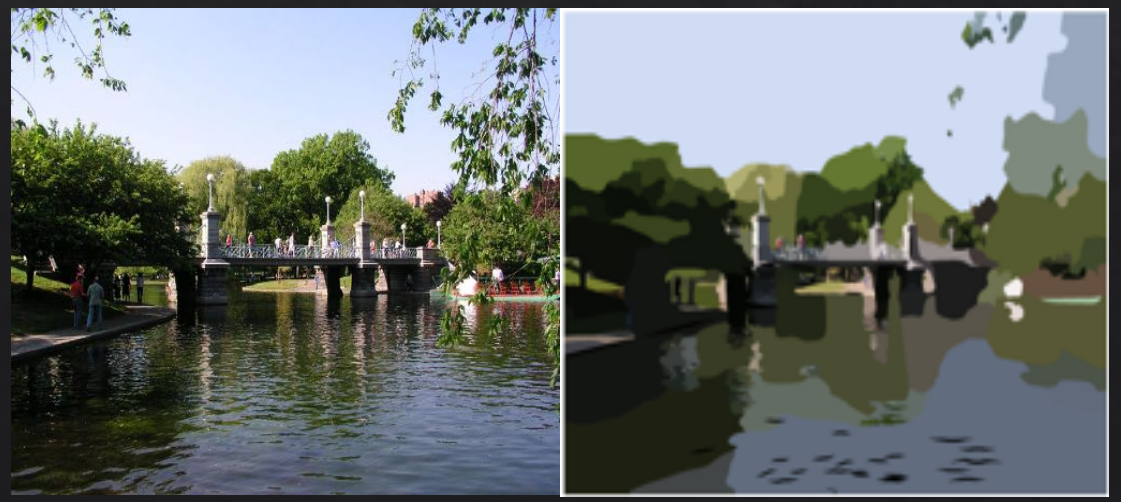
Department of ECE [1], Department of CISE [2]

Visual Saliency in Comic Art

Jain et al., IEEE CG&A Special Issue 16-to appear



Thirunarayanan et al., ACM SAP Poster 16



Judd et al., ICCV 09

Outline

- Benchmark of four *existing* saliency models on *comic art*
  - Data-driven approach - *Judd et al., ICCV 09* (LSVM)
  - Graph based bottom-up saliency model - *Harel et al., NIPS 06* (GBVS)
  - Difference of Gaussian based bottom up algorithm - *Frintrop et al., CVPR 15* (VOCUS2)
  - Region contrast based salient object detection - *Cheng et al., PAMI 15* (RC)

Eye Tracking Data – Comic Images

- Eye tracking data collected by Thirunarayanan et al., ACM SAP Poster 16
- 23 comic images
- 5 viewers (3 male)

Our Test Setup – Outline

Models	Image Categories	Eye Tracking Data	Metrics

• LSVM
• GBVS
• VOCUS2
• RC
	• Comics
	• CAT2000
	  • Outdoor Natural
	  • Outdoor Man-made
	  • Social
		• Comics eyetracking
		• CAT2000
		  • Training Data
		  • Testing Data
			• Normalized Scanpath Saliency (NSS)
			• Area Under Curve (AUC)

CAT2000 [Borji et al., CVPR 15]

Our Test Setup – Phase I

Models	Image Categories	Eye Tracking Data	Metrics

• LSVM
• GBVS
• VOCUS2
• RC
	• Comics
	• CAT2000
	  • Outdoor Natural
	  • Outdoor Man-made
	  • Social
		• Comics eyetracking
		• CAT2000
		  • Training Data
		  • Testing Data
			• Normalized Scanpath Saliency (NSS)
			• Area Under Curve (AUC)

CAT2000 [Borji et al., CVPR 15]

Our Test Setup – Phase I

Normalized Scanpath Saliency [NSS]	Outdoor ManMade	Outdoor Natural	Social	Overall
LSVM	1.22 (0.22)	1.28 (0.21)	1.21 (0.19)	1.24 (0.21)
GBVS	1.11 (0.28)	1.17 (0.34)	1.14 (0.27)	1.14 (0.3)
RC	0.88 (0.29)	0.92 (0.28)	0.83 (0.23)	0.88 (0.27)
VOCUS2	0.73 (0.32)	0.75 (0.29)	0.81 (0.26)	0.76 (0.29)

Our Test Setup – Phase I

Normalized Scanpath Saliency [NSS]	Outdoor ManMade	Outdoor Natural	Social	Overall
LSVM	1.22 (0.22)	1.28 (0.21)	1.21 (0.19)	1.24 (0.21)
GBVS	1.11 (0.28)	1.17 (0.34)	1.14 (0.27)	1.14 (0.3)
RC	0.88 (0.29)	0.92 (0.28)	0.83 (0.23)	0.88 (0.27)
VOCUS2	0.73 (0.32)	0.75 (0.29)	0.81 (0.26)	0.76 (0.29)

Area Under Curve [AUC]	Outdoor ManMade	Outdoor Natural	Social	Overall
LSVM	0.83 (0.04)	0.84 (0.04)	0.83 (0.05)	0.83 (0.04)
GBVS	0.79 (0.06)	0.79 (0.05)	0.79 (0.05)	0.79 (0.05)
RC	0.72 (0.06)	0.73 (0.73)	0.72 (0.63)	0.72 (0.07)
VOCUS2	0.68 (0.09)	0.67 (0.08)	0.70 (0.07)	0.68 (0.08)

Our Test Setup – Benchmark

Models	Image Categories	Eye Tracking Data	Metrics

• LSVM
• GBVS
• VOCUS2
• RC
	• Comics
	• CAT2000
	  • Outdoor Natural
	  • Outdoor Man-made
	  • Social
		• Comics eyetracking
		• CAT2000
		  • Training Data
		  • Testing Data
			• Normalized Scanpath Saliency (NSS)
			• Area Under Curve (AUC)

Our Test Setup – Benchmark

NSS	CAT2000	Comics
LSVM	1.24 (0.21)	0.88 (0.32)
GBVS	1.14 (0.3)	0.87 (0.44)
RC	0.88 (0.27)	0.64 (0.34)
VOCUS2	0.76 (0.29)	0.59 (0.34)

AUC	CAT2000	Comics
LSVM	0.83 (0.04)	0.72 (0.10)
GBVS	0.79 (0.05)	0.71 (0.10)
RC	0.72 (0.07)	0.66 (0.12)
VOCUS2	0.68 (0.08)	0.65 (0.12)

Our Test Setup – Benchmark

NSS	CAT2000	Comics
LSVM	1.24 (0.21)	0.88 (0.32)
GBVS	1.14 (0.3)	0.87 (0.44)
RC	0.88 (0.27)	0.64 (0.34)
VOCUS2	0.76 (0.29)	0.59 (0.34)

AUC	CAT2000	Comics
LSVM	0.83 (0.04)	0.72 (0.10)
GBVS	0.79 (0.05)	0.71 (0.10)
RC	0.72 (0.07)	0.66 (0.12)
VOCUS2	0.68 (0.08)	0.65 (0.12)

Our Test Setup – Benchmark

NSS	CAT2000	Comics
LSVM	1.24 (0.21)	0.88 (0.32)
GBVS	1.14 (0.3)	0.87 (0.44)
RC	0.88 (0.27)	0.64 (0.34)
VOCUS2	0.76 (0.29)	0.59 (0.34)

AUC	CAT2000	Comics
LSVM	0.83 (0.04)	0.72 (0.10)
GBVS	0.79 (0.05)	0.71 (0.10)
RC	0.72 (0.07)	0.66 (0.12)
VOCUS2	0.68 (0.08)	0.65 (0.12)

Our Test Setup – Benchmark

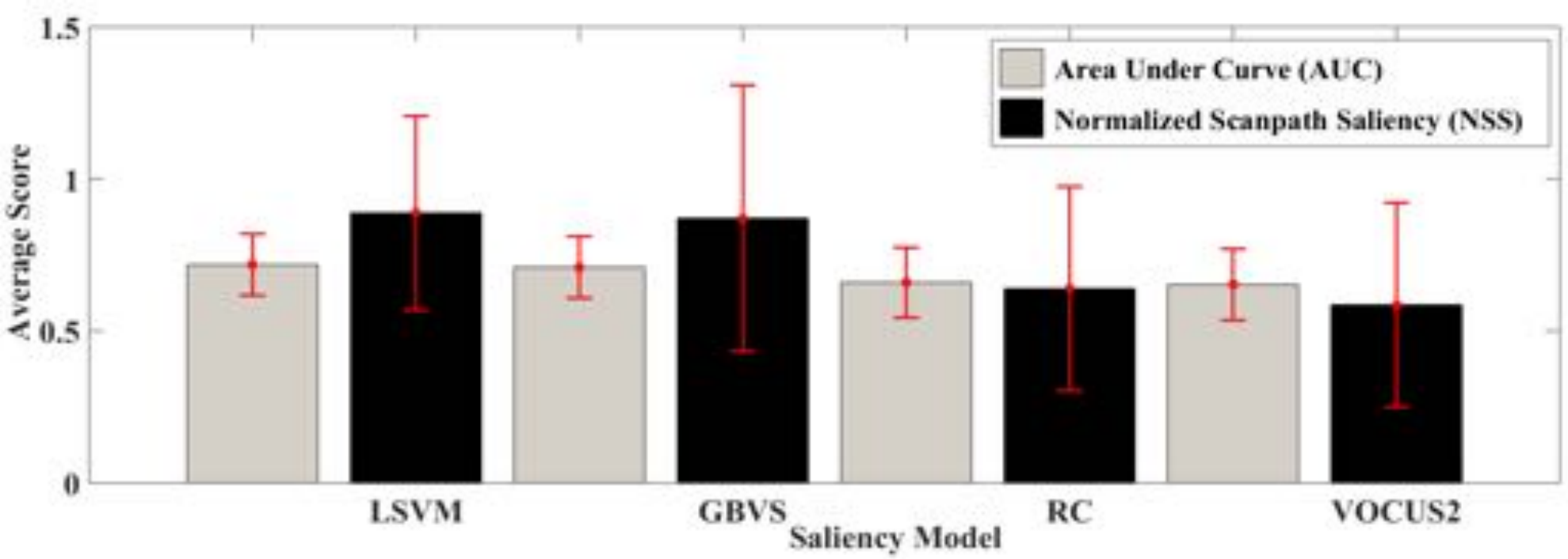
NSS	CAT2000	Comics
LSVM	1.24 (0.21)	0.88 (0.32)
GBVS	1.14 (0.3)	0.87 (0.44)
RC	0.88 (0.27)	0.64 (0.34)
VOCUS2	0.76 (0.29)	0.59 (0.34)

AUC	CAT2000	Comics
LSVM	0.83 (0.04)	0.72 (0.10)
GBVS	0.79 (0.05)	0.71 (0.10)
RC	0.72 (0.07)	0.66 (0.12)
VOCUS2	0.68 (0.08)	0.65 (0.12)

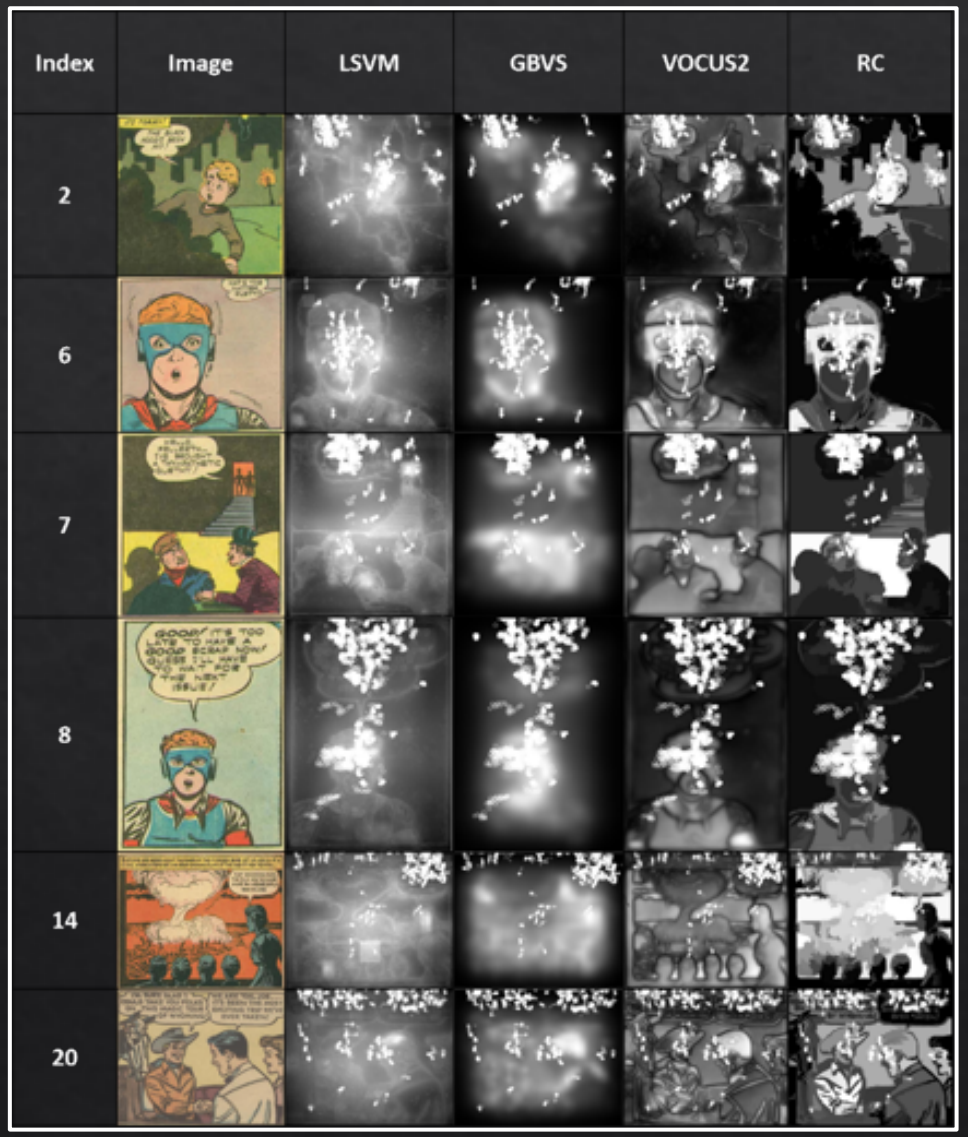
Our Test Setup – Benchmark

NSS	CAT2000	Comics
LSVM	1.24 (0.21)	0.88 (0.32)
GBVS	1.14 (0.3)	0.87 (0.44)
RC	0.88 (0.27)	0.64 (0.34)
VOCUS2	0.76 (0.29)	0.59 (0.34)

AUC	CAT2000	Comics
LSVM	0.83 (0.04)	0.72 (0.10)
GBVS	0.79 (0.05)	0.71 (0.10)
RC	0.72 (0.07)	0.66 (0.12)
VOCUS2	0.68 (0.08)	0.65 (0.12)



Experiment & Evaluation



Gaze data from 5 participants is overlaid using white circles on the saliency maps

Conclusion

LSVM is a leading candidate for visual saliency on comic art

Discussion

Why does LSVM model perform so well on comic images ?

Exploratory Analysis

High Level

- Cars [30]
- People [31]
- Faces [32]

Mid Level

- Horizon Line [29]

Center Prior

- Distance to center [33]

Low Level

- Steerable Pyramids [1-13]
- Intensity [14]
- Orientation [15]
- Color [16]
- Values of R, G, B [17-19]
- Prob. of R, G, B [20-22]
- 3D Histograms to find the Prob. of a color [23-27]
- Torralba Saliency [28]

Judd et al., ICCV 09

Exploratory Analysis

High Level

- Cars [30]
- People [31]
- Faces [32]

Mid Level

- Horizon Line [29]

Center Prior

- Distance to center [33]

Low Level

- Steerable Pyramids [1-13]
- Color [14]
- Intensity [15]
- Orientation [16]
- Values of R, G, B [17-19]
- Prob. of R, G, B [20-22]
- 3D Histograms to find the Prob. of a color [23-27]
- Torralba Saliency [28]

Judd et al., ICCV 09

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

Exploratory Analysis

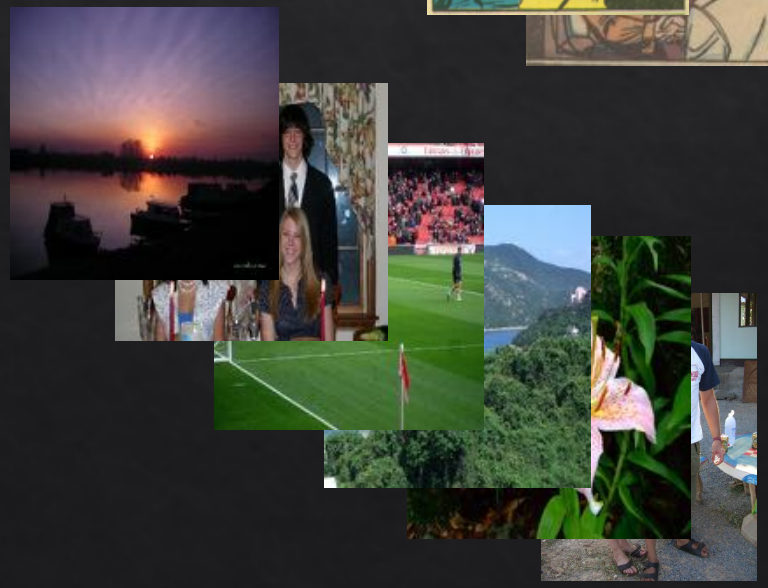
Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images

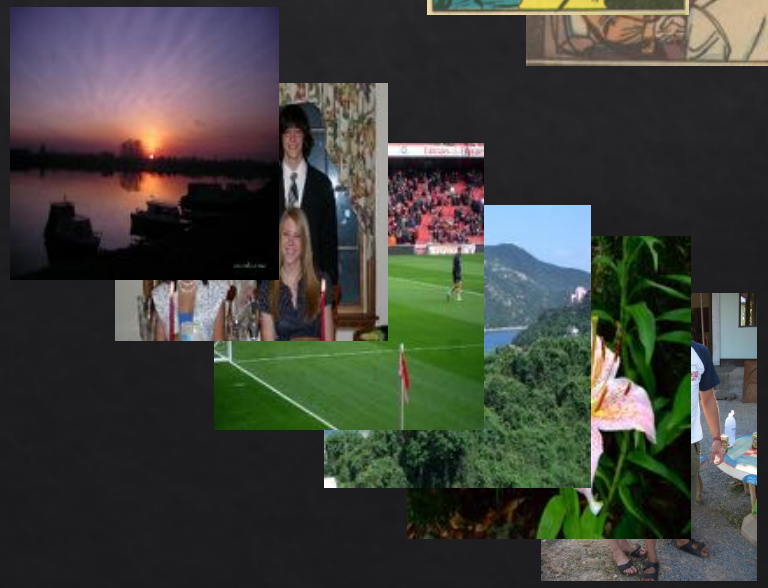


1003 Natural Images

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images



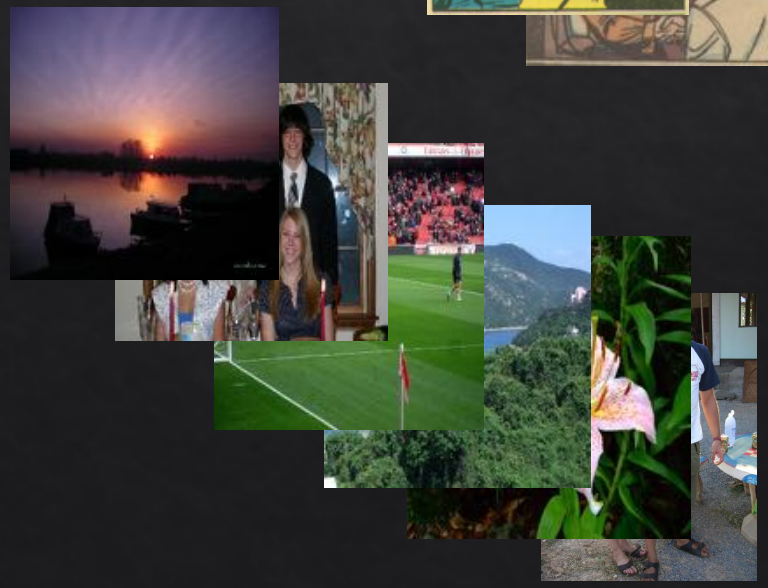
1003 Natural Images

3 most weighted ***feature channels*** are chosen

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images



1003 Natural Images

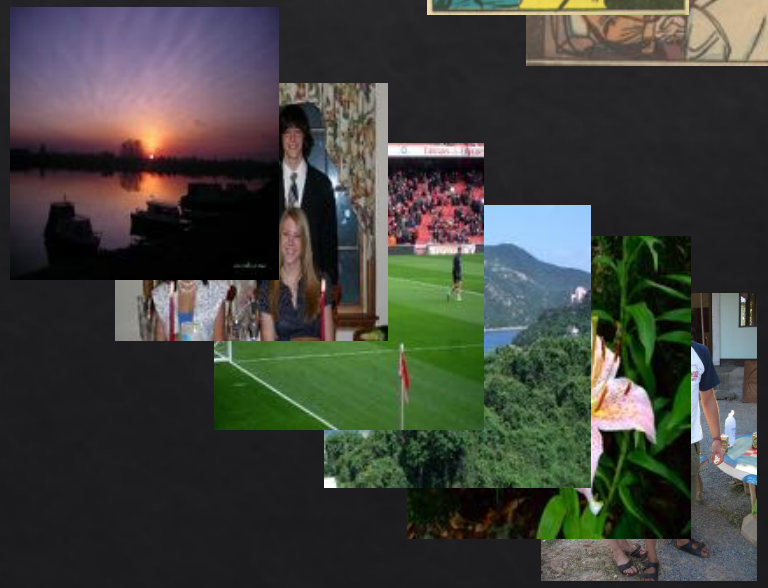
3 most weighted ***feature channels*** are chosen

Computation of ***feature vectors*** using the implementation in LSVM

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images



1003 Natural Images

Computation of ***feature vectors*** using the implementation in LSVM

3 most weighted ***feature channels*** are chosen

Feature Vectors

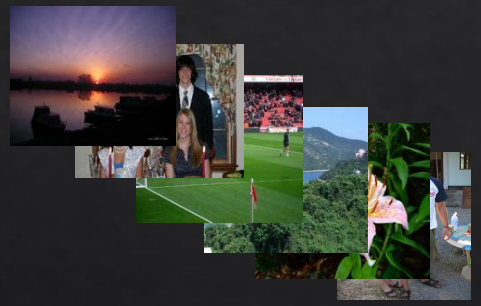
23	N*N
..	N*N
2	N*N
1	N*N

1	N*N
2	N*N
..	N*N
1003	N*N

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images



1003 Natural Images

Computation of ***feature vectors*** using the implementation in LSVM

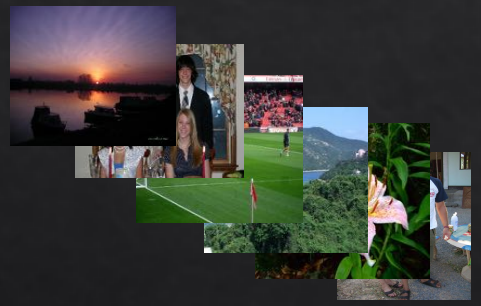
23	N*N
..	N*N
2	N*N
1	N*N

1	N*N
2	N*N
..	N*N
1003	N*N

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images



1003 Natural Images

Computation of ***feature vectors*** using the implementation in LSVM

23 N*N

.. N*N

2 N*N

1 N*N

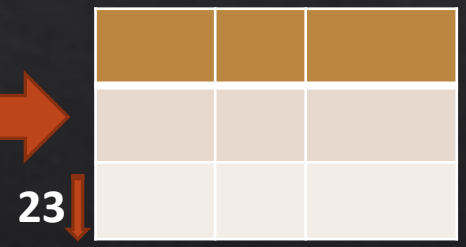
1 N*N

2 N*N

.. N*N

1003 N*N

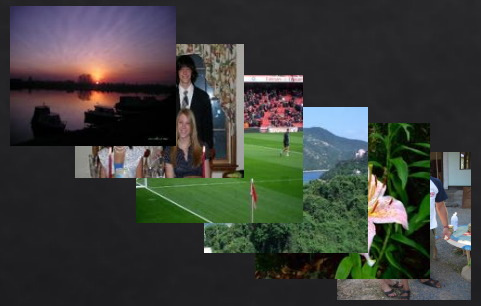
Euclidean Distance



Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***

23 Comic Images



1003 Natural Images

Computation of ***feature vectors*** using the implementation in LSVM

23 N*N

.. N*N

2 N*N

1 N*N

1 N*N

2 N*N

.. N*N

1003 N*N

Euclidean Distance

23

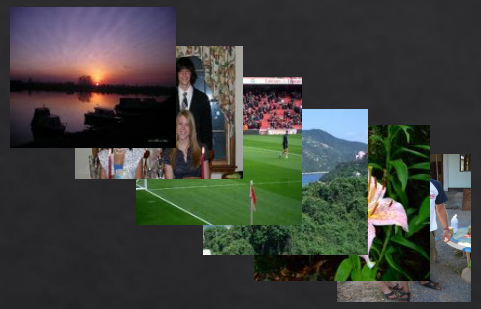
1003

Averaged across all comic images

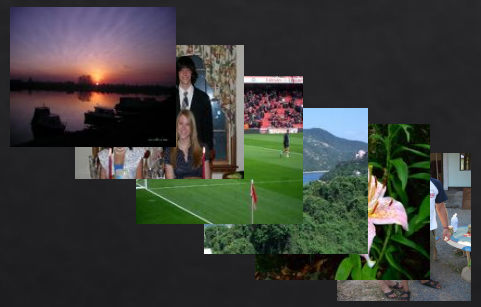
1*1003

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***



1003 Natural Images



1003 Natural Images

Computation of ***feature vectors*** using the implementation in LSVM

1003 N*N

.. N*N

2 N*N

1 N*N

1 N*N

2 N*N

.. N*N

1003 N*N

Euclidean Distance

1003

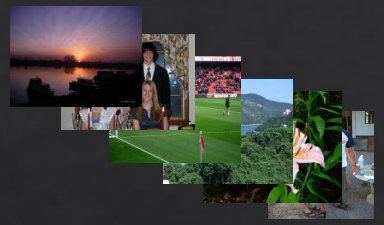
1003

Averaged across all natural images

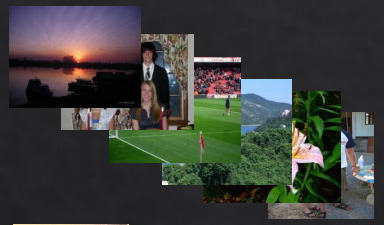
1*1003

Exploratory Analysis

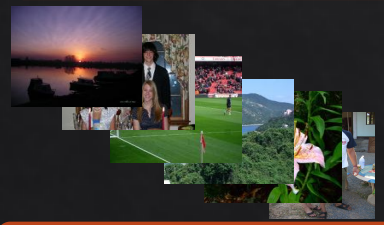
Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***



1003 Natural Images



23 Comic Images



1003 Natural Images

1*1003

Average distance between natural and natural images

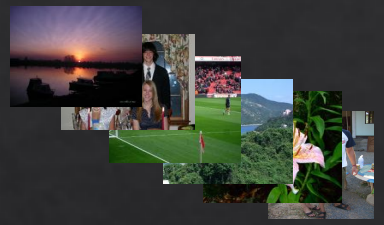
Average distance between comic and natural images

1*1003

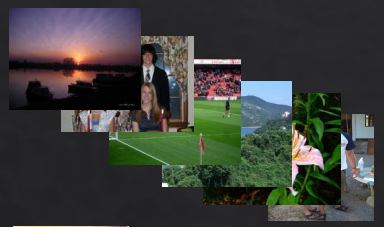
Similarity Score

Exploratory Analysis

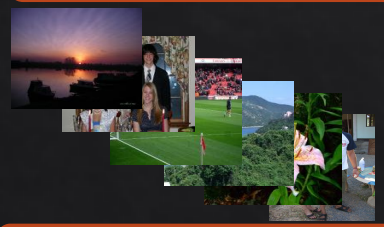
Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***



1003 Natural Images



23 Comic Images



1003 Natural Images

1*1003

Average distance between natural and natural images

Average distance between comic and natural images

1*1003

Similarity Score

High Positive

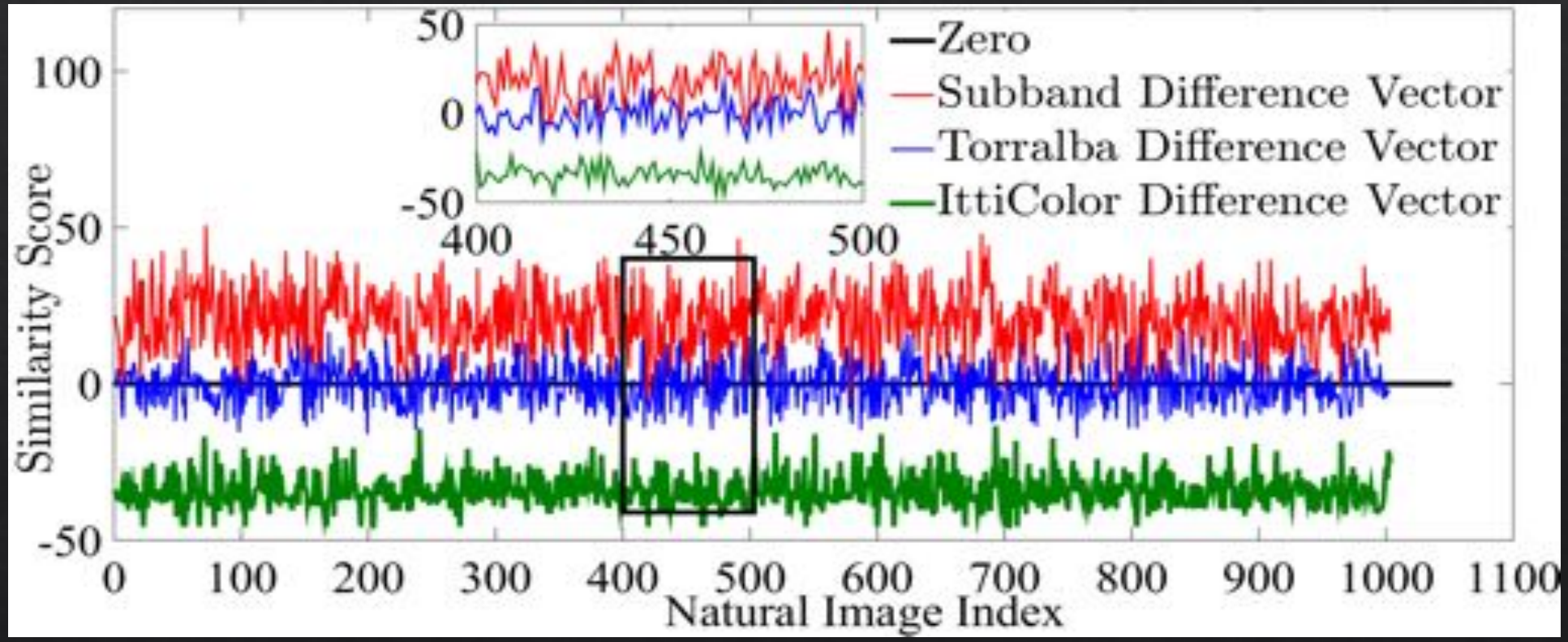
Negative

Higher similarity

Low Similarity

Exploratory Analysis

Hypothesis I : ***Comic panels in our dataset are similar to natural images in LSVM's dataset in terms of features used for saliency prediction***



Similarity between natural and comic images in context of 3 out of 33 features from LSVM model of saliency

Exploratory Analysis

High Level

- Cars [30]
- People [31]
- Faces [32]

Mid Level

- Horizon Line [29]

Center Prior

- Distance to center [33]

Low Level

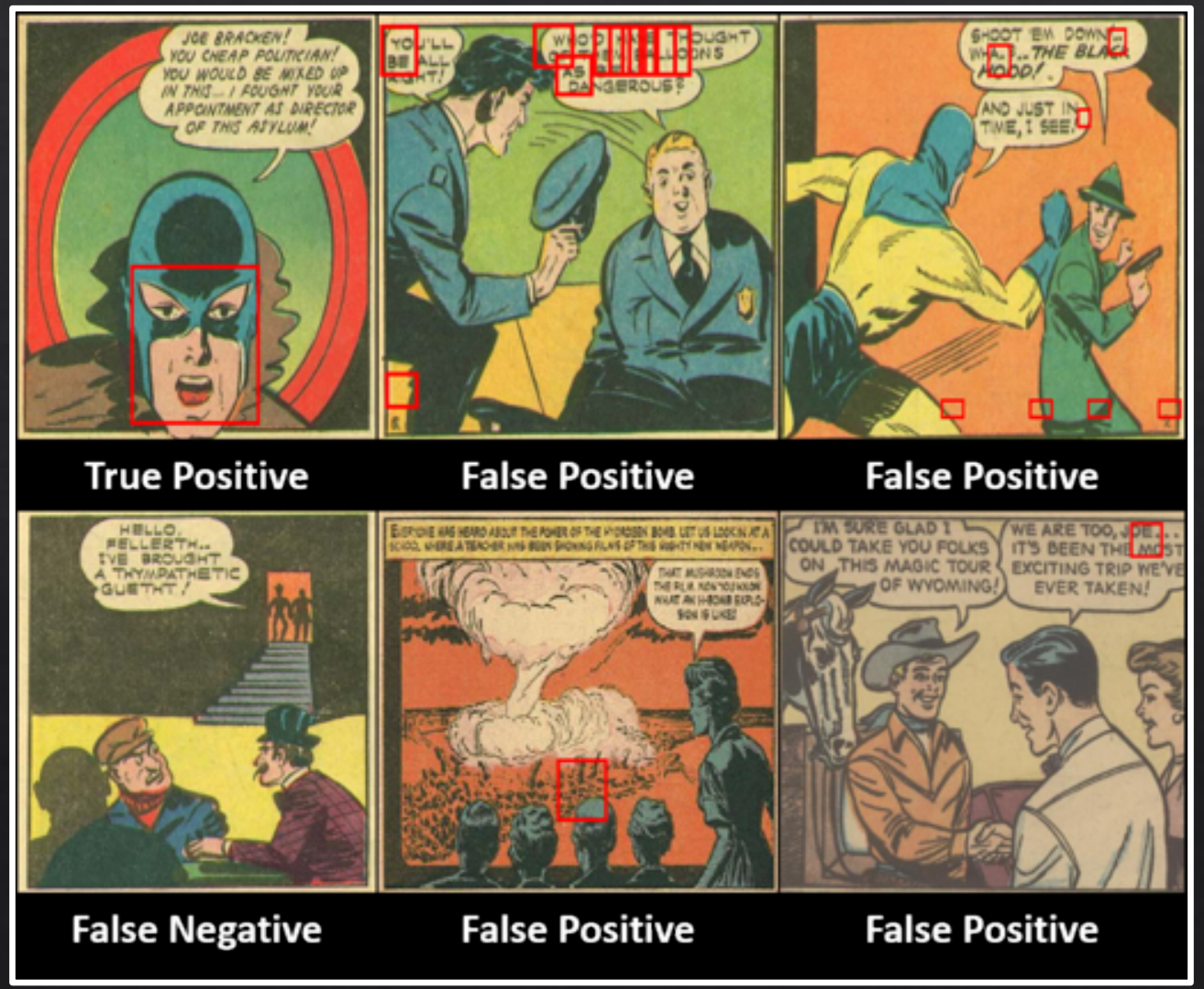
- Steerable Pyramids [1-13]
- Color [14]
- Intensity [15]
- Orientation [16]
- Values of R, G, B [17-19]
- Prob. of R, G, B [20-22]
- 3D Histograms to find the Prob. of a color [23-27]
- Torralba Saliency [28]

Exploratory Analysis

Hypothesis II : Face detection algorithms used by saliency models work well for natural images but not for comic art

Exploratory Analysis

Hypothesis II : Face detection algorithms used by saliency models work well for natural images but not for comic art



Take Aways

- Benchmarked *four existing saliency algorithms* on *comic images* performing a comparative analysis with two metrics
- Models in our study have the *same order of performance* for both *natural images* as well as *comic art* – a data driven model outperforms all
- All models show *relatively lower performances* on *comic art* – need for saliency models targeted for comics !
- Faces in comic art are different in style than natural images - *Feature engineering* needed for visual saliency in comics

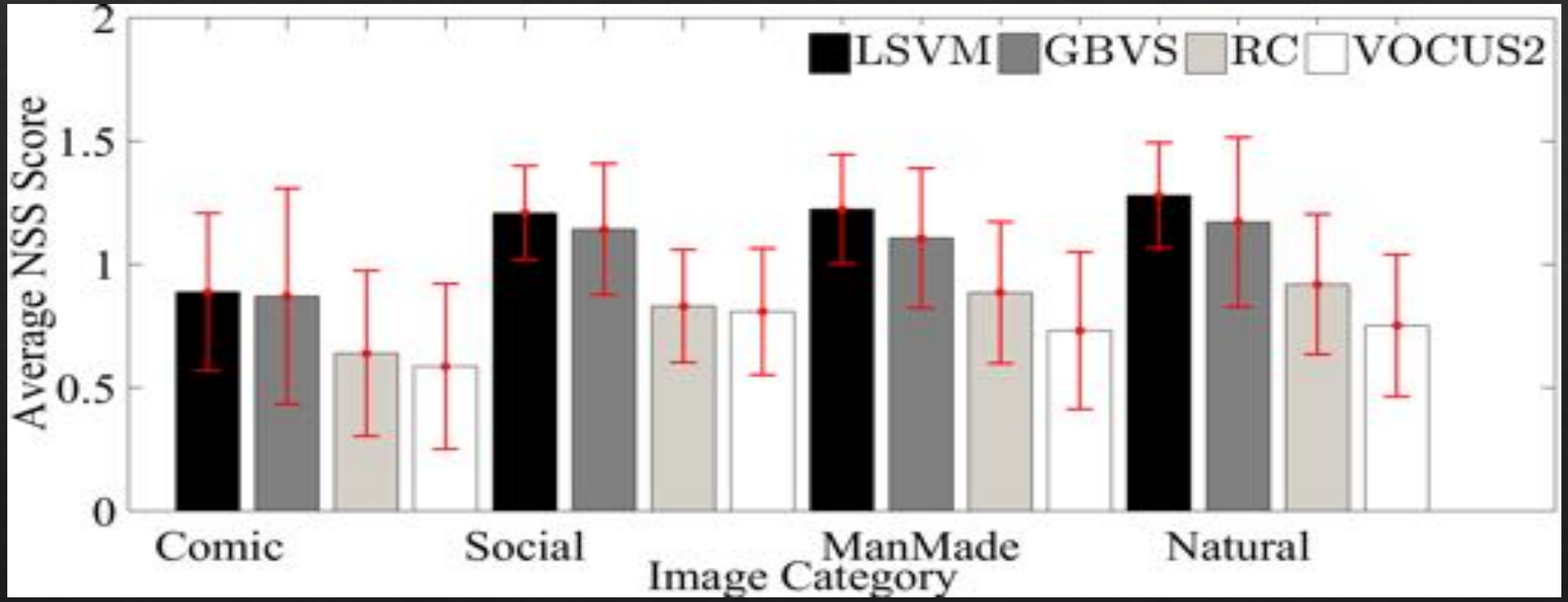
Future Work

- Extension by comparing a *bigger pool of saliency models* existing in the literature
- *Evaluation of deep learning based saliency*
- *More features* and *bigger datasets* including manga images for example
- Applying these saliency algorithms to the *applications for comic art*

Thank You

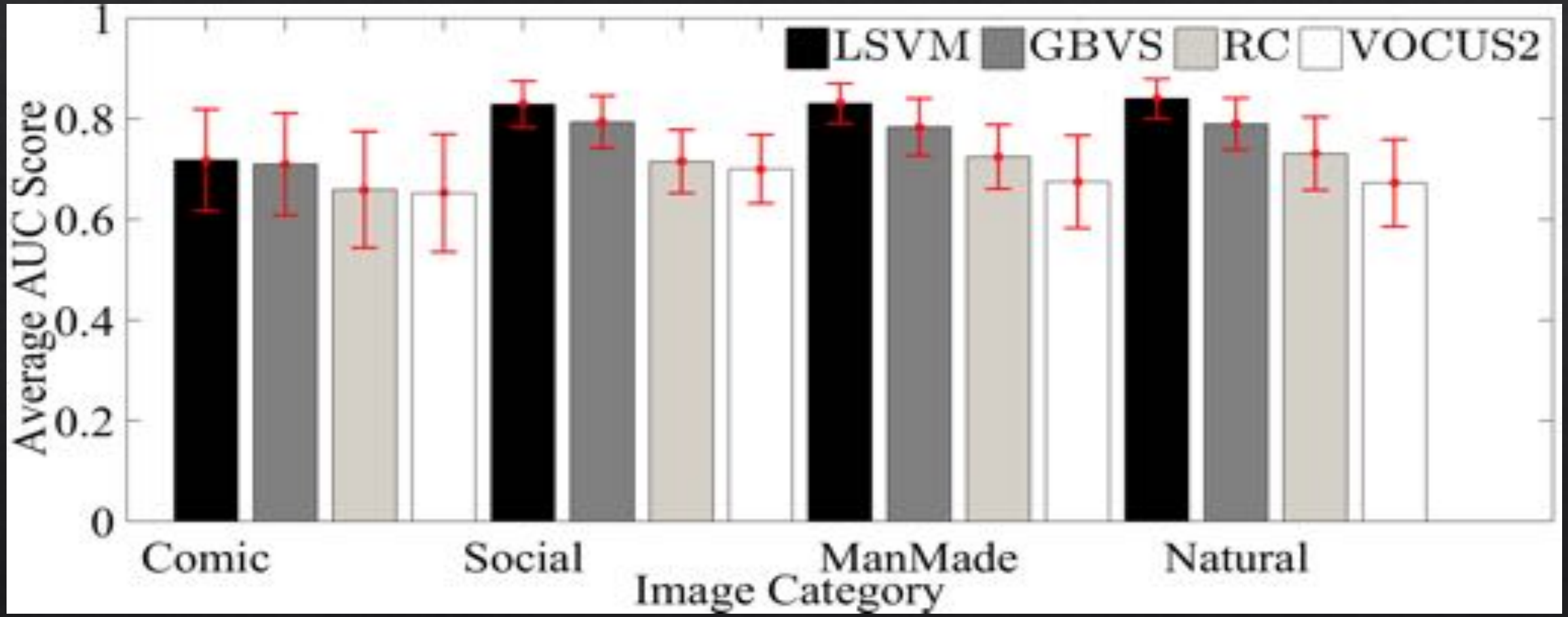
Extra Slides

Performance Comparison - NSS



Mean NSS score where error bars are standard deviation

Performance Comparison – AUC



Mean AUC score with error bars as standard deviation

Our Test Setup – Phase I

NSS	ManMade	Natural	Social	Overall
LSVM	1.22 (0.22)	1.28 (0.21)	1.21 (0.19)	1.24 (0.21)
	1.27 (0.2)	1.23 (0.21)	1.18 (0.19)	1.23 (0.2)
GBVS	1.11 (0.28)	1.17 (0.34)	1.14 (0.27)	1.14 (0.3)
	1.18 (0.34)	1.06 (0.36)	1.1 (0.32)	1.11 (0.34)

AUC	ManMade	Natural	Social	Overall
LSVM	0.83 (0.04)	0.84 (0.04)	0.83 (0.05)	0.83 (0.04)
	0.84 (0.04)	0.84 (0.04)	0.83 (0.04)	0.84 (0.04)
GBVS	0.79 (0.06)	0.79 (0.05)	0.79 (0.05)	0.79 (0.05)
	0.8 (0.05)	0.78 (0.06)	0.79 (0.06)	0.79 (0.06)

MIT Saliency Benchmark

Data : CAT2000

Method	MSS	CA	LSVM	RC
AUC	0.683	0.844	0.849	0.830

Cheng et al, PAMI 2015

Data : Judd et al

MSS Achanta et al, IEEE ICIP [2010]

CA Goferman et al, IEEE TPAMI [2012]

Saliency Models

- LSVM
  - Hybrid of *bottom up and top down*
  - Employs *33 features* ranging from low level to high level features
  - A Linear Support Vector Machine is trained using the eye tracking data itself

- GBVS
  - Given an image I, GBVS uses a *Markovian approach* to compute activation maps from feature maps.
  - The second step; *normalization aims at providing more weight to the salient regions*, thus resulting in more informative saliency map.

Saliency Models

- VOCUS2
  - A pyramid structure consisting of *twin pyramids with multiple scales per layer*
  - As opposed to one pyramid with one scale per layer used in iNVT model.
  - Features are computed in parallel,
  - *center surround* contrast is computed by *difference-of-gaussians* on different scales.

- RC
  - In RC, the input image is first *segmented into regions*,
  - Then *color contrast at the region level is computed*, and finally *saliency for each region is defined as the weighted sum of the regions contrasts to all other regions in the image.*
  - The weights are set according to the spatial distances with farther regions being assigned smaller weights.

Normalized Scanpath Saliency

- Each saliency map was *linearly normalized* to have zero mean and unit standard deviation.
- The normalized *saliency values were extracted corresponding to the fixation locations* for each subject
- The *mean of these values* was taken as a measure of the correspondence between the saliency map and Scanpath.
- For each model, mean of all the five subjects' NSS score results in the average NSS score for an image. The same is repeated to compute the mean NSS score for all four models across 23 comic images.

Area Under Curve

- A *binary map* with 1 corresponding to the fixation locations.
- Jitter is introduced to saliency maps that come from saliency models that have a lot of zero values.
- The saliency map is then normalized - values are then sorted to be used as *threshold values.*
- Furthermore, true positive rate is calculated by *ratio of the salience map values at fixation locations above the threshold*, where as false positive rate is given by ratio of other salience map vales above threshold.
- The AUC score is calculated by the *area under the curve governed by true positive and false positive rates.*

CAT2000

Outdoor Natural





CAT2000

Outdoor Manmade

CAT2000

Social