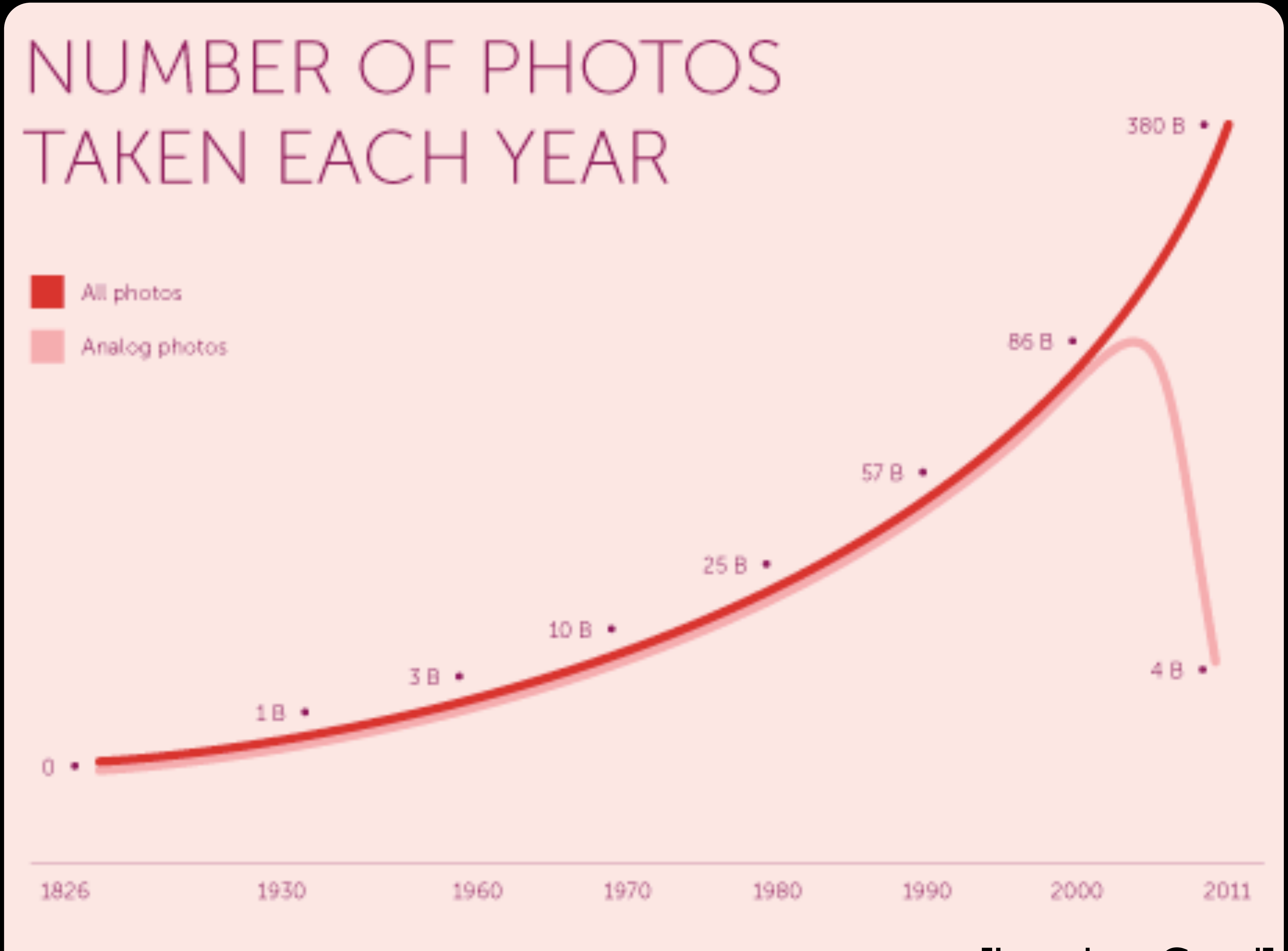
Semantic Awareness for Automatic Image Interpretation

Albrecht Lindner

Private defense – Dec 21, 2012

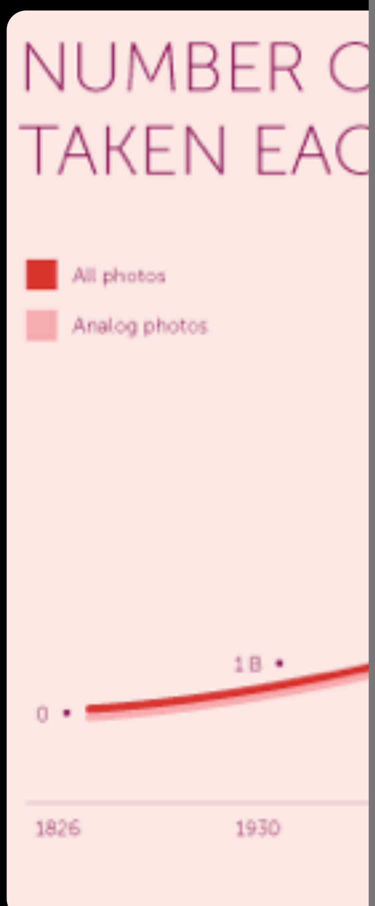
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

A Vast Challenge/Opportunity



[Jonathan Good]

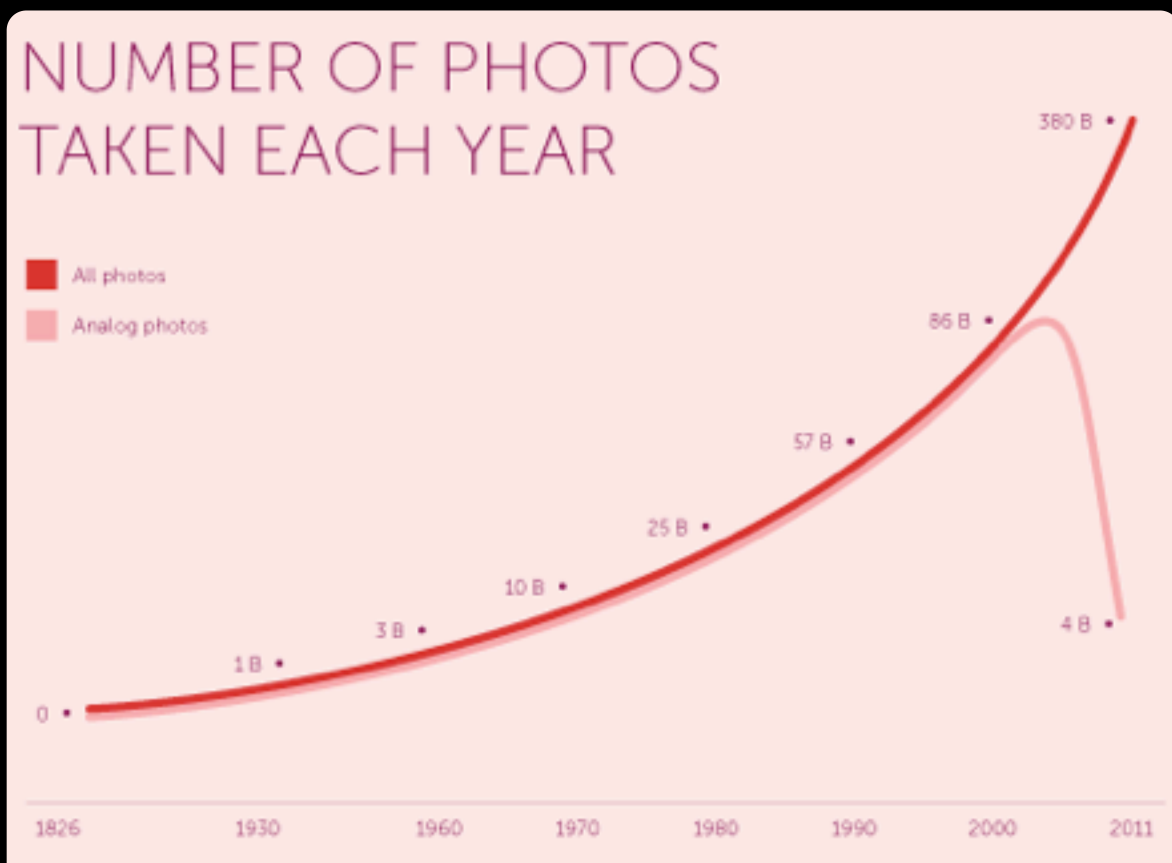
A Vast Challenge/Opportunity



"This suggests that there are, at the very least, a quarter of a million distinct English words, excluding inflections, and words from technical and regional vocabulary not covered by the OED ..."

[Oxford English Dictionary]

A Vast Challenge/Opportunity



[Jonathan Good]

"This suggests that there are, at the very least, a quarter of a million distinct English words, excluding inflections, and words from technical and regional vocabulary not covered by the OED ..."

[Oxford English Dictionary]

Novel methods and applications to link digital image content with human language.

Thesis Overview

method

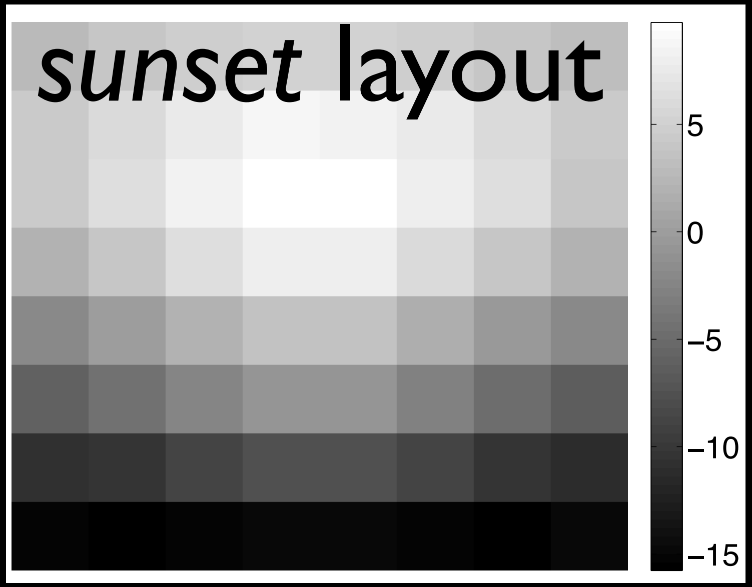
applications

Thesis Overview

Method:

Applications:

Thesis Overview

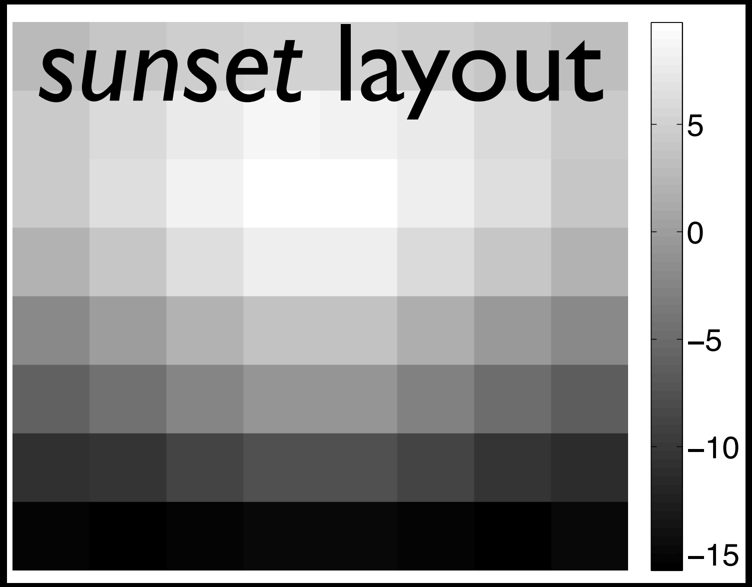


Method:
statistical framework

Applications:

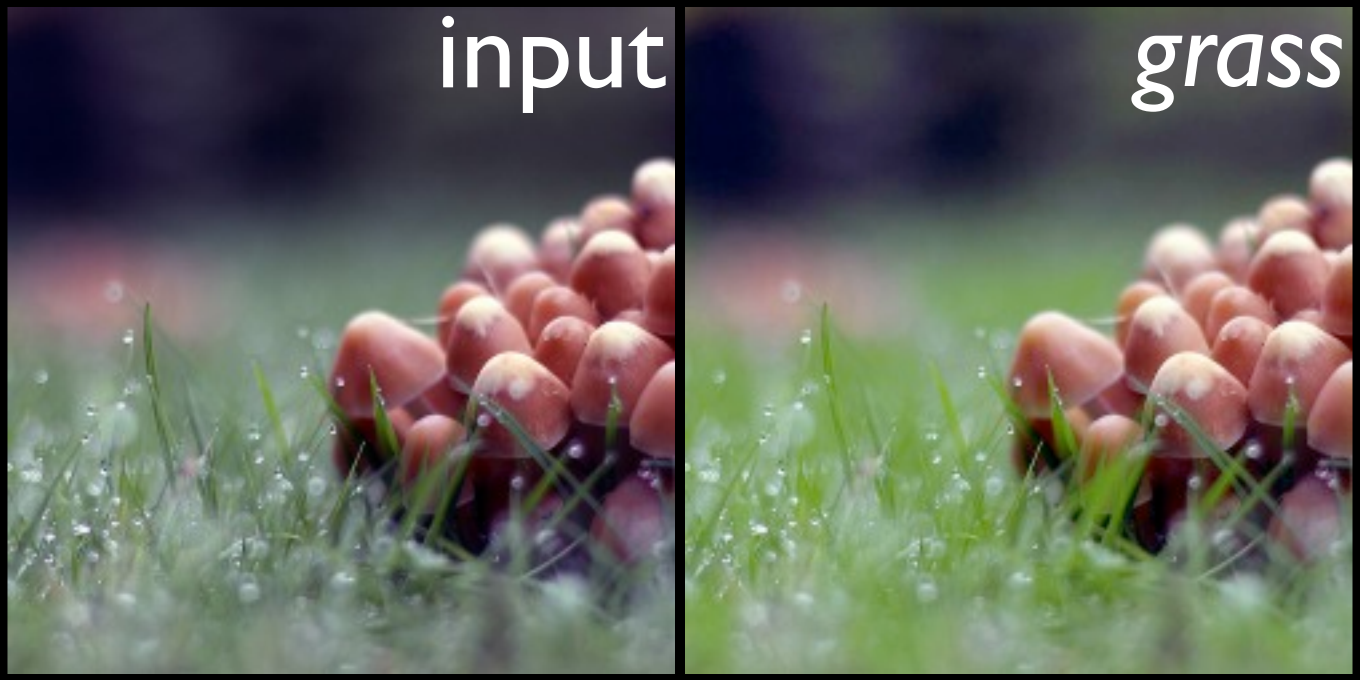
Thesis Overview

Method:
statistical framework

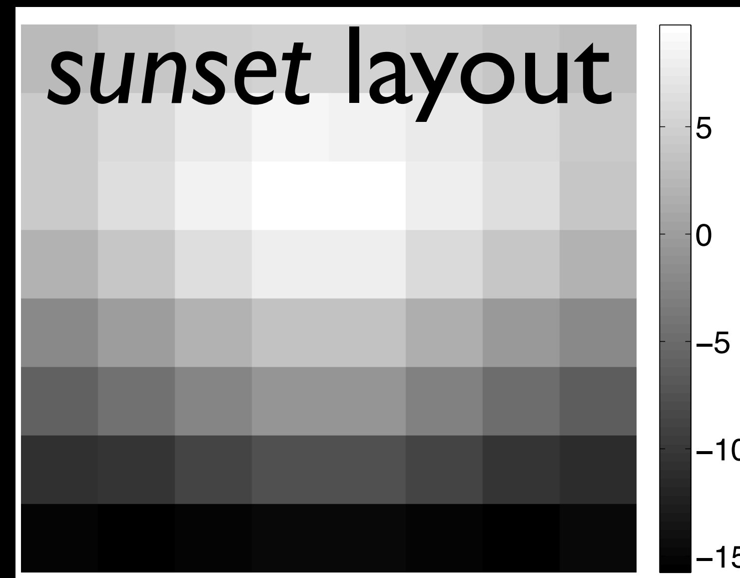


Applications:

1. semantic image enhancement



Thesis Overview

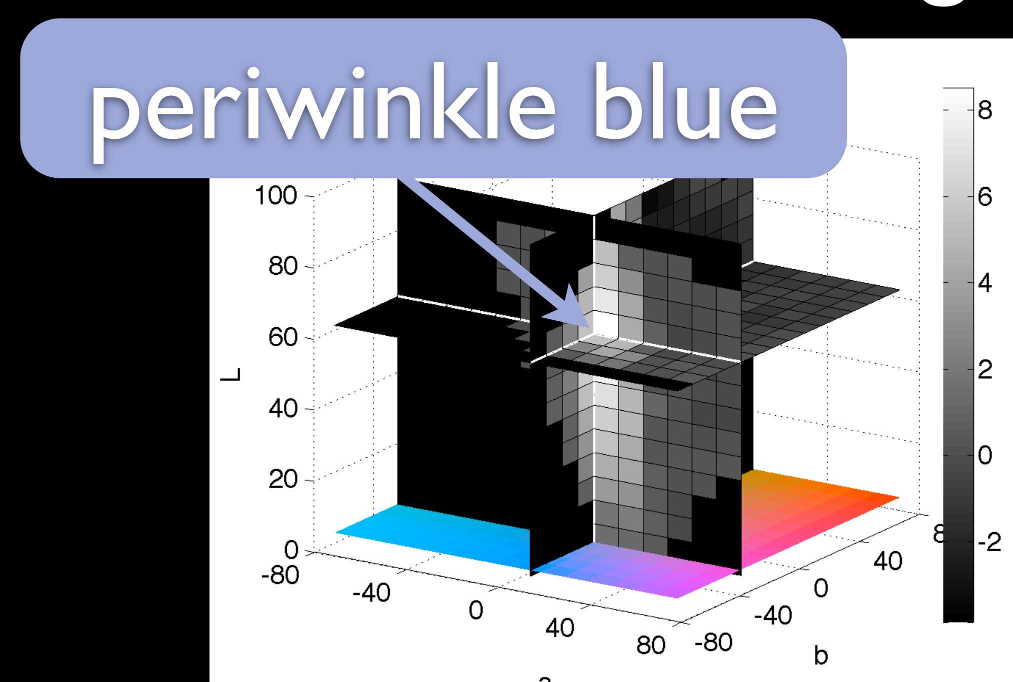


Method:
statistical framework

Applications:

1. semantic image enhancement

2. color naming



Statistical Framework

Link image characteristics with keywords.

Image Database

- MIR Flickr database, 1 Million annotated images.
- Selection based on Flickr's "interestingness" score.
- 1 MegaPixel, assume sRGB.

gold, oregoncoast, fortstevens, astoria, outside, lightroom, sigma, 1020mm, nikon, d40, diamondclassphotographer, grass, yellow, blue, sky, clouds, singlecloud, color, saturated, happy, field

[Huiskes et al., ACMMM, 2010]

Statistical Framework

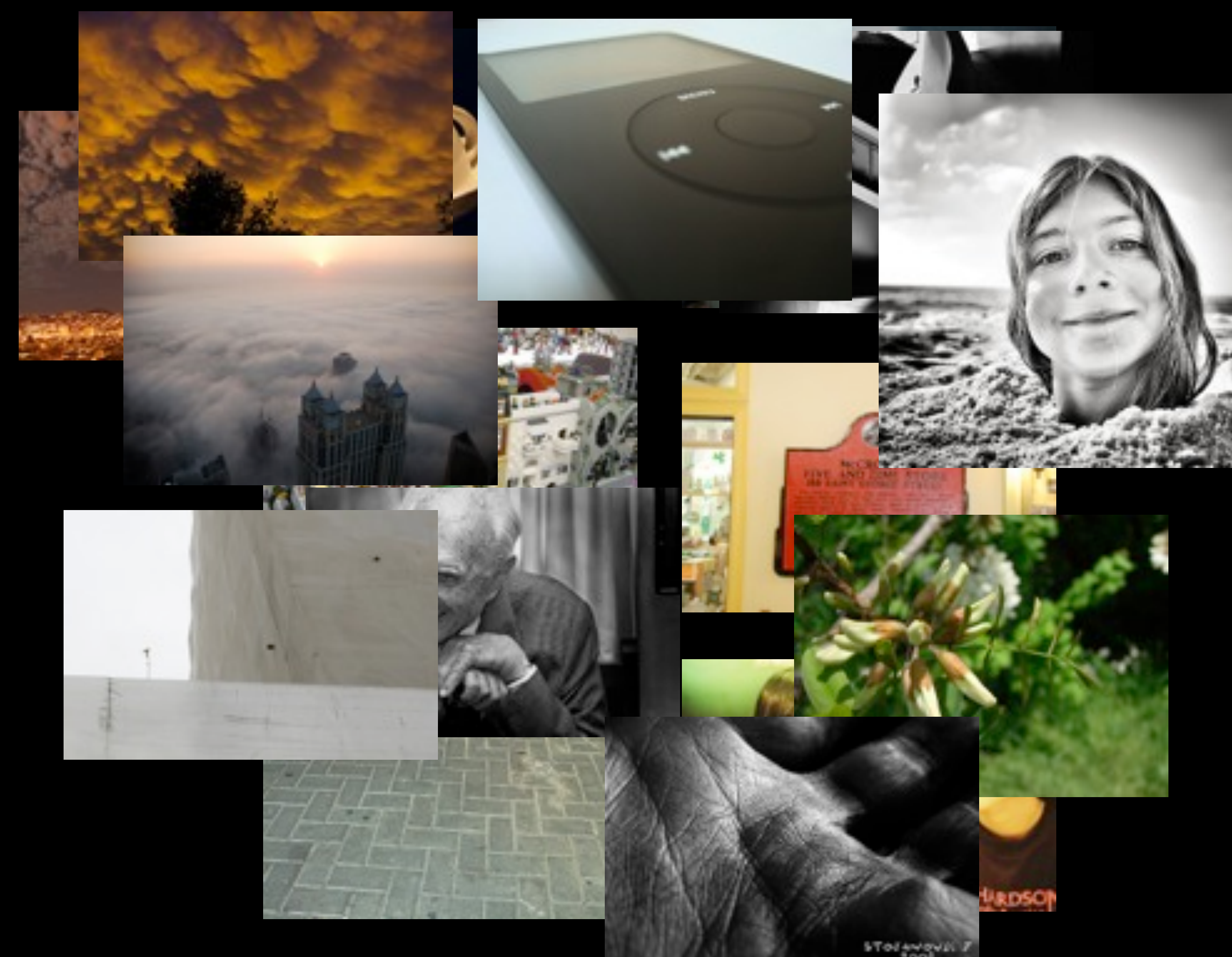
1M images + keywords

Statistical Framework

gold

3312

$\overline{\textit{gold}}$

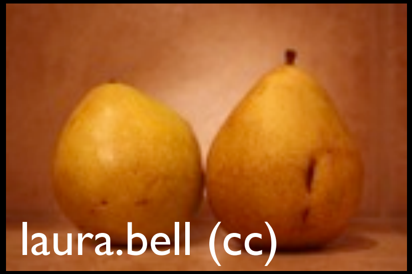


996'688

Statistical Framework

gold


stevewhis (cc)


laura.bell (cc)


paige_eliz (cc)


Arty Smokes (cc)

4

$\overline{gold}$


TW Collins (cc)


raketentim (cc)


Dunechaser (cc)


golfnride (cc)


Tal Bright (cc)


10863752@N00 (cc)

6

Statistical Framework

gold

$\overline{gold}$

percentage of yellow pixels

Statistical Framework

sorted list: 0% 2% 3% 5% 8% 9% 10% 30% 70% 90%

Statistical Framework

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

ranksum: $T = 4 + 7 + 9 + 10 = 30$

Statistical Framework

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

ranksum: $T = 4 + 7 + 9 + 10 = 30$

Mann-Whitney-Wilcoxon ranksum test

$$\mu_T = \frac{n_w(n_w + n_{\overline{w}} + 1)}{2} \qquad \sigma_T^2 = \frac{n_w n_{\overline{w}}(n_w + n_{\overline{w}} + 1)}{12}$$

$n_w, n_{\overline{w}}$ cardinalities of both sets

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{30 - 22}{4.69} \approx 1.71$$

[F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin, 1(6):80–83, 1945]

Statistical Framework

sorted list: 0% 2% 3% 5% 8% 9% 10% 30% 70% 90%

rank index: 1 2 3 4 5 6 7 8 9 10

ranksum: $T = 4 + 7 + 9 + 10 = 30$

Mann-Whitney-Wilcoxon ranksum test

$$\mu_T = \frac{n_w(n_w + n_{\overline{w}} + 1)}{2} \qquad \sigma_T^2 = \frac{n_w n_{\overline{w}}(n_w + n_{\overline{w}} + 1)}{12}$$

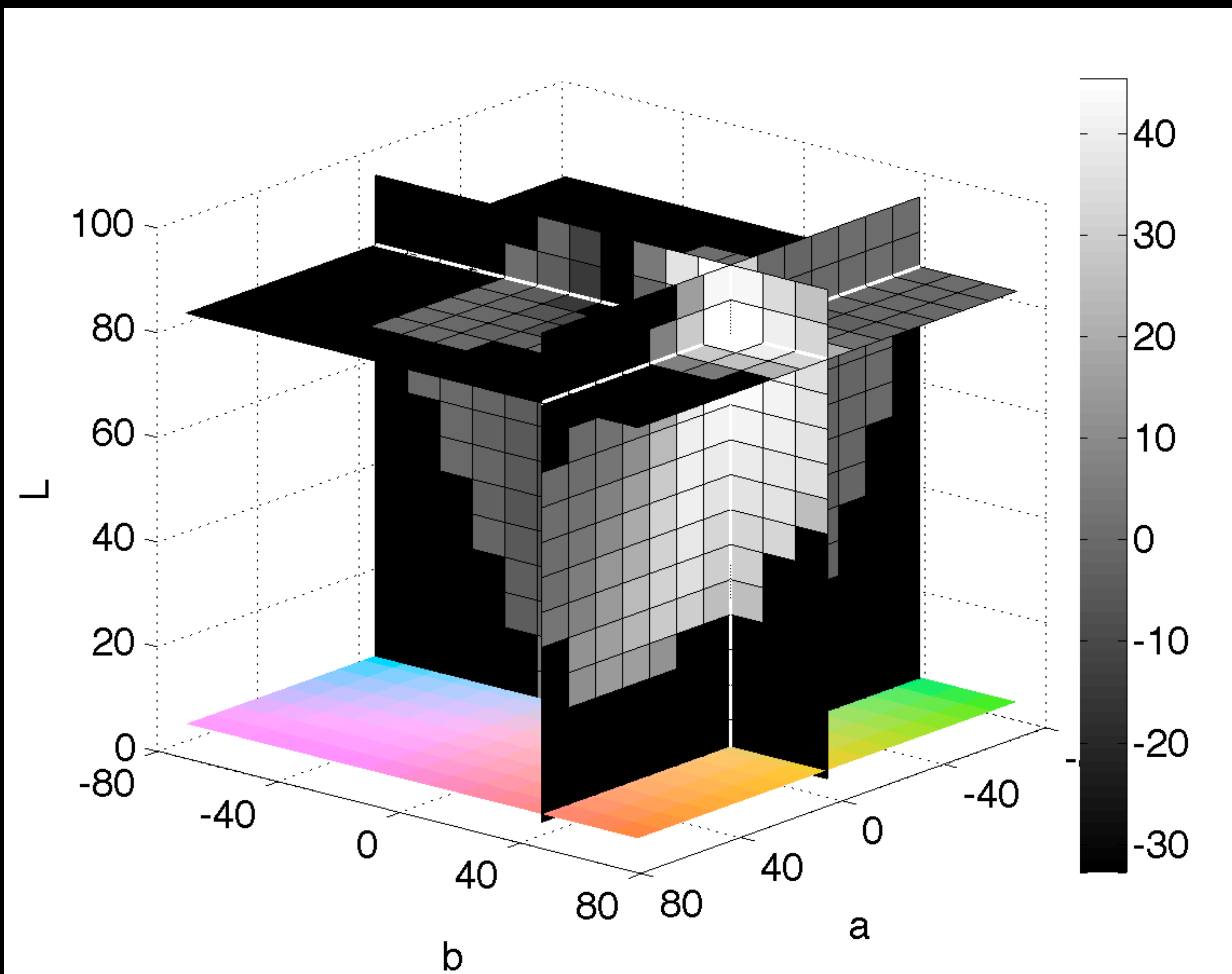
$n_w, n_{\overline{w}}$ cardinalities of both sets

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{30 - 22}{4.69} \approx 1.71$$

$z > 0$ ➔ significantly more yellow pixels in *gold* images.

z Distribution

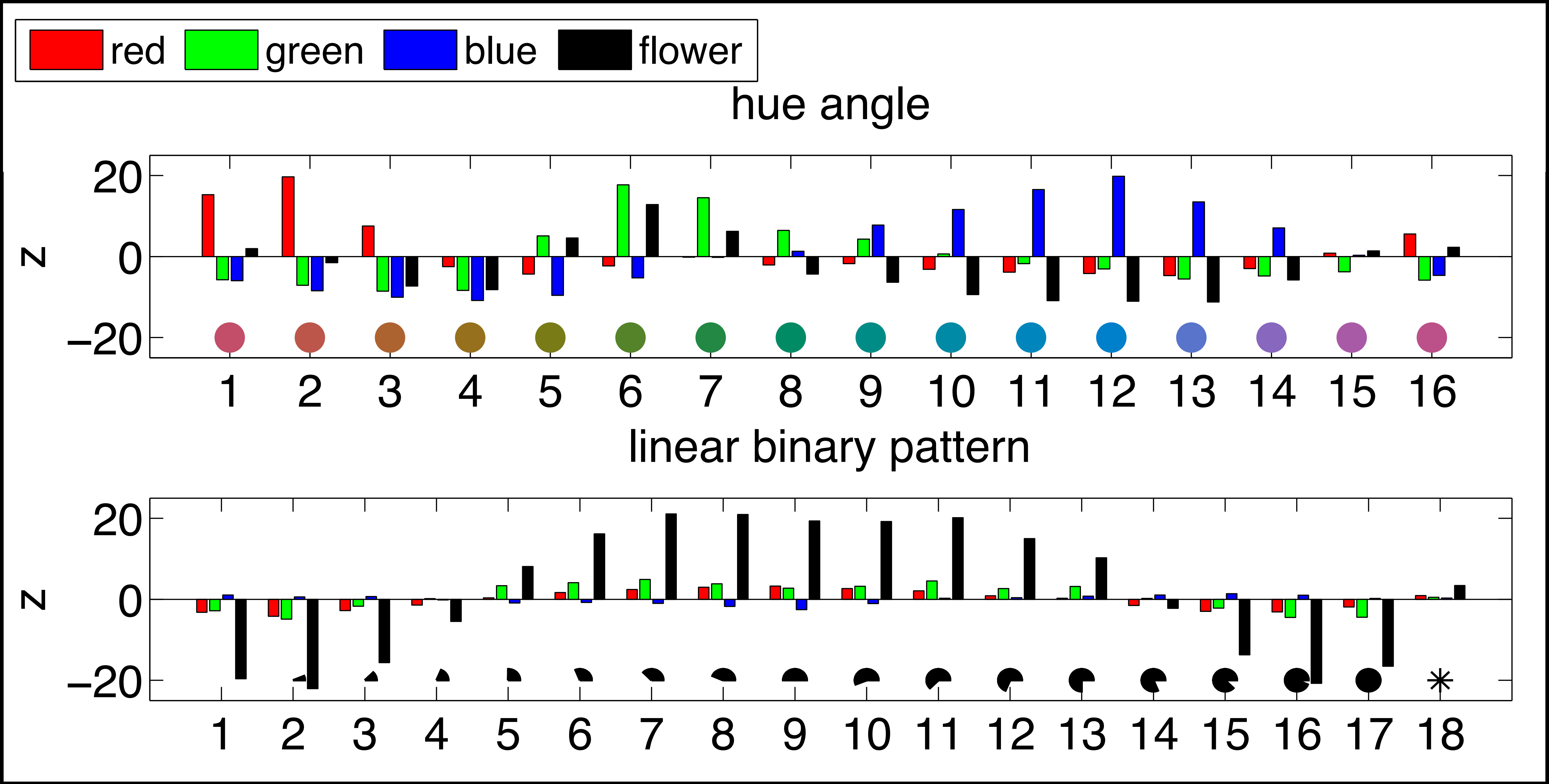
gold



- CIELAB histogram 15x15x15 bins.
- z values indicate significance of a keyword w.r.t. to a characteristic.

Other Characteristics

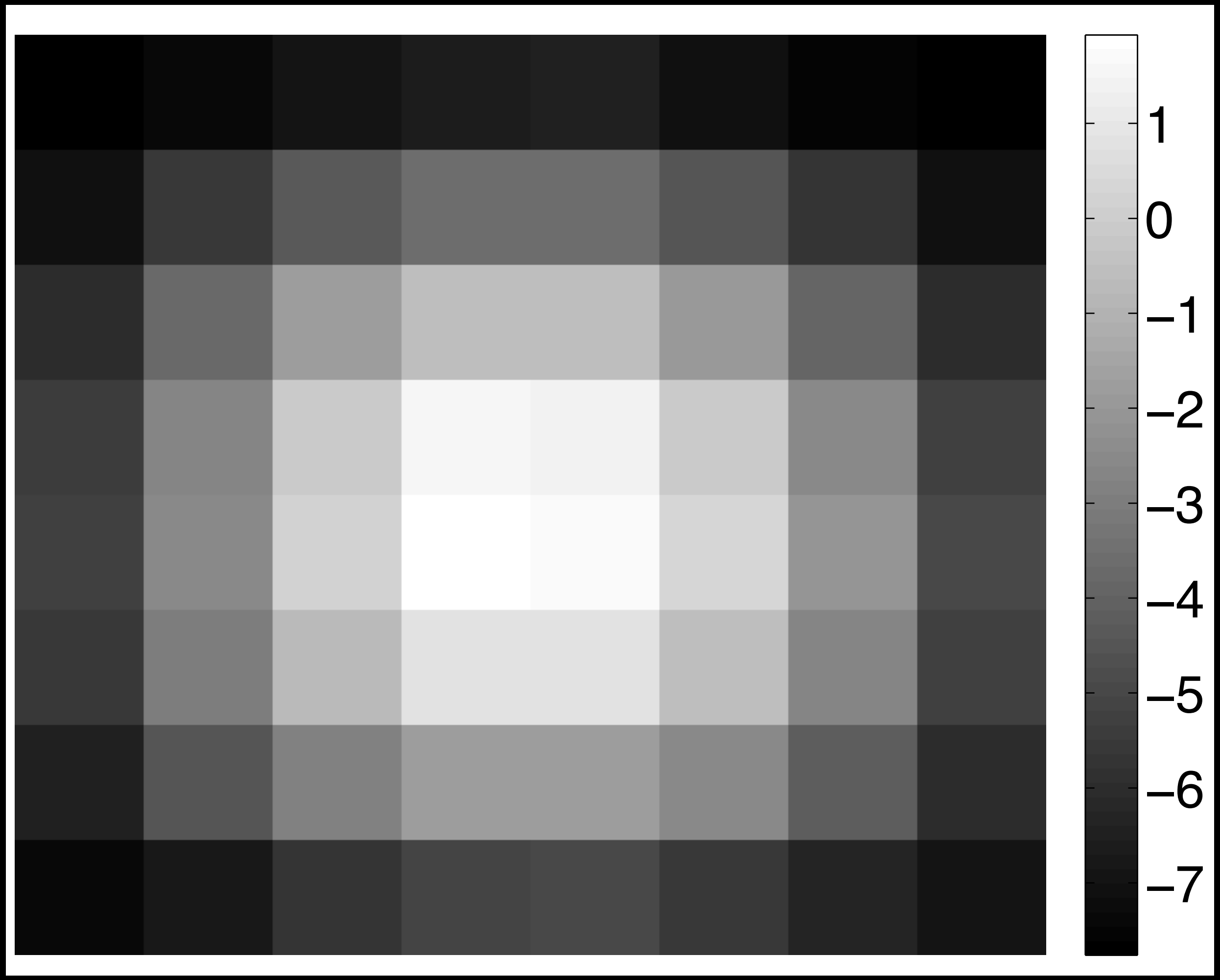
Hue angle and linear binary pattern.



Other Characteristics

Spatial lightness layout.

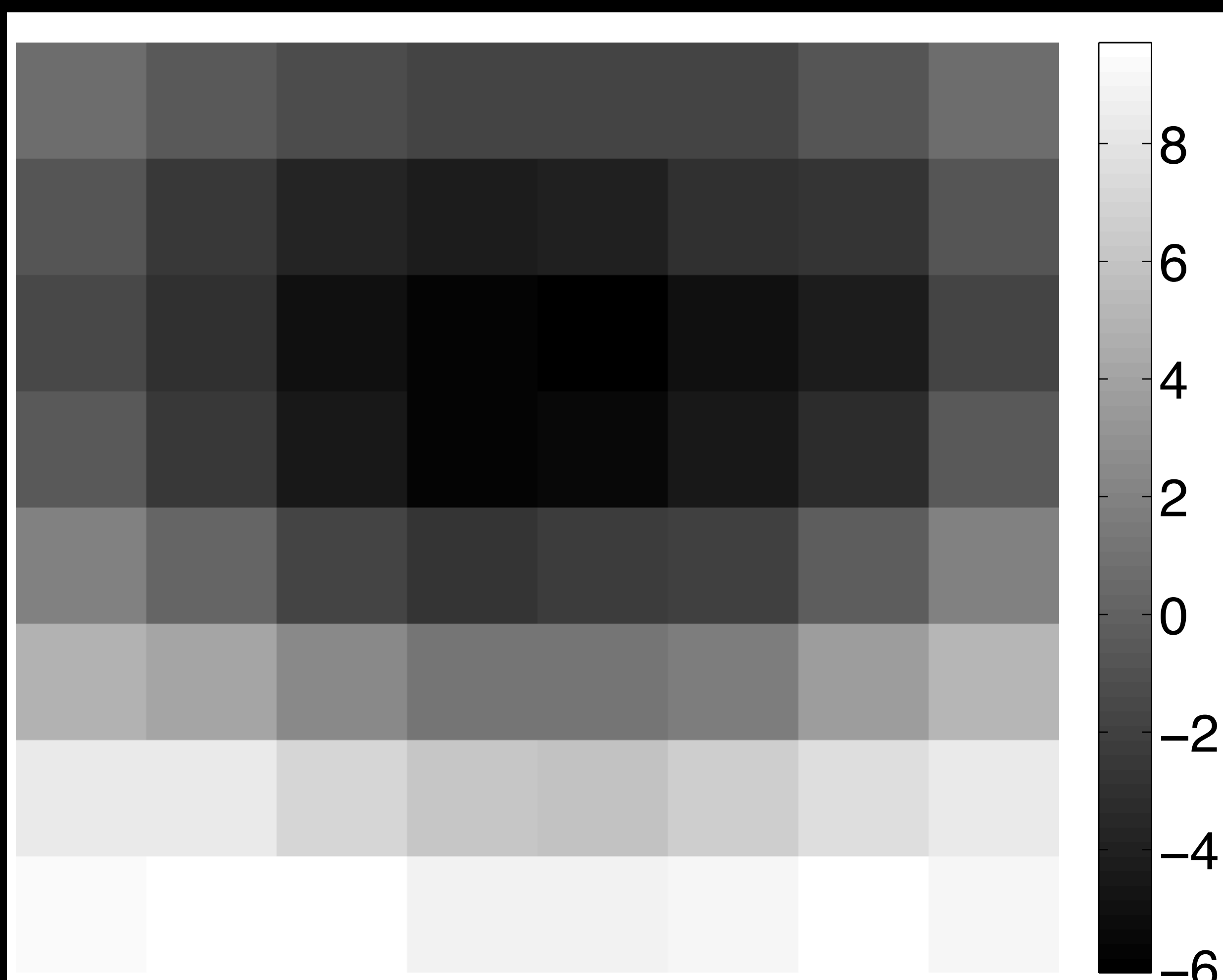
light



Other Characteristics

Spatial chroma layout.

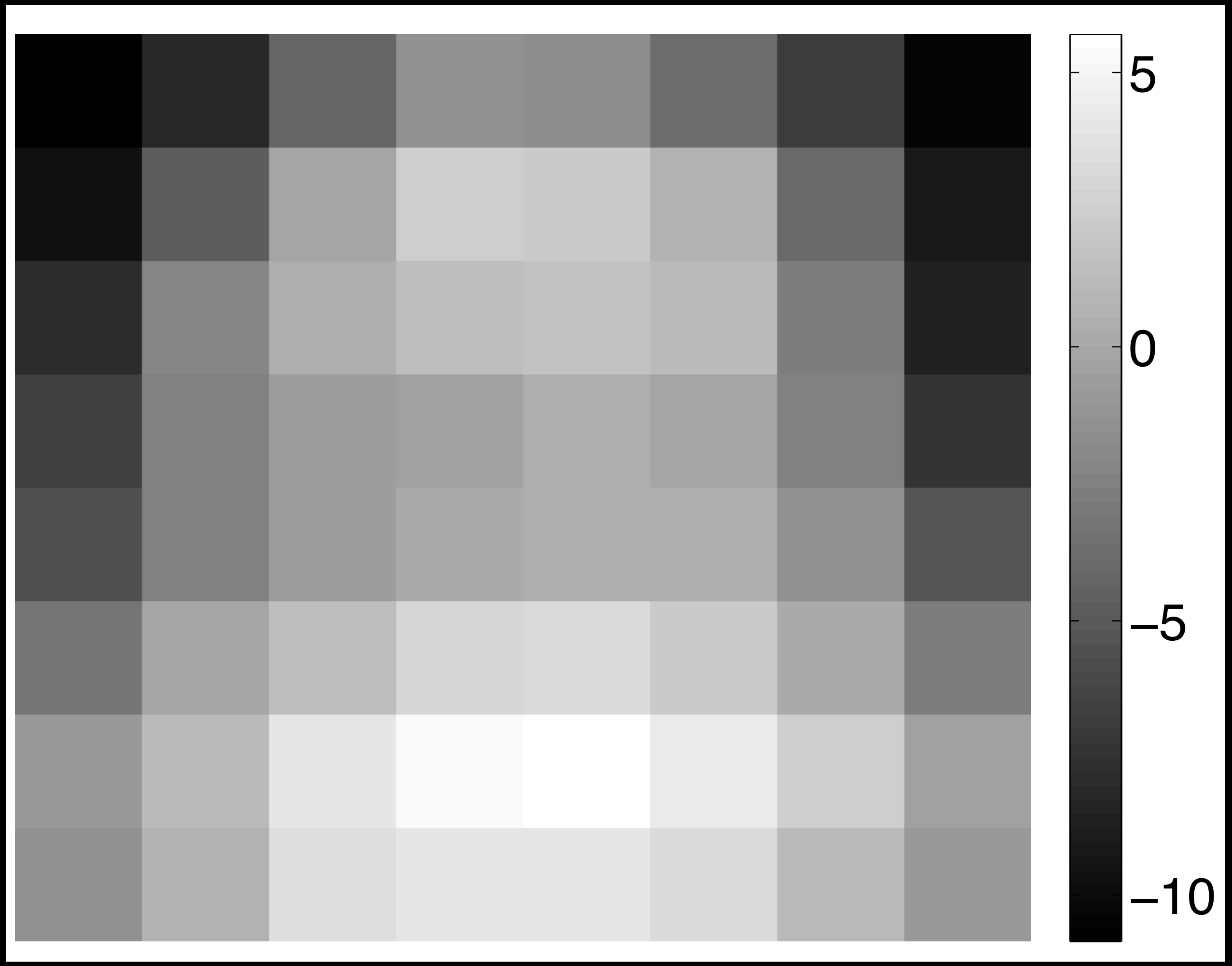
barn



Other Characteristics

Spatial Gabor filter layout.

fireworks



Summary

- Link any characteristic to any keyword.
- Fast and highly scalable:
  millions of images and thousands of keywords.
- Base for subsequent imaging applications with semantic awareness.

Semantic Image Enhancement

[Lindner et al., ACM Multimedia 2012, long paper]

Which image is better?

Which image is better?

dark

snow

Which image is better?

sand

sunset

Which image is better?

sand

sunset

No decision possible based on pixel values only.

Which image is better?

sand

sunset

No decision possible based on pixel values only.

Auto-adjust contrast/colors.

Which image is better?

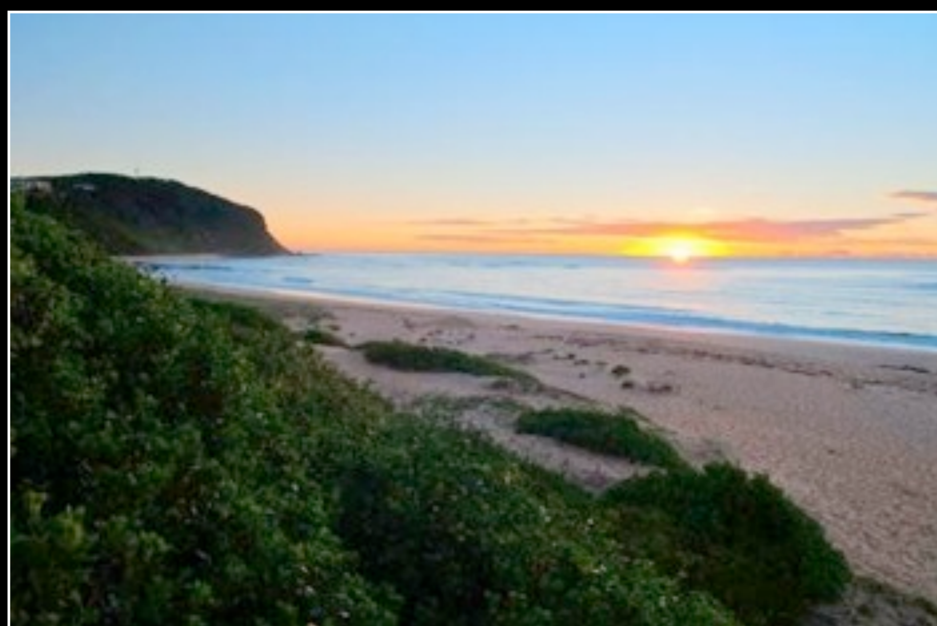
sand

sunset

No decision possible based on pixel values only.

Manual editing.

Which image is better?



sand

sunset

No decision possible based on pixel values only.

Automatic Enhancement with **Semantics.**

Today’s Solutions

- Modes:
  Camera: “portrait”, “nature”, “firework”.
  Printer: “draft”, “presentation”, “text”.

Today’s Solutions

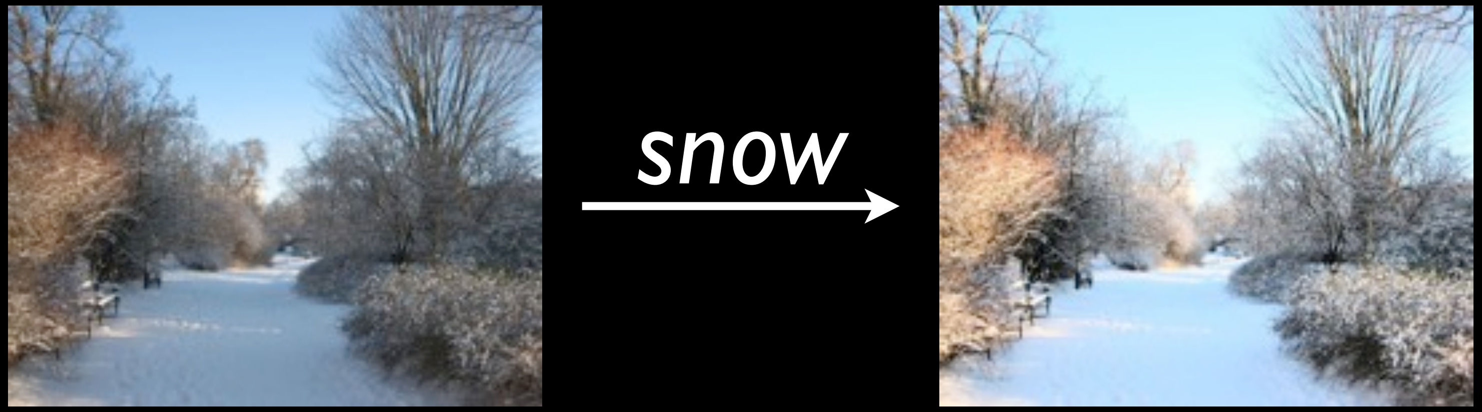
- Modes:
  Camera: “portrait”, “nature”, “firework”.
  Printer: “draft”, “presentation”, “text”.
- Classification + enhancement:
  skin, sky or other classes.
  Park et al. 06, Ciocca et al. 07, Kaufman et al. 12.

Today's Solutions

- Modes:
Camera: "portrait", "nature", "firework".
Printer: "draft", "presentation", "text".
- Classification + enhancement:
skin, sky or other classes.
Park et al. 06, Ciocca et al. 07, Kaufman et al. 12.
- **Difficult to scale to large vocabularies.**

Semantic Image Enhancement

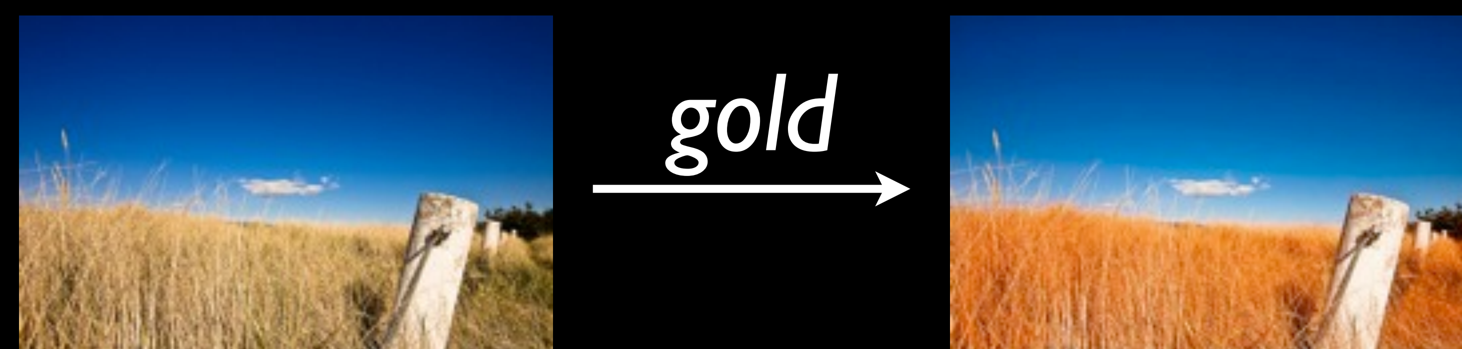
Gray scale tone mapping



Semantic Image Enhancement

Gray scale tone mapping

Color enhancement

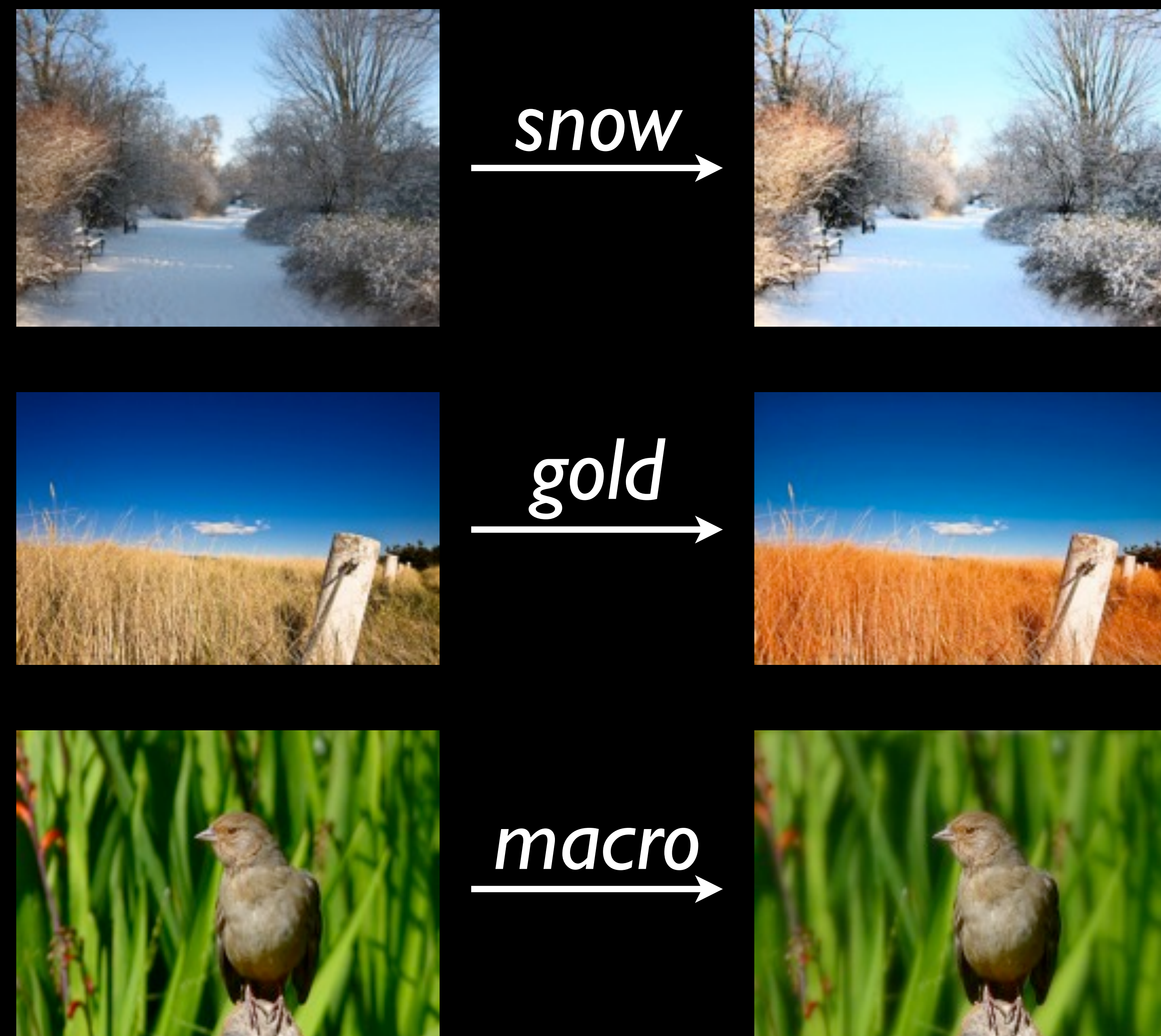


Semantic Image Enhancement

Gray scale tone mapping

Color enhancement

Change depth-of-field
[Zhuo and Sim, 2011]



Semantic Image Enhancement

Gray scale tone mapping

snow

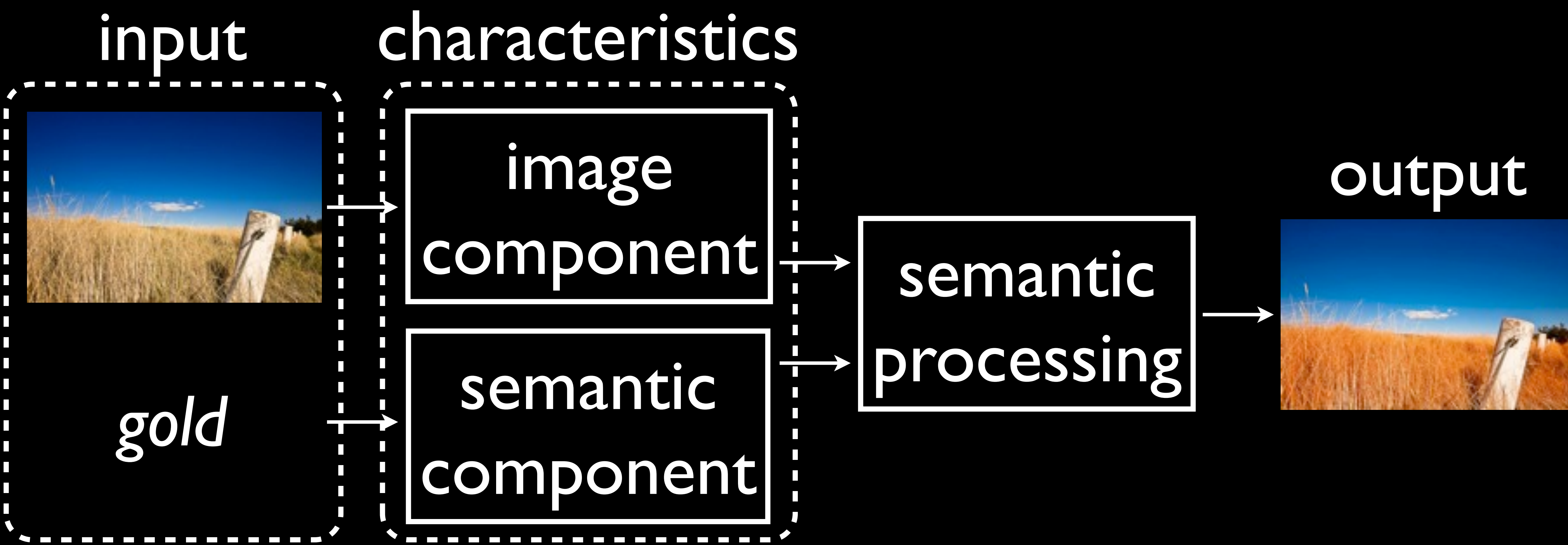
Color enhancement

gold

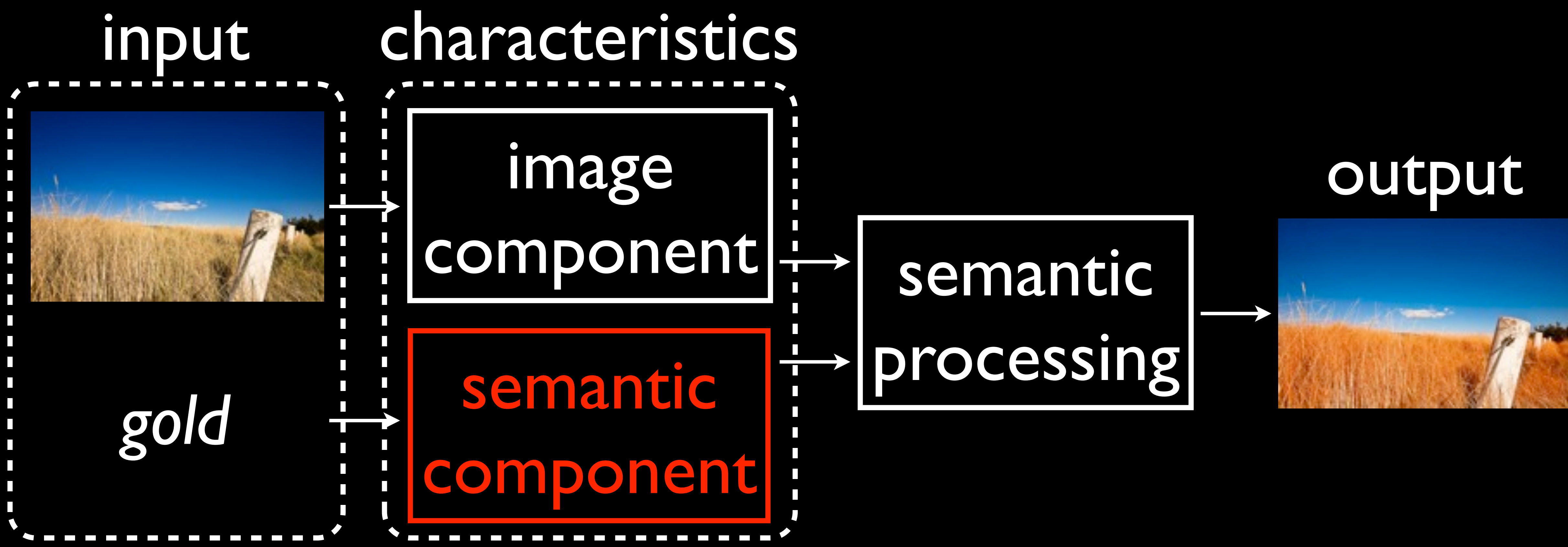
Change depth-of-field
[Zhuo and Sim, 2011]

macro

Semantic Enhancement

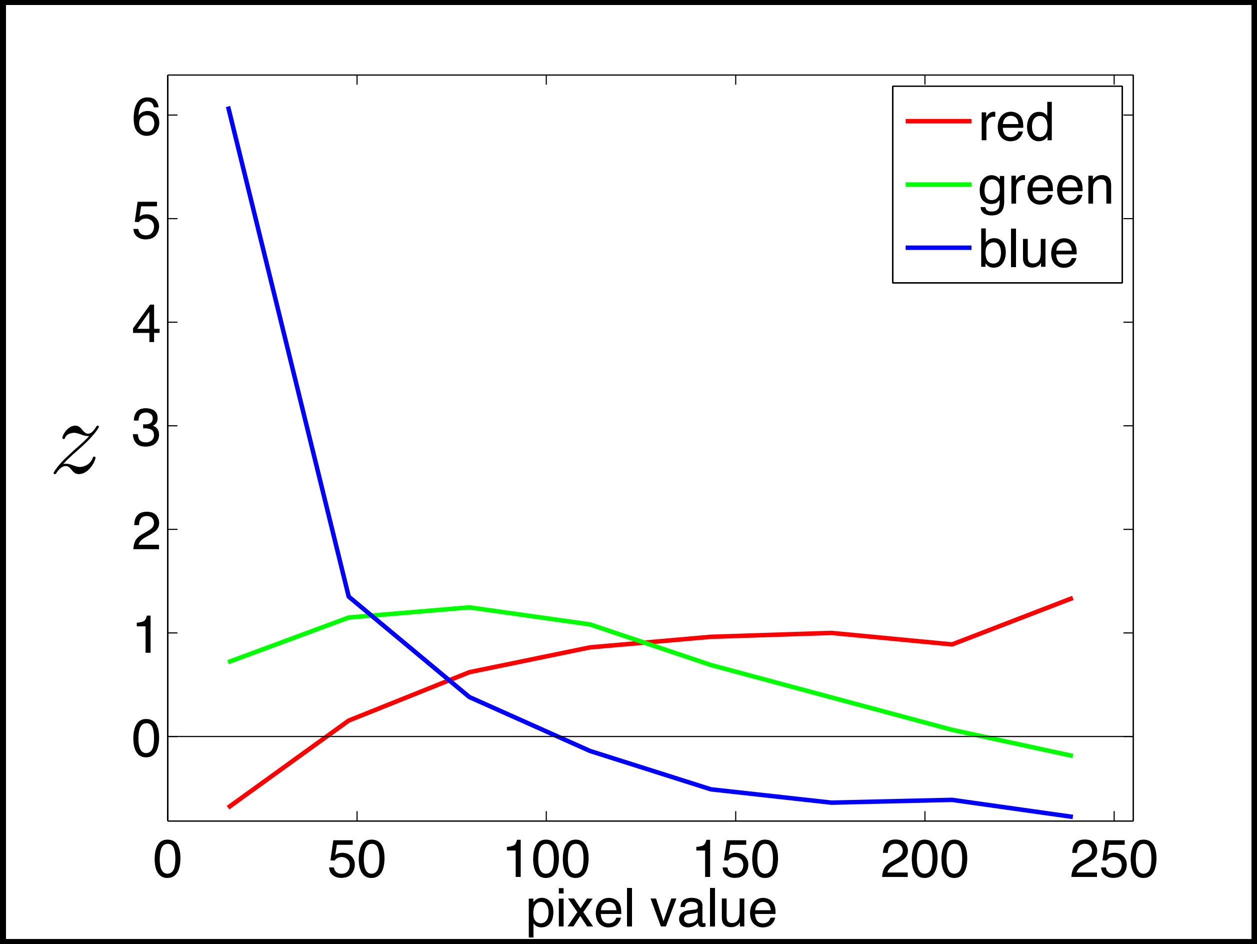


Semantic Enhancement



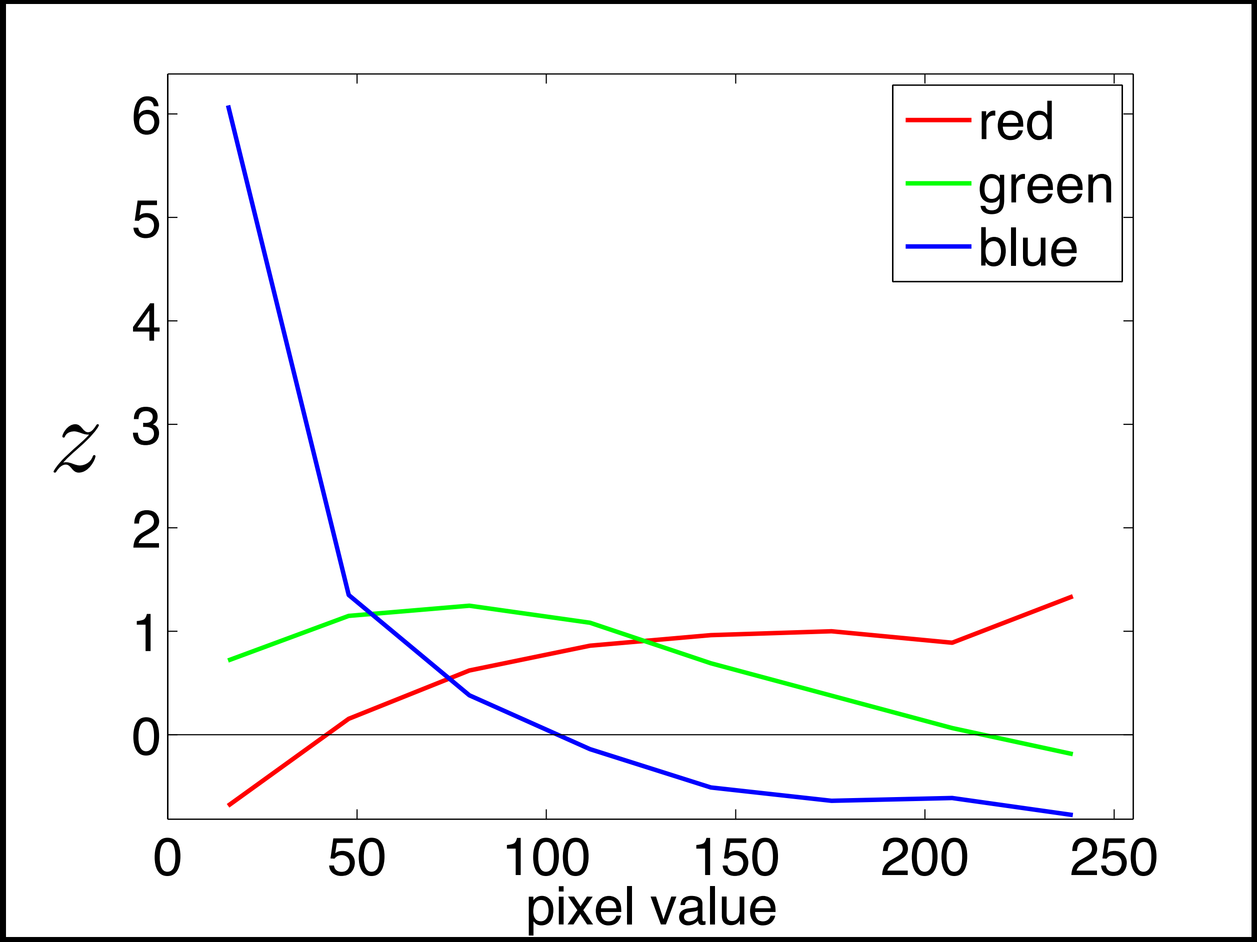
Semantic Component

significance values for *gold*



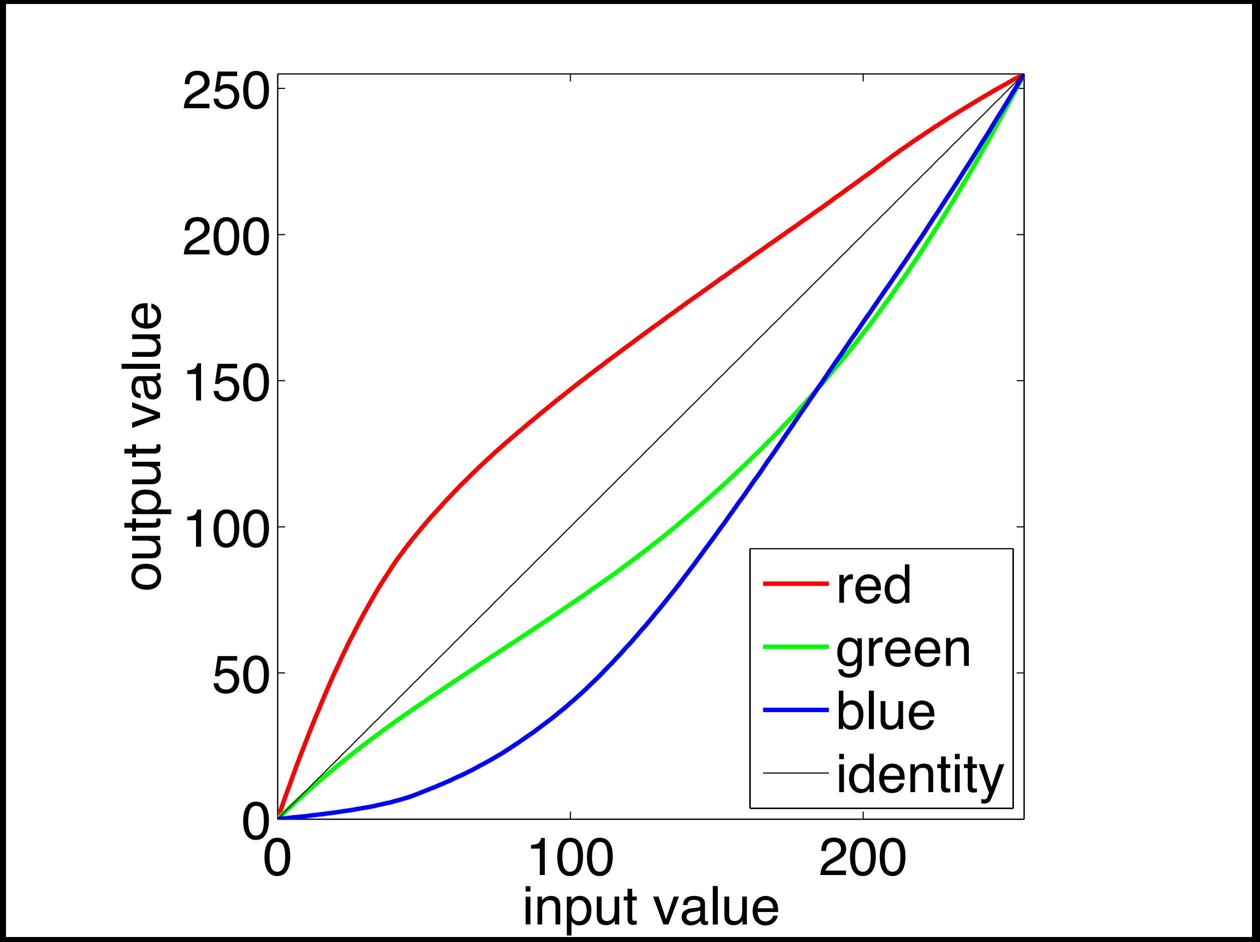
Semantic Component

significance values for *gold*



$\longrightarrow$

Tone mapping function f



$$f' = \begin{cases} 1/\left(1 + Sz\right) & \text{if } z \geq 0 \\ 1 + S|z| & \text{if } z < 0 \end{cases}$$

S global scale parameter

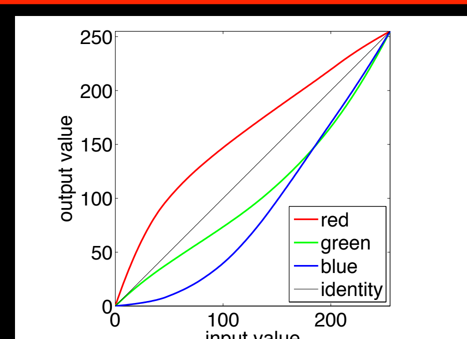
Semantic Enhancement

input

gold

characteristics

image component



semantic processing

output

Image Component

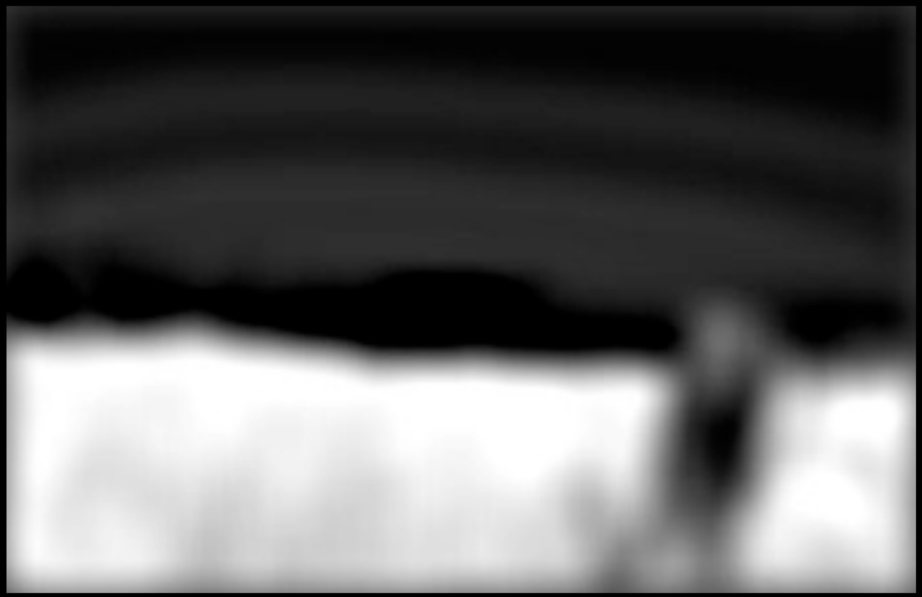
gold

Image Component

gold

$\longrightarrow$

weight map ω

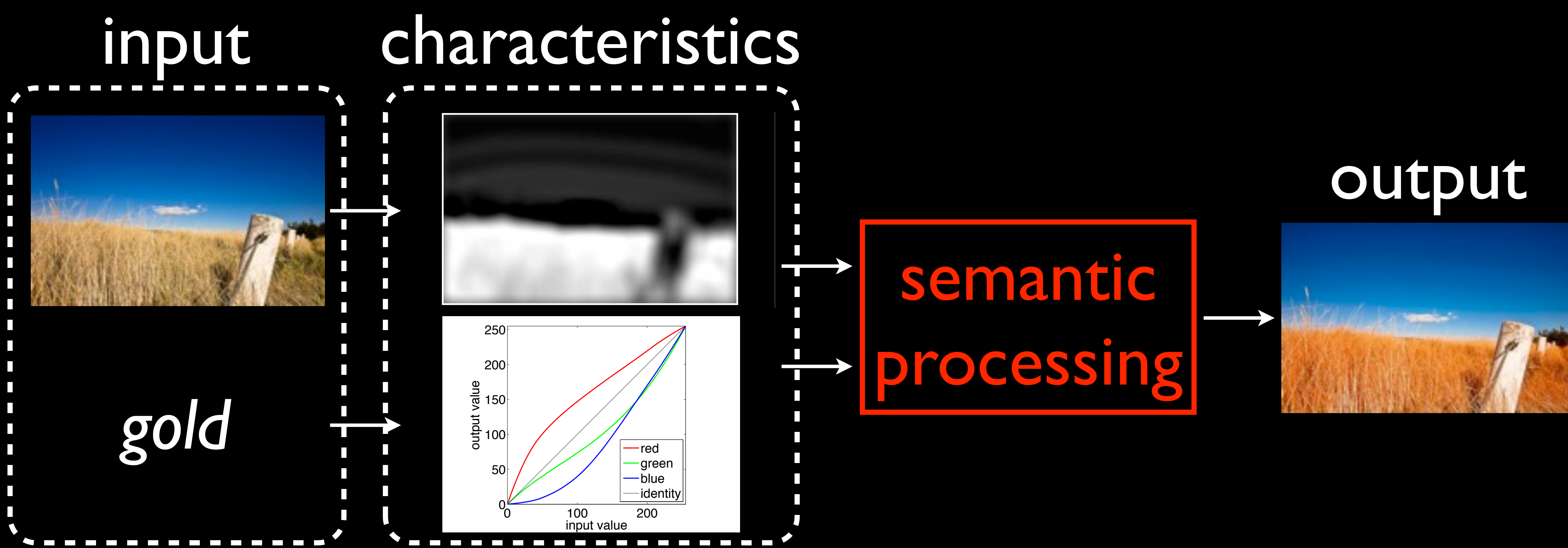


$$\omega = \left[g_\sigma * z_w\big(\,\mathrm{col}(p)\big)\right]_0^1$$

g_σ Gaussian blurring kernel (1% of image diagonal)

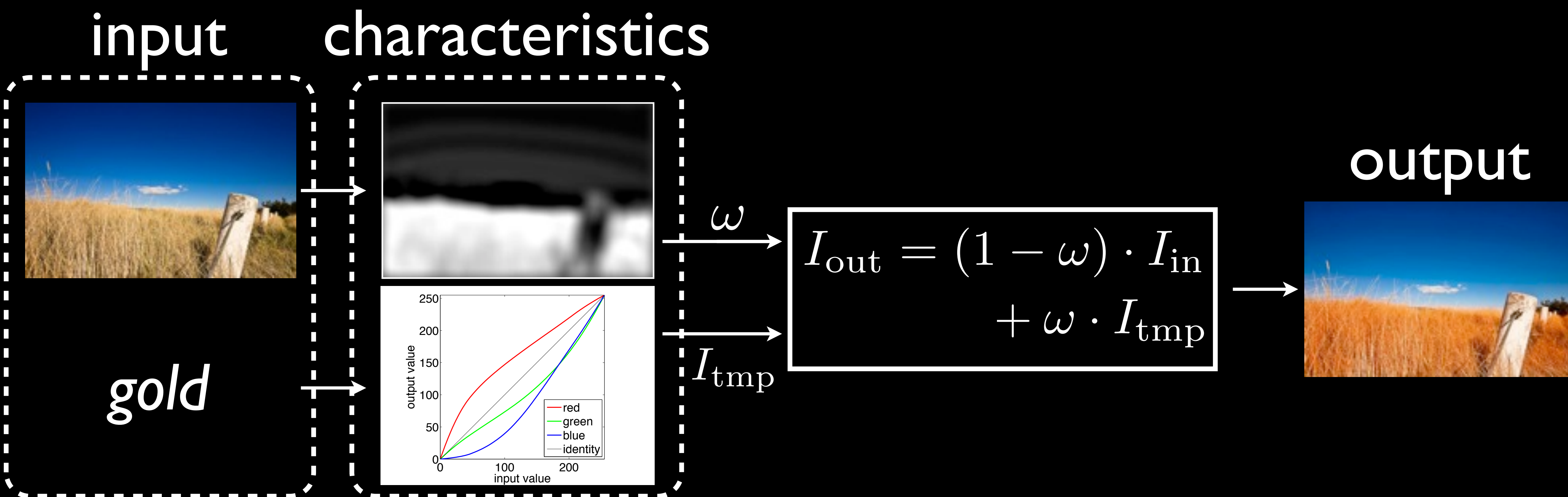
$[\,\cdot\,]_0^1$ normalization operator

Semantic Enhancement



Semantic Enhancement

Enhance relevant characteristics in relevant regions.



sand

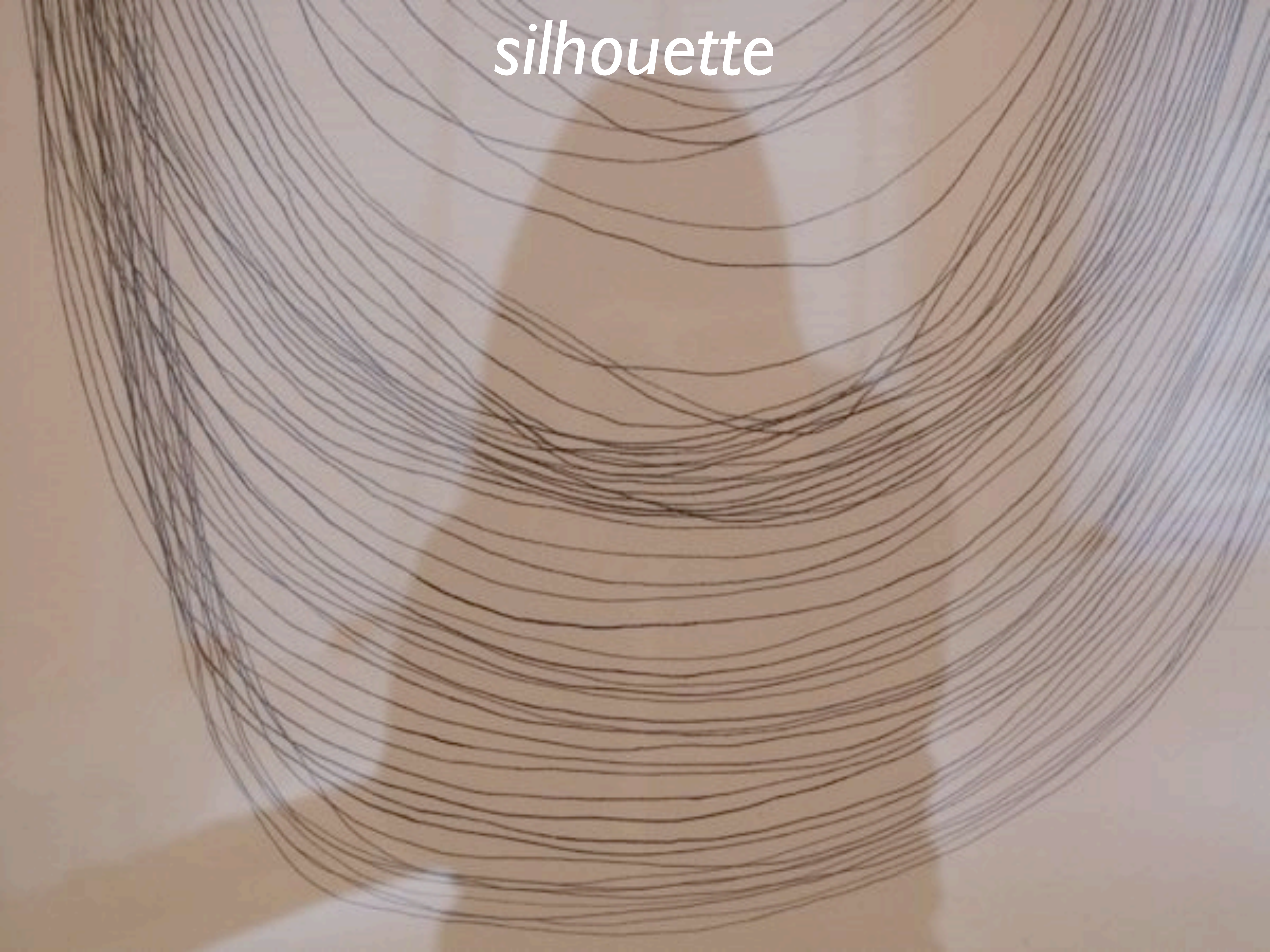
sand


snow

snow


dark


dark


silhouette

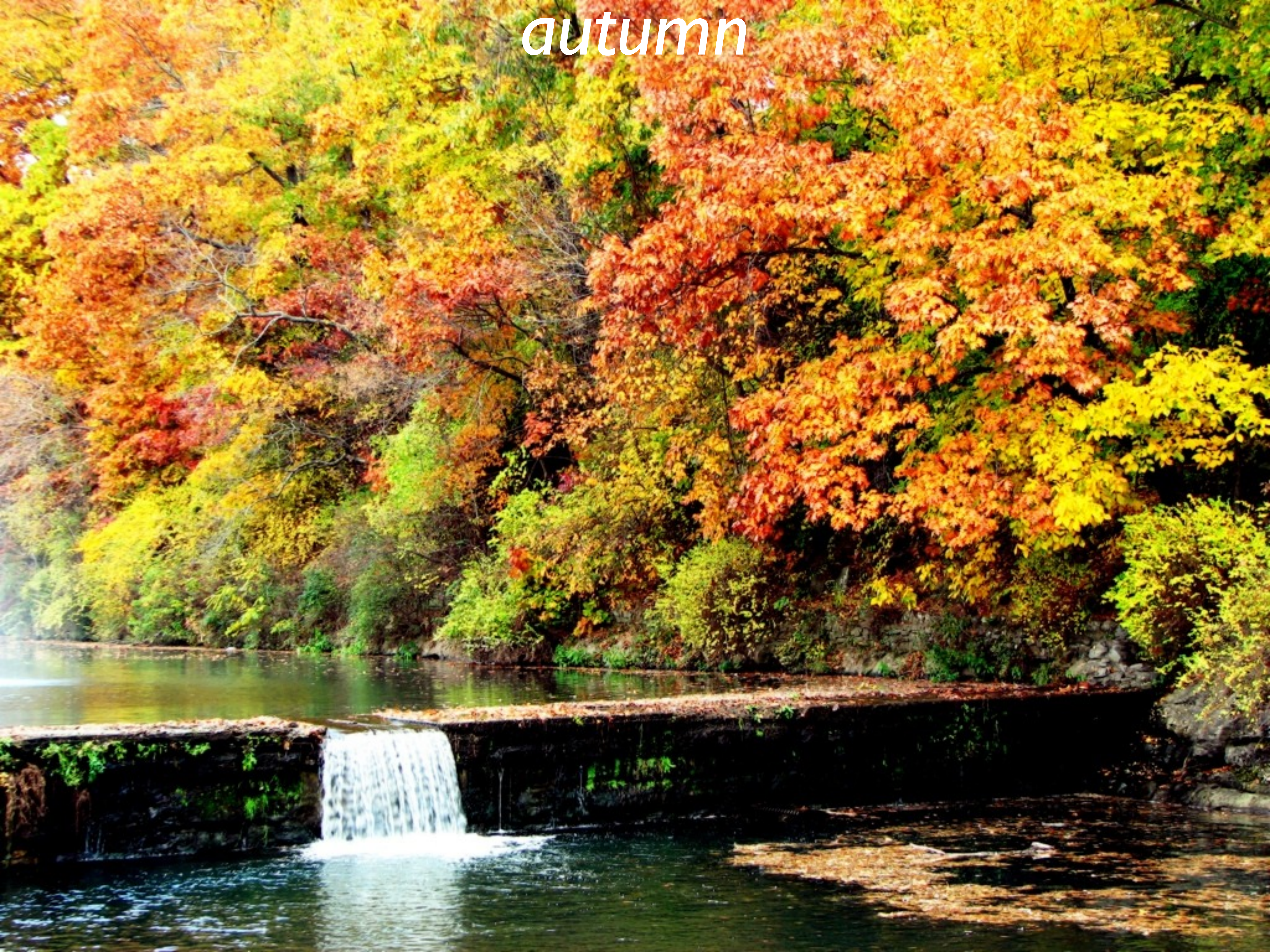

silhouette

sunset

sunset

grass

grass


autumn


autumn

strawberry

strawberry


sky


sky

banana


banana


macro


macro


flower


flower


macro

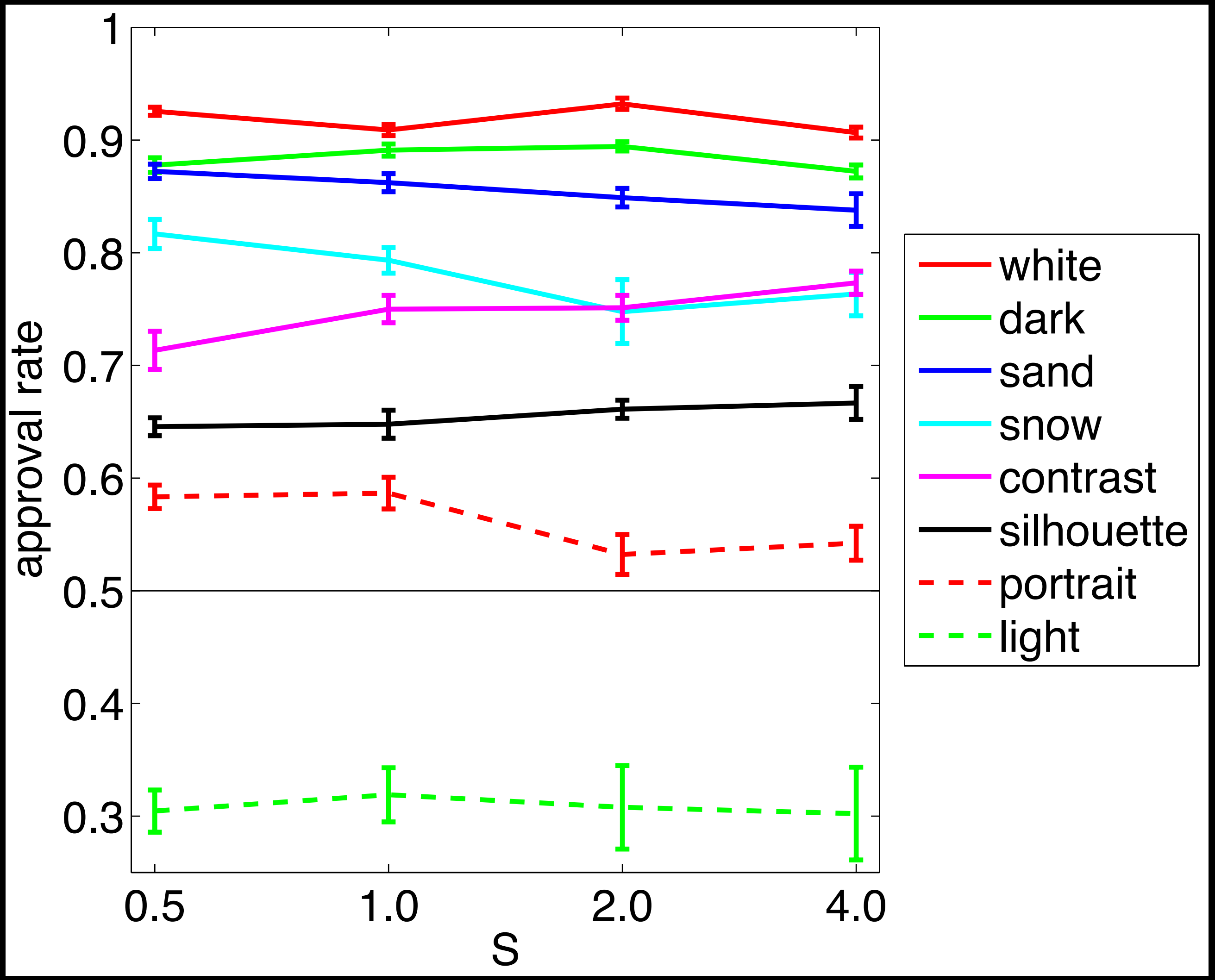

macro

Psychophysical Experiment

sand

8 keywords, 30 images, $S = \{0.5,\ 1,\ 2,\ 4\}$, 30 observers
=28'800 image comparisons

Amazon Mechanical Turk

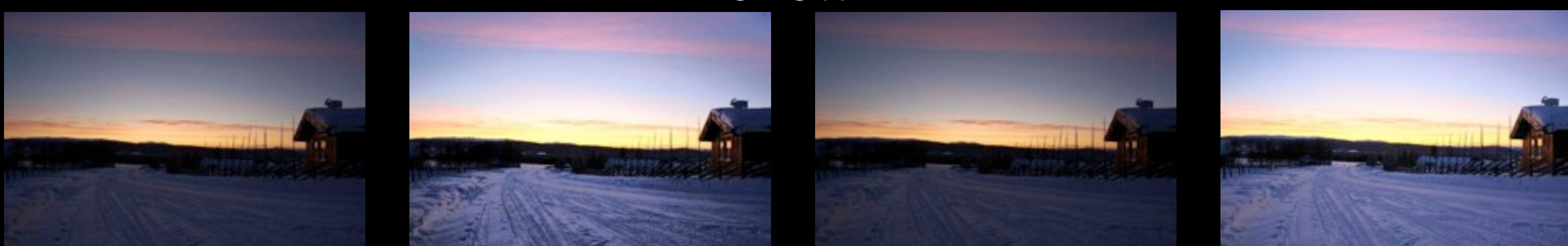


8 keywords, 30 images, $S = \{0.5,\ 1,\ 2,\ 4\}$, 30 observers =28'800 image comparisons

Reciprocal Keywords

dark

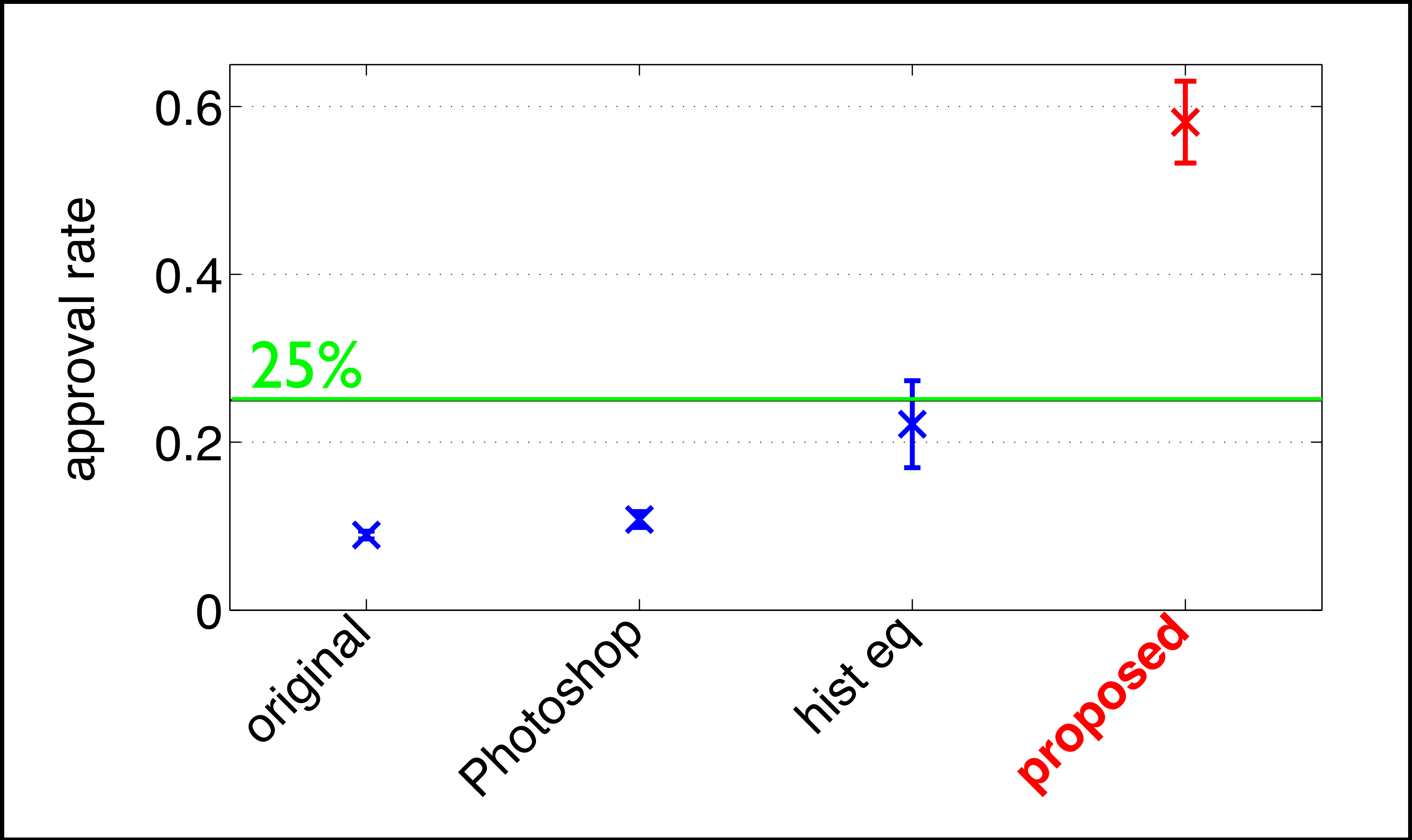
snow



original | histogram equalization | Photoshop auto contrast | proposed

29 image/keyword pairs, 40 observers.

Reciprocal Keywords



29 image/keyword pairs, 40 observers.

Limitations and Future Work

- Keyword without significant characteristics: *friendship, boredom, happy, statue.*

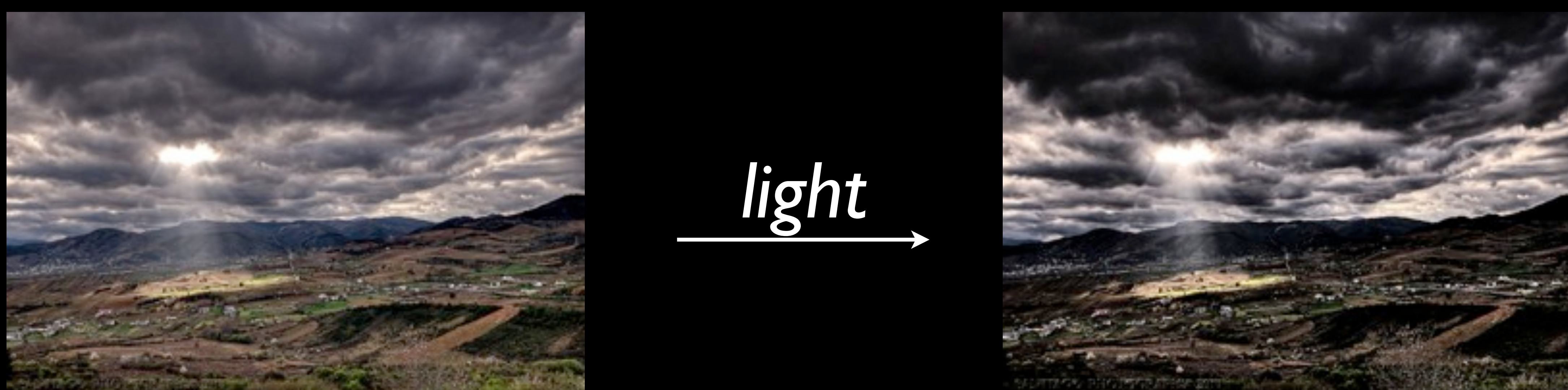
Limitations and Future Work

- Keyword without significant characteristics: *friendship, boredom, happy, statue*.
- Keywords with conflicting interpretations.

light →

Limitations and Future Work

- Keyword without significant characteristics: *friendship, boredom, happy, statue.*
- Keywords with conflicting interpretations.



- Multiple or machine-generated keywords.

Automatic Color Naming

[Lindner et al., IS&T CIC 2012]
MERL best student paper award

[Lindner et al., IS&T CGIV 2012]

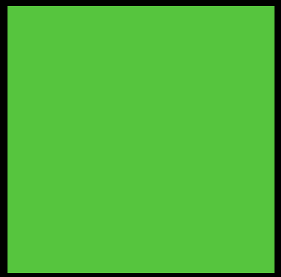
Introduction

Standard psychophysical color naming experiment:

Introduction

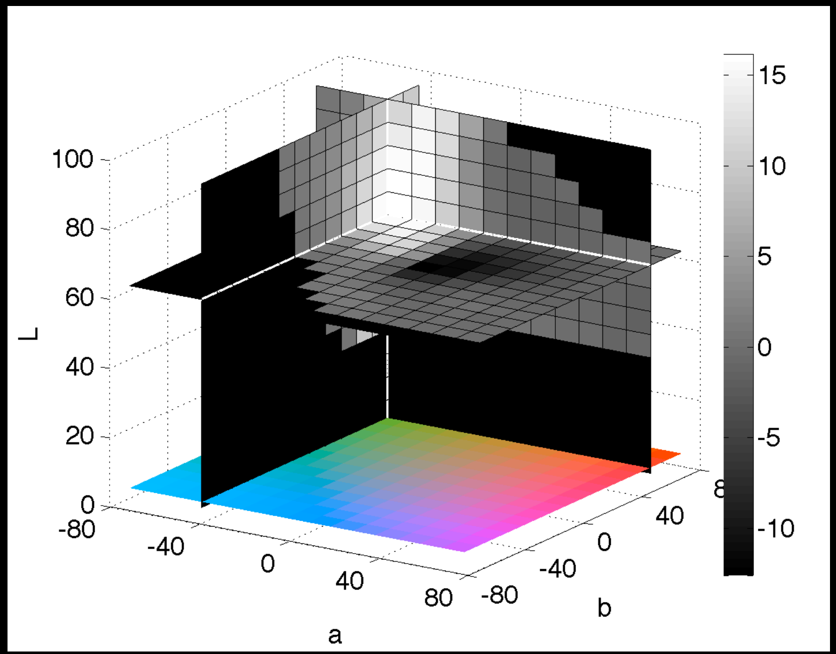
Standard psychophysical color naming experiment:

green ⟷ observer



Our approach:

green → statistical framework



9000+ Color Names

- XKCD color survey, psychophysical experiment.

9000+ Color Names

- XKCD color survey, psychophysical experiment.
- 950 English color names + color values.

9000+ Color Names

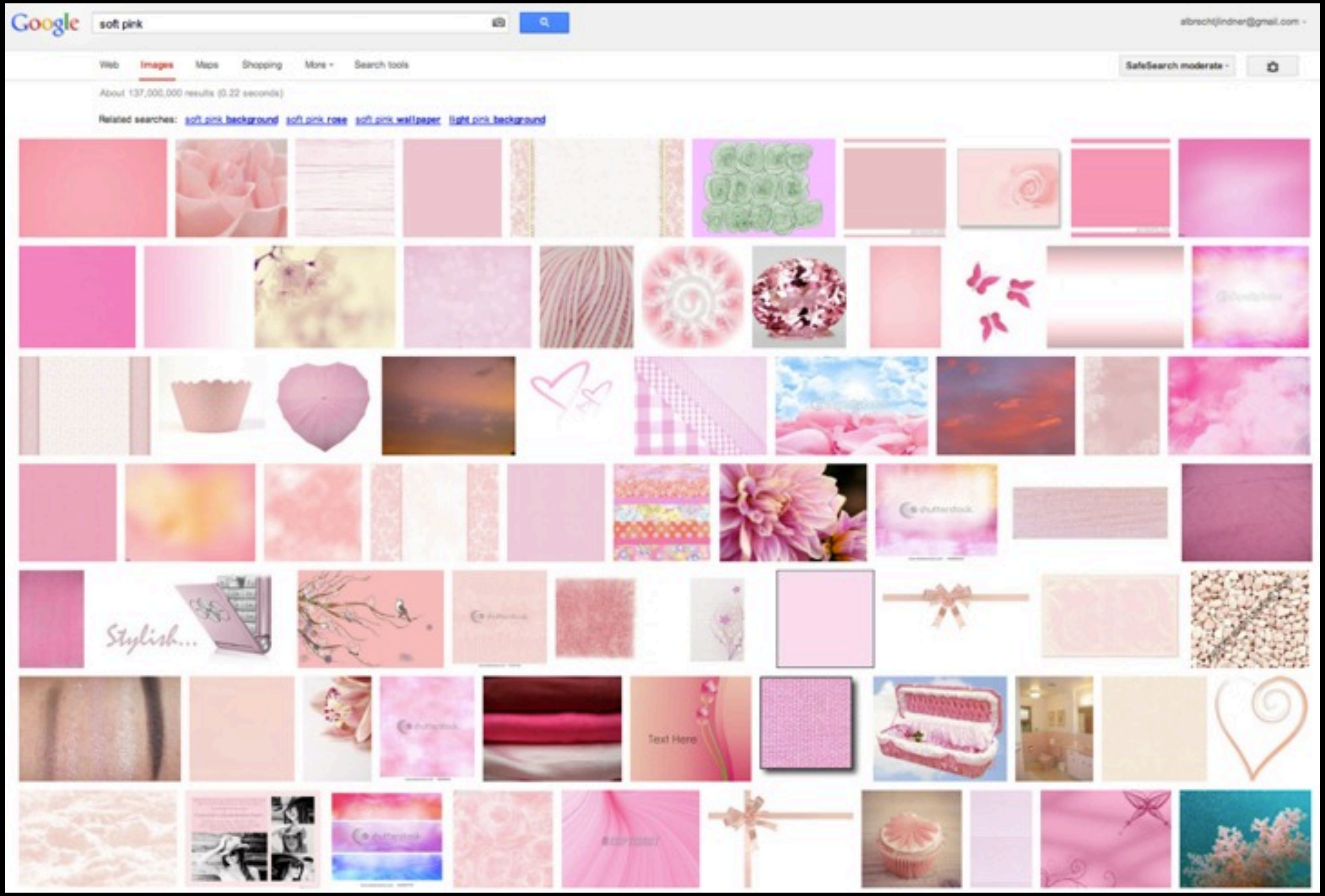
- XKCD color survey, psychophysical experiment.
- 950 English color names + color values.
- Translate to 9 other languages:
  Chinese, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish.

9000+ Color Names

- XKCD color survey, psychophysical experiment.
- 950 English color names + color values.
- Translate to 9 other languages: Chinese, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish.
- Example: 柔和的粉红色, *soft pink, rose tendre, sanftes pink, rosa tenue,* ソフトピンク, 부드러운 녹색, *rosa suave, нежно розовый, rosa suave.*

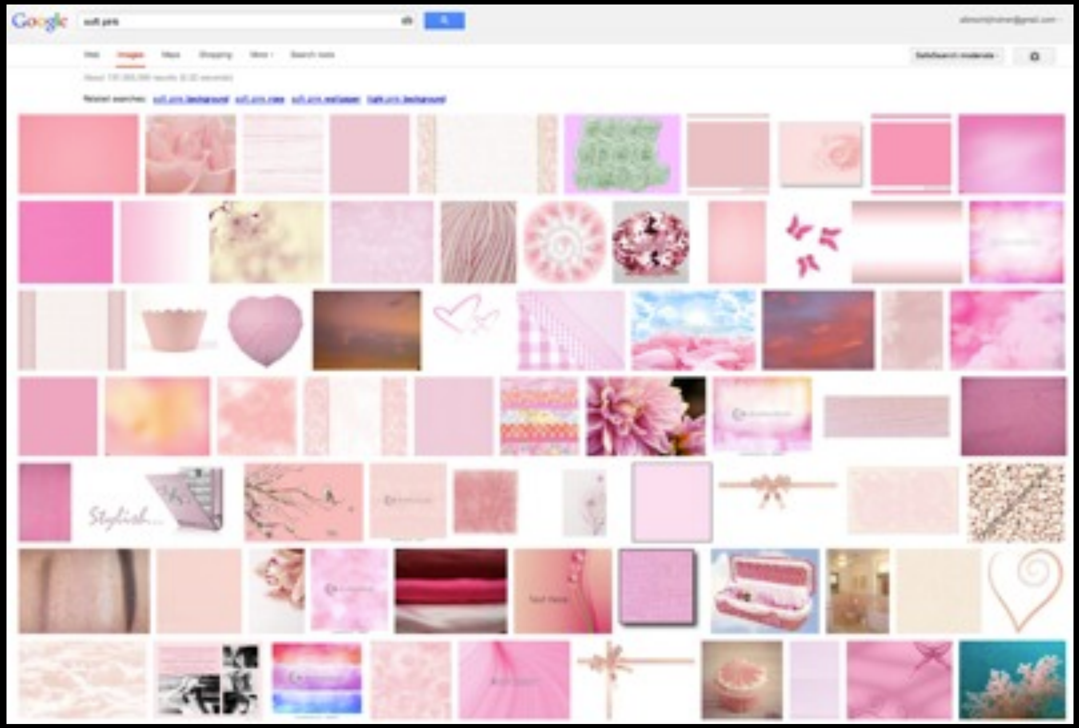
Data Acquisition

Google Image: *soft pink*



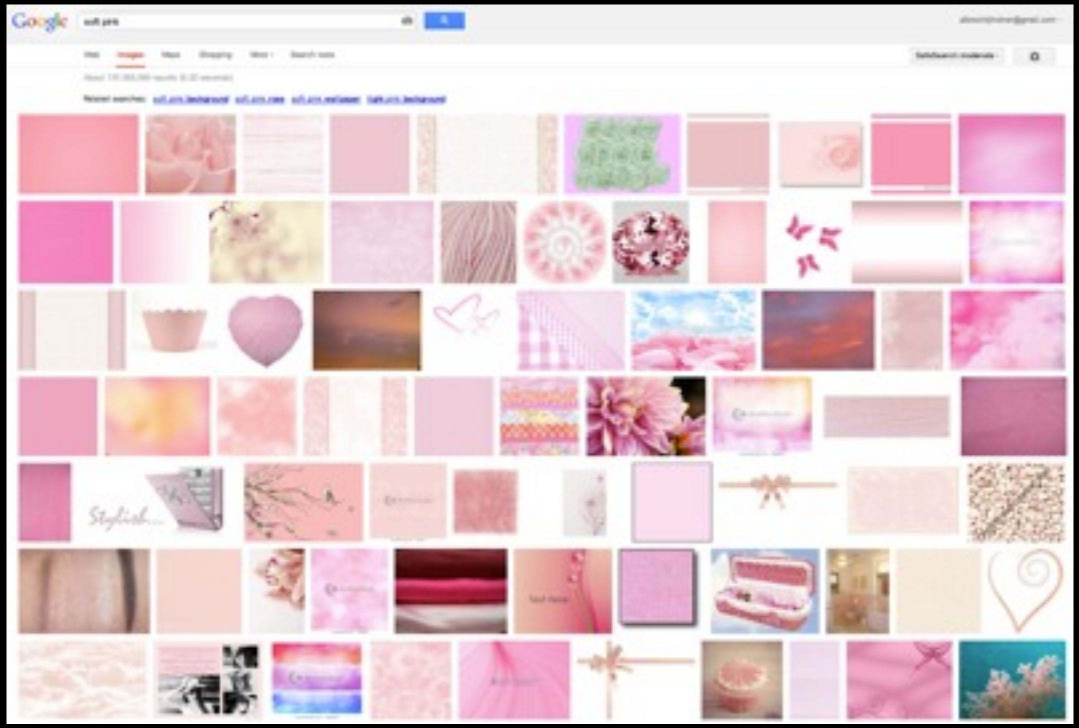
Data Acquisition

Google Image: *soft pink*



Data Acquisition

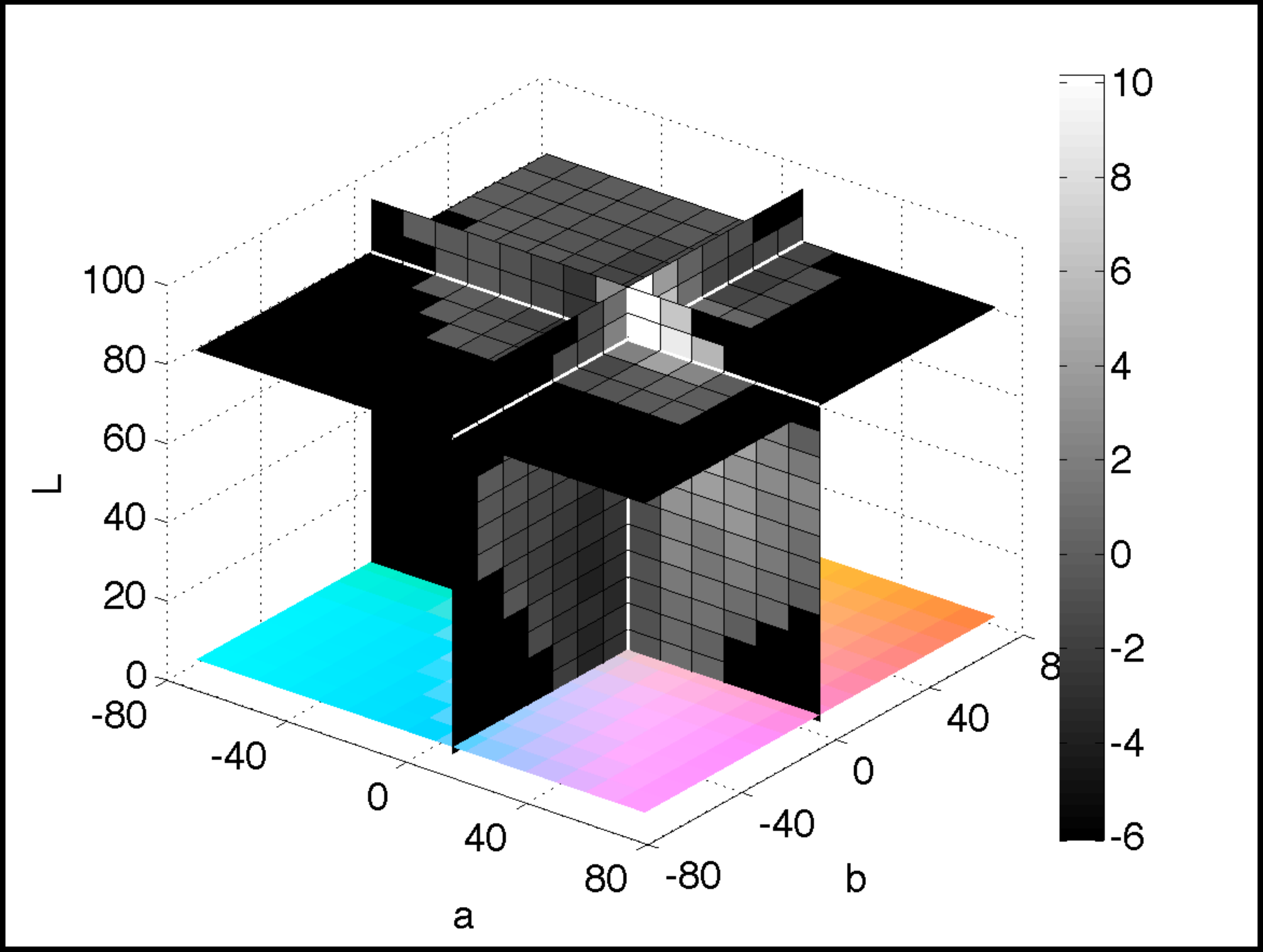
Google Image: *soft pink*



- 100 images per color name.
- Language and country restrict.
- Assume sRGB encoding.
- Almost 1M images.

z Distribution

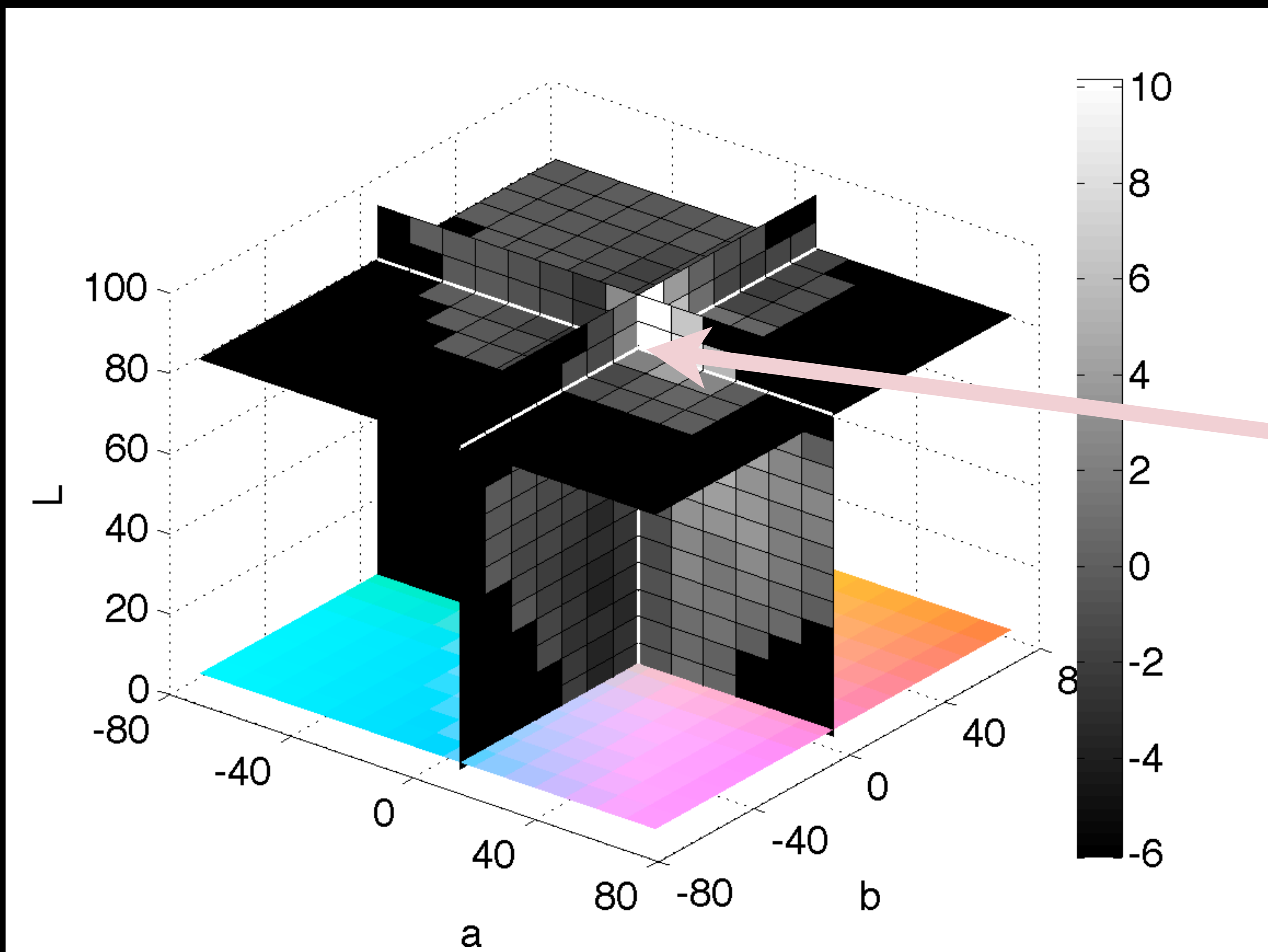
soft pink, English



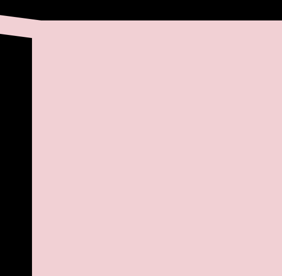
- CIELAB histogram 15x15x15 bins.

z Distribution

soft pink, English



- CIELAB histogram 15x15x15 bins.



sRGB: 238, 197, 203

Soft Pink

柔和的粉红色, cn	ソフトピンク, jp
soft pink, en	부드러운 녹색, ko
rose tendre, fr	rosa suave, pt
sanftes pink, de	нежно розовый, ru
rosa tenue, it	rosa suave, es

Soft Pink

柔和的粉红色, cn

soft pink, en

rose tendre, fr

sanftes pink, de

rosa tenue, it

ソフトピンク, jp

부드러운 녹색, ko

rosa suave, pt

rosa suave, es

Soft Pink

柔和的粉红色, cn

soft pink, en

rose tendre, fr

sanftes pink, de

rosa tenue, it

ソフトピンク, jp

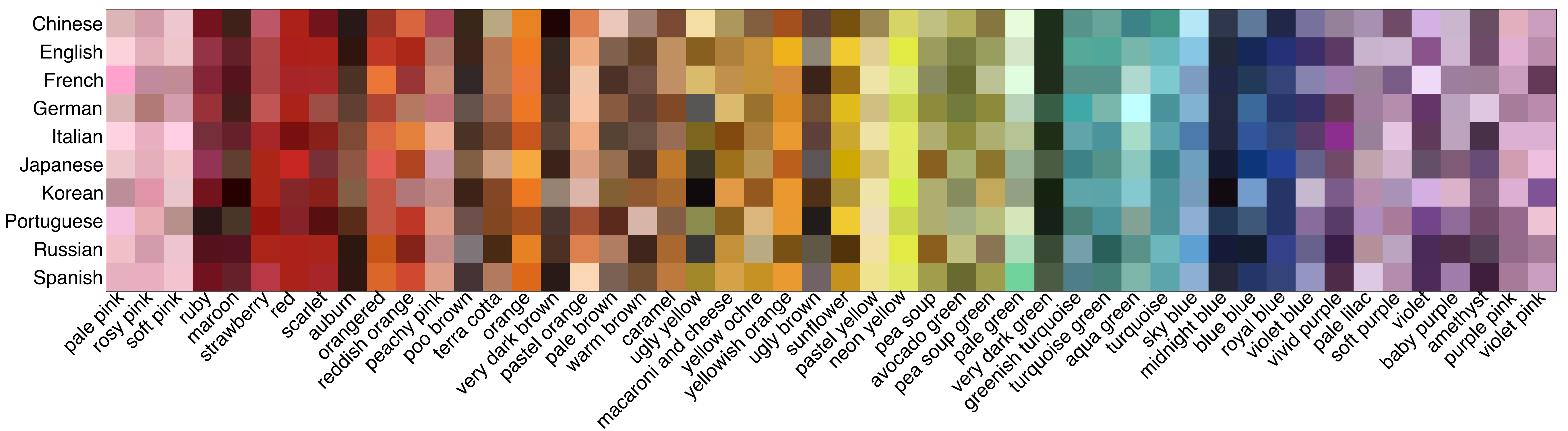
부드러운 녹색, ko

rosa suave, pt

rosa suave, es

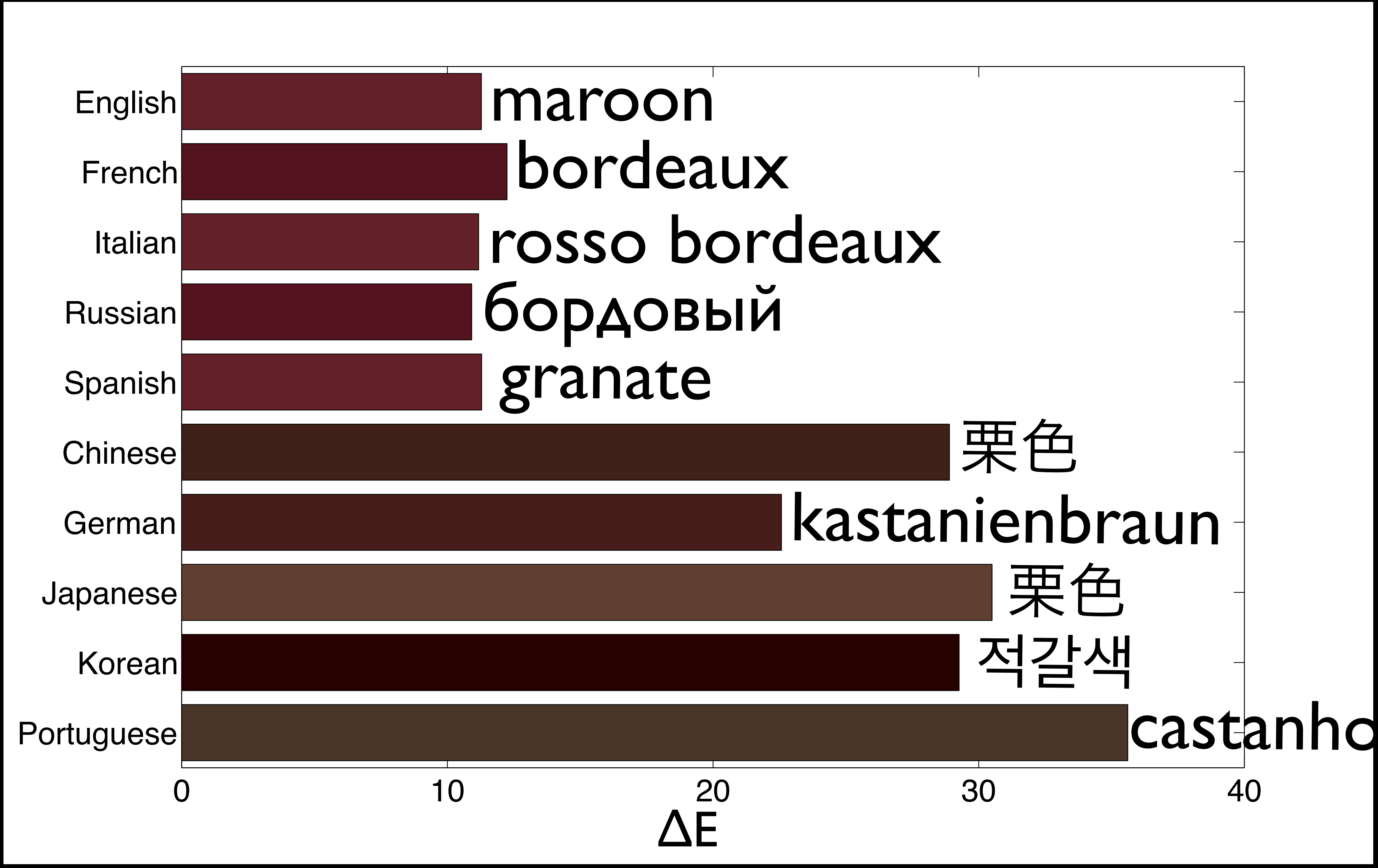
Language and **country** restrict.

Color Estimations



Accuracy for *maroon*

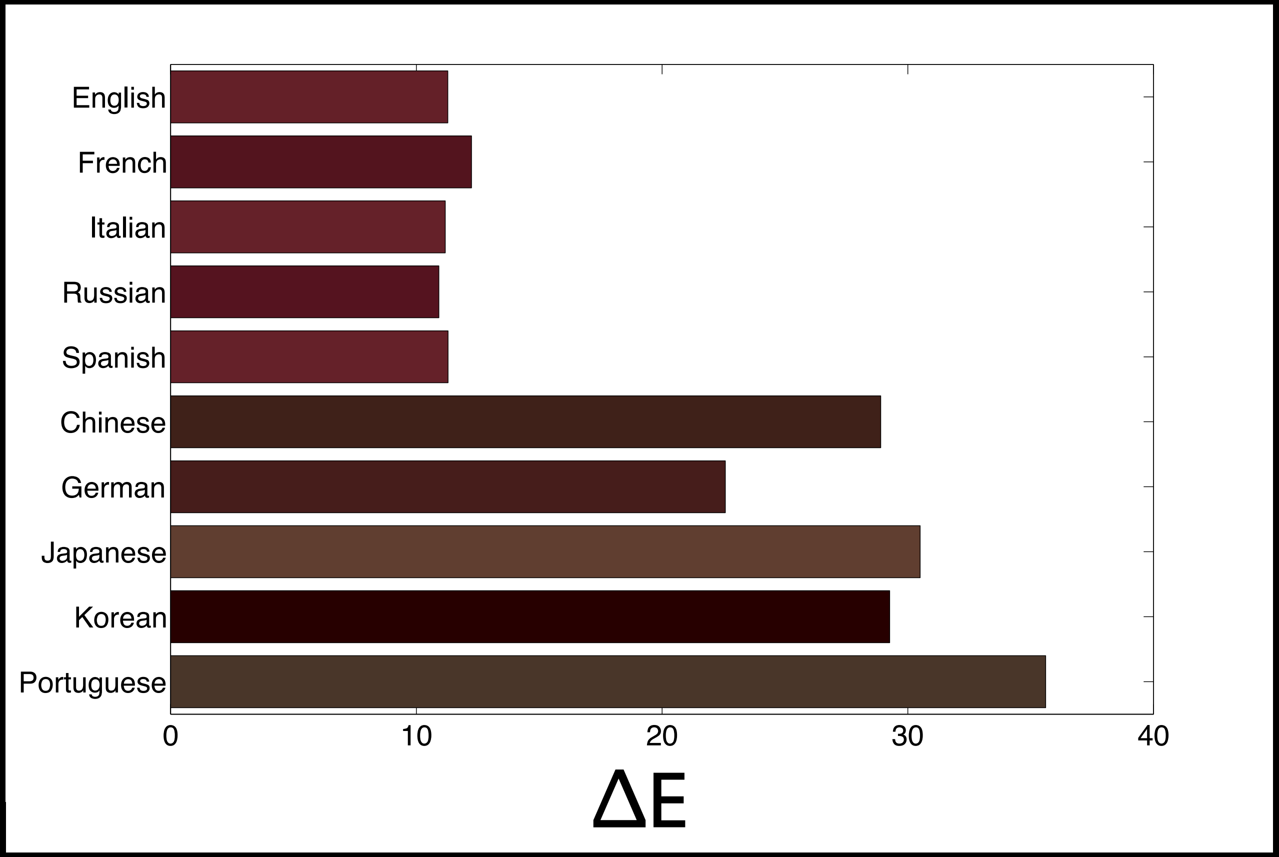
ΔE distances to **English** XKCD ground truth.



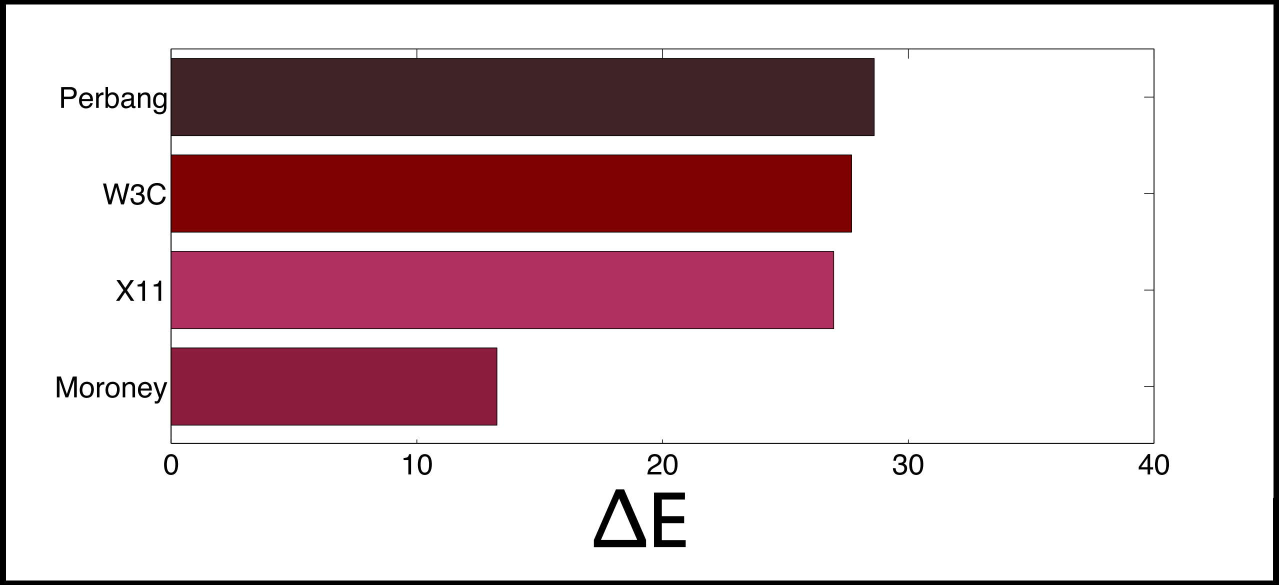
Accuracy for *maroon*

ΔE distances to **English** XKCD ground truth.

maroon, ours

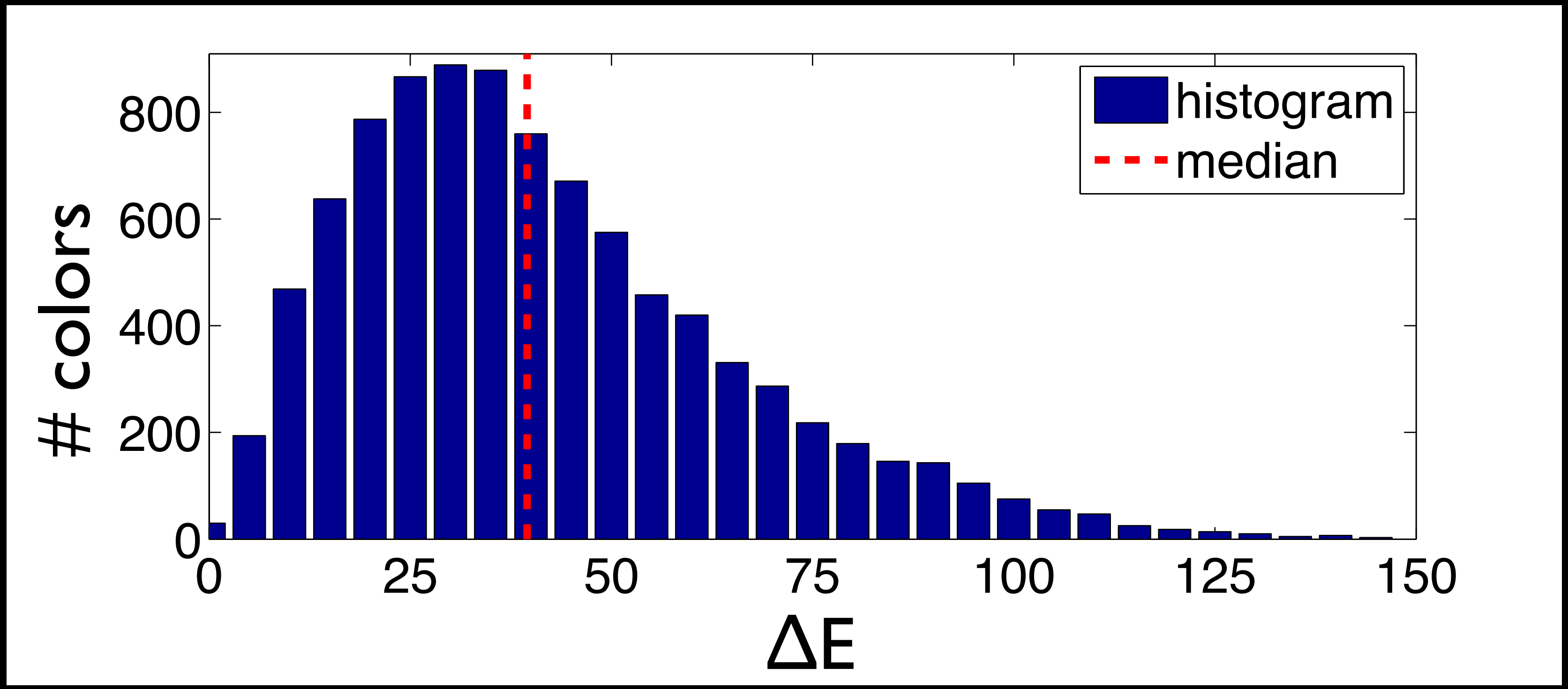


maroon, others



Accuracy

ΔE distances to **English** XKCD ground truth.



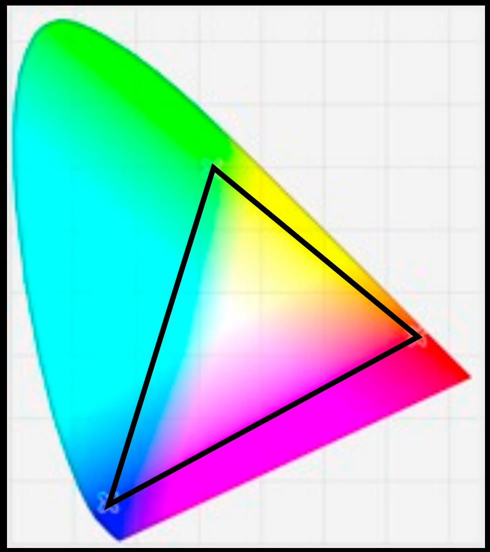
Our estimations are reasonably accurate considering:

- Disagreements between other databases.
- Language translations.

DEMO

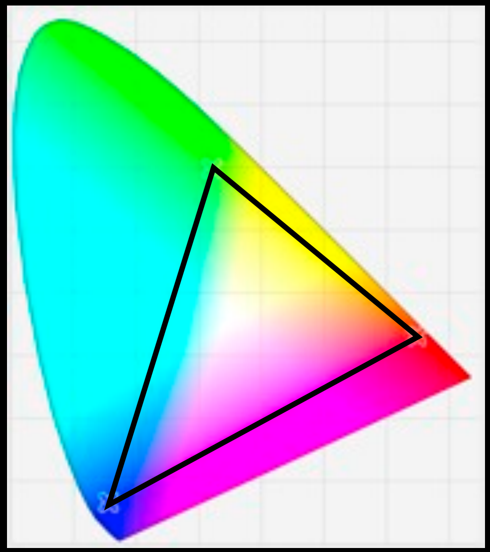
Limitations and Future Work

- No colors outside gamut.
- Only languages that have active online community.



Limitations and Future Work

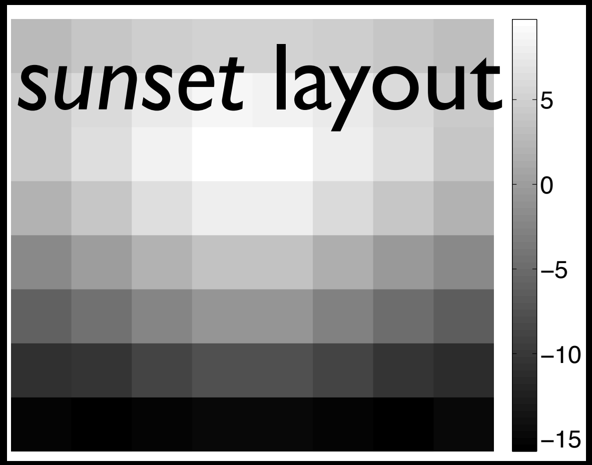
- No colors outside gamut.
- Only languages that have active online community.



- Color palettes.
- Color of an entire paragraph/text.

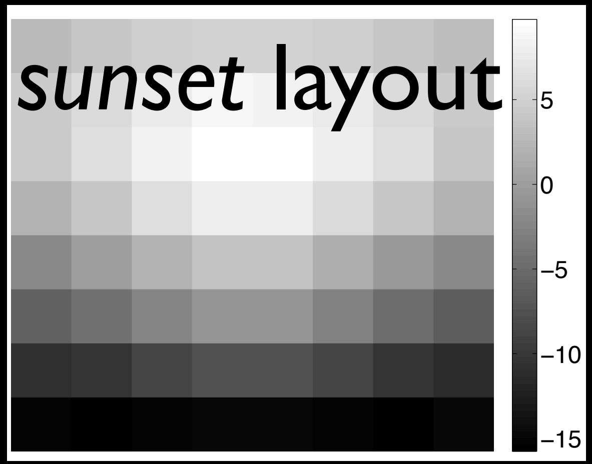
Romeo & Juliet ⟶ 

Conclusions & Future Work

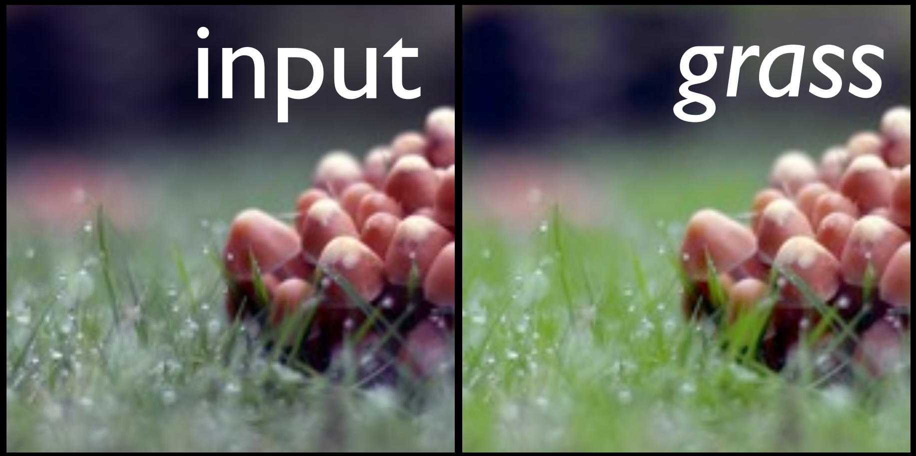


Easily scalable statistical framework.

Conclusions & Future Work

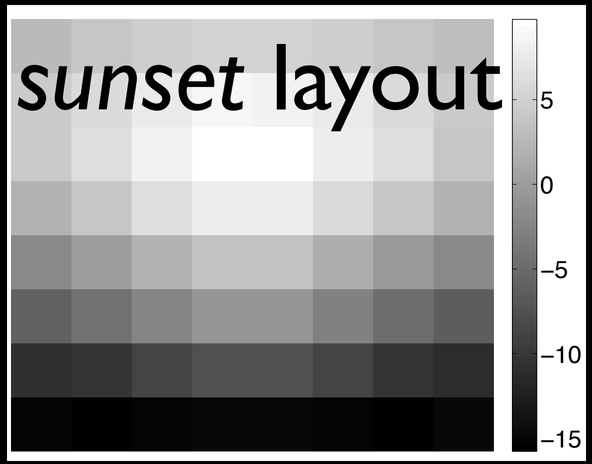


Easily scalable statistical framework.

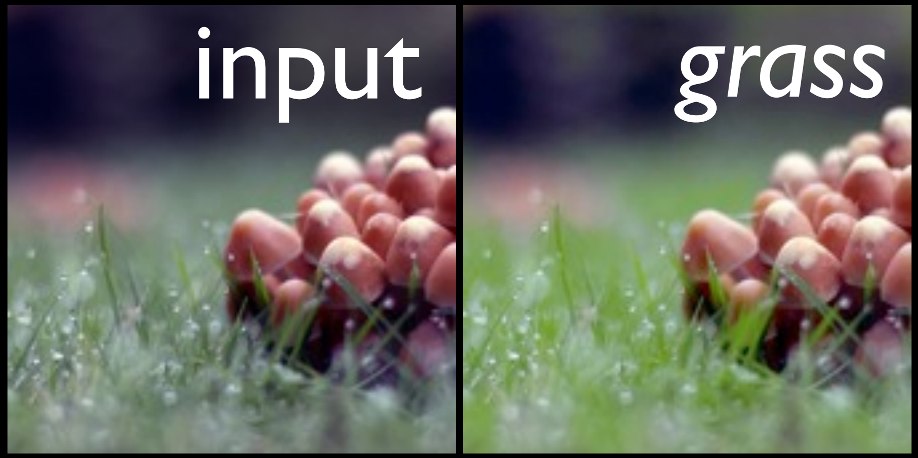


Semantic image enhancement for tone-mapping, color and depth-of-field.

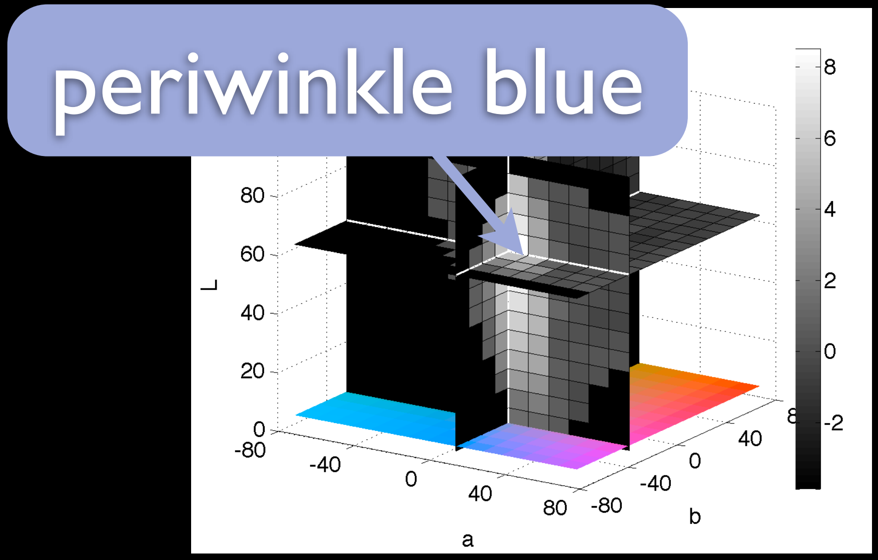
Conclusions & Future Work



Easily scalable statistical framework.



Semantic image enhancement for tone-mapping, color and depth-of-field.



Automatic color naming and an interactive online color thesaurus.

Conclusions & Future Work

- Multi-dimensional significance tests.

Conclusions & Future Work

- Multi-dimensional significance tests.
- Multiple keywords and word sense disambiguation.

Conclusions & Future Work

- Multi-dimensional significance tests.
- Multiple keywords and word sense disambiguation.
- Enhancement for other characteristics, specific devices, people with vision deficiencies, movies, etc.

Conclusions & Future Work

- Multi-dimensional significance tests.
- Multiple keywords and word sense disambiguation.
- Enhancement for other characteristics, specific devices, people with vision deficiencies, movies, etc.
- Color palettes.

Conclusions & Future Work

- Multi-dimensional significance tests.
- Multiple keywords and word sense disambiguation.
- Enhancement for other characteristics, specific devices, people with vision deficiencies, movies, etc.
- Color palettes.
- Broaden to other signals such as sound or gestures.

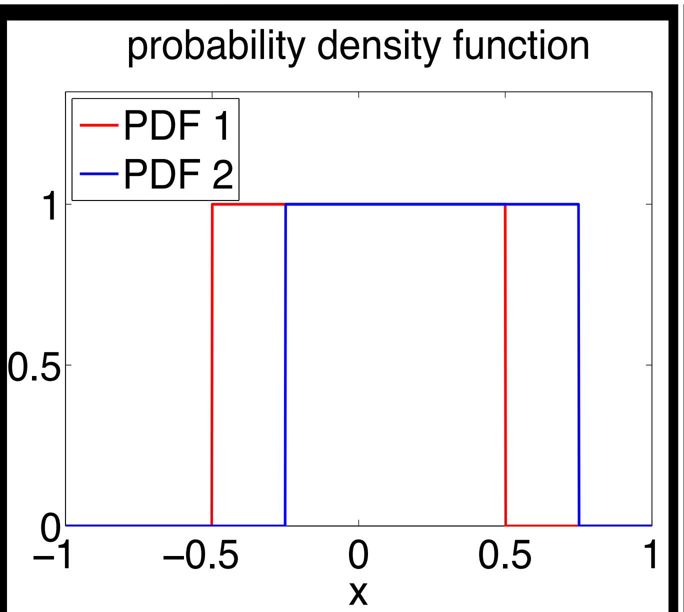
Thank you for your attention.

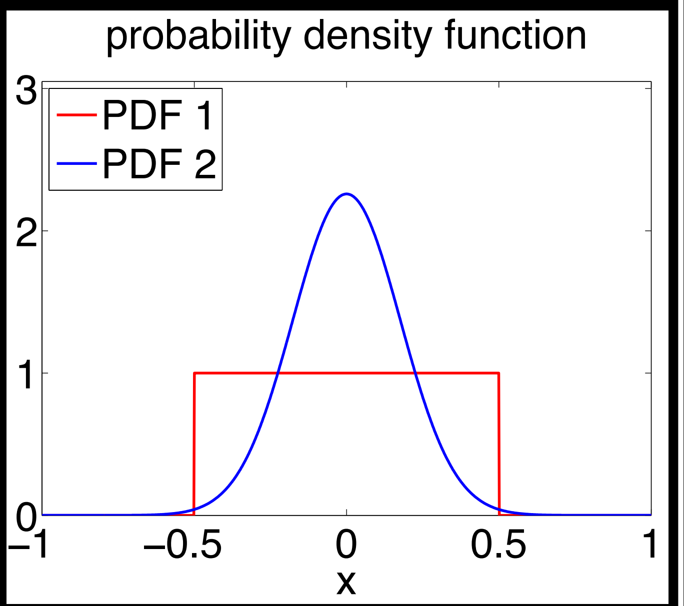
Q&A

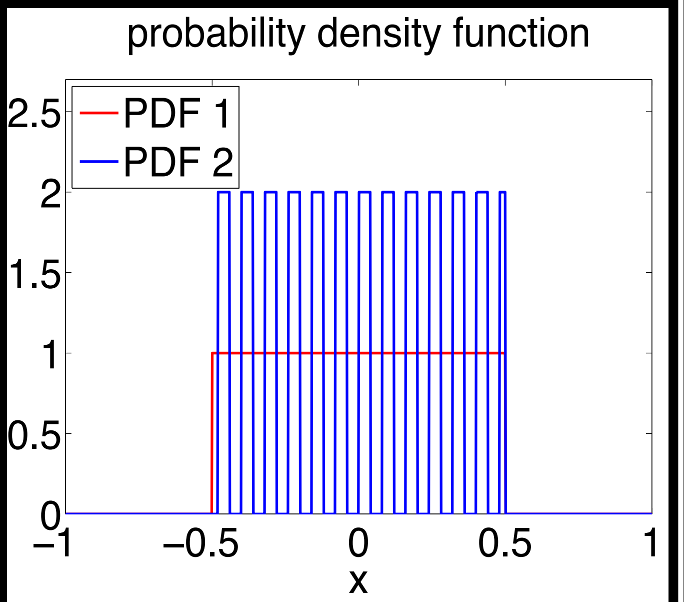
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

- Du-Sik Park, Youngshin Kwak, Hyunwook Ok and Chang-Yeong Kim, *Preferred skin color reproduction on the display*, JEI, 2006.
- Gianluigi Ciocca, Claudio Cusano, Francesca Gasparini and Raimondo Schettini, *Content Aware Image Enhancement*, Artificial Intelligence and Human-Oriented Computing, 2007.
- Liad Kaufman. Dani Lischinski and Michael Werman, *Content-Aware Automatic Photo Enhancement*, Computer Graphics Forum, 2012.
- Baoyuan Wang, Yizhou Yu, Tien-Tsin Wong, Chun Chen and Ying-Qing Xu, *Data-Driven Image Color Theme Enhancement*, ACM SIGGRAPH, 2010.
- Naila Murray, Sandra Skaff and Luca Marchesotti, *Towards Automatic Concept Transfer*, SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering, 2011.
- Frank Wilcoxon, *Individual Comparisons by Ranking Methods*, Biometrics Bulletin, 1945.
- Shaojie Zhuo and Terence Sim, *Defocus map estimation from a single image*, Pattern Recognition, 2011.
- Sung Ju Hwang, Ashish Kapoor and Sing Bing Kang, *Context-Based Automatic Local Image Enhancement*, ECCV, 2012.

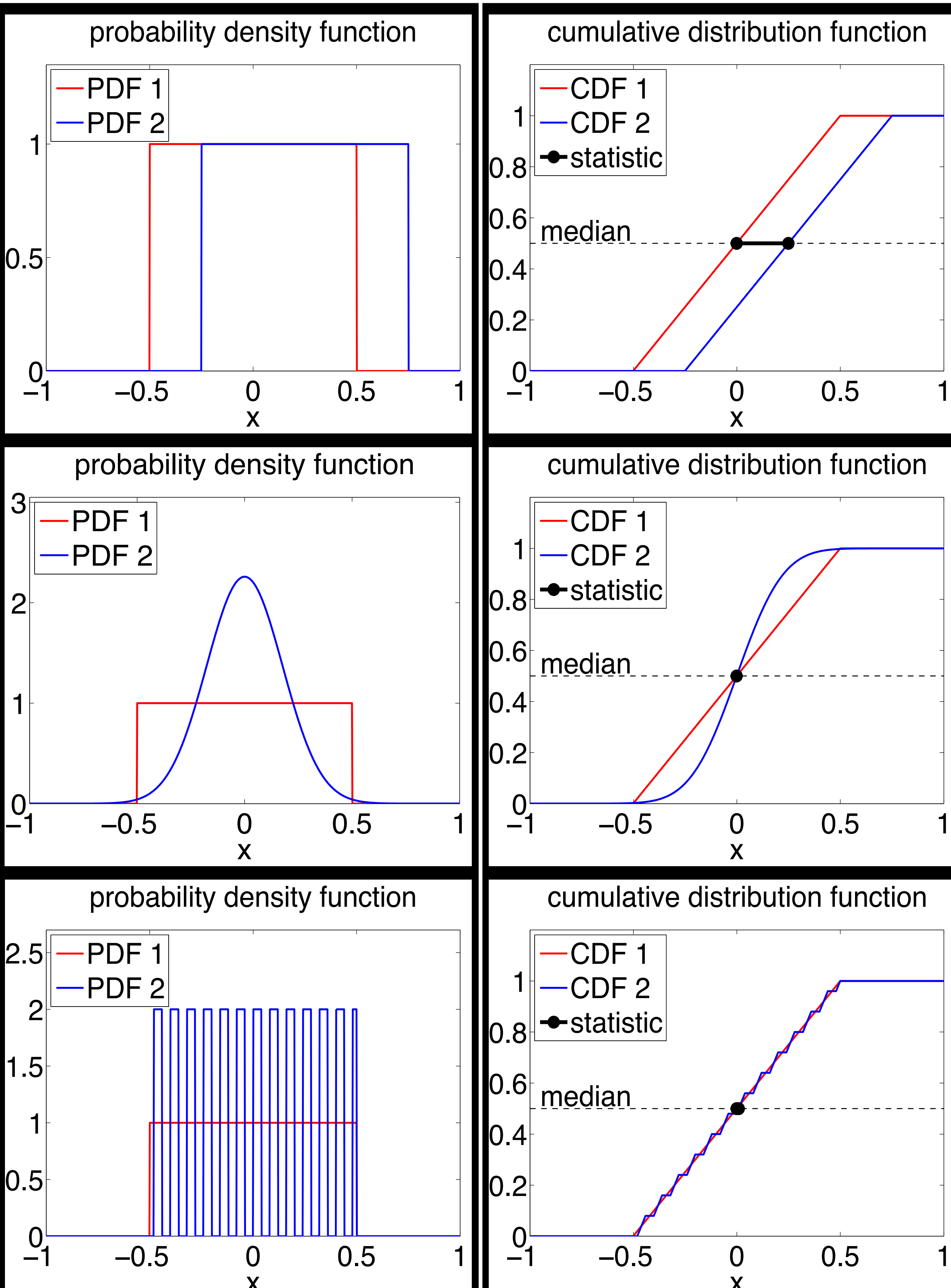
Input	Wilcoxon	Kolmogorow Smirnow	Chi-square



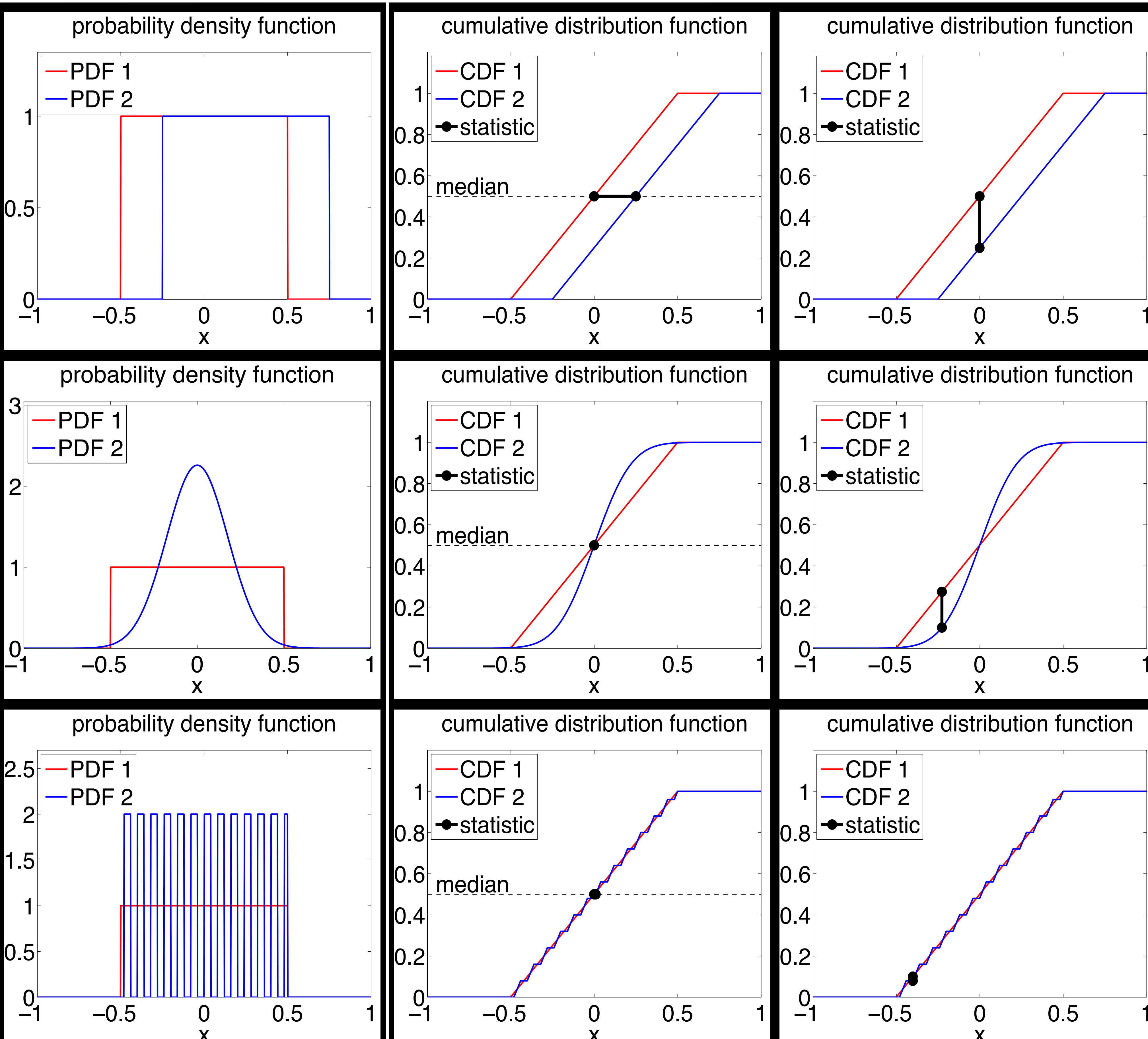




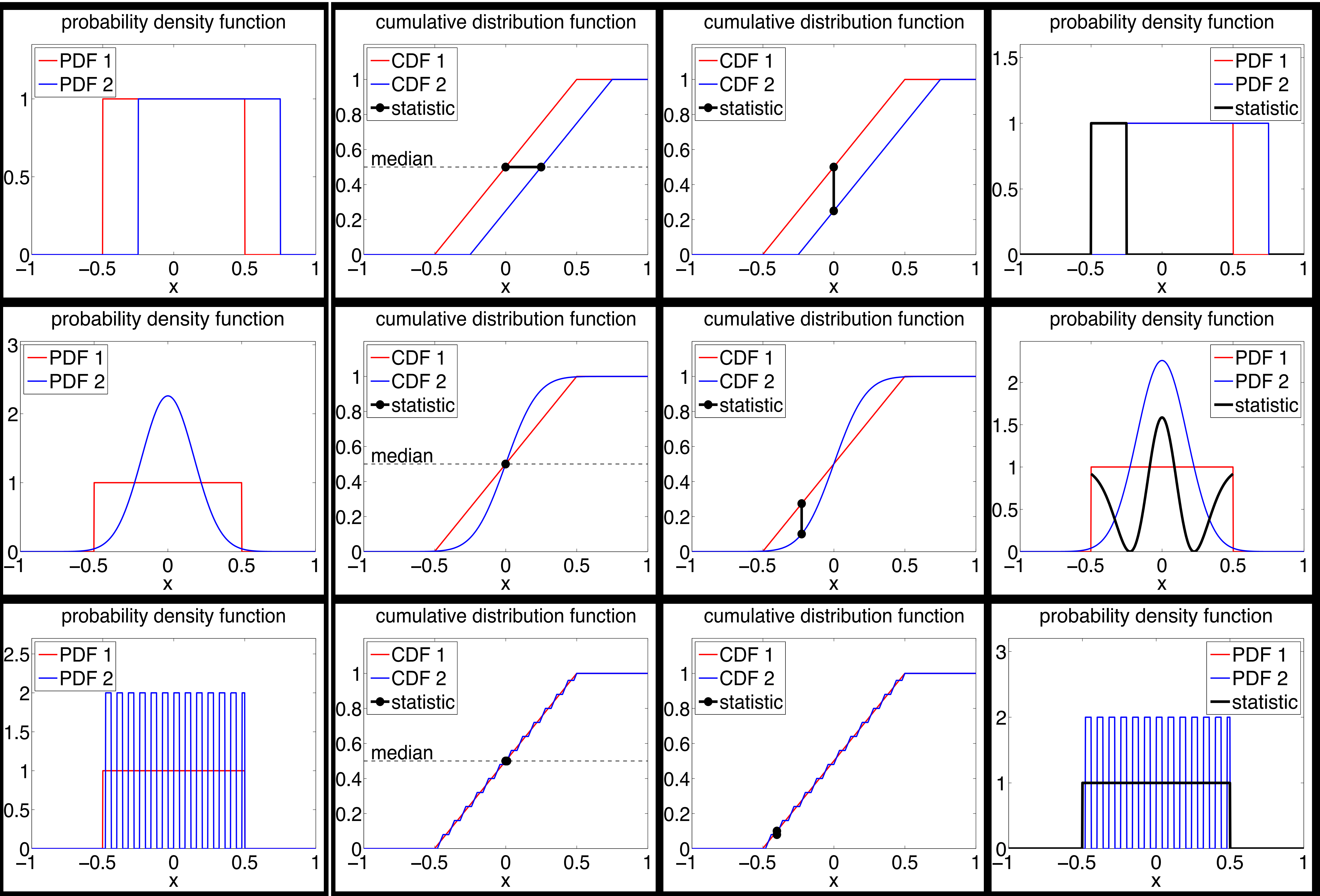
Input	Wilcoxon	Kolmogorow Smirnow	Chi-square



Input	Wilcoxon	Kolmogorow Smirnow	Chi-square



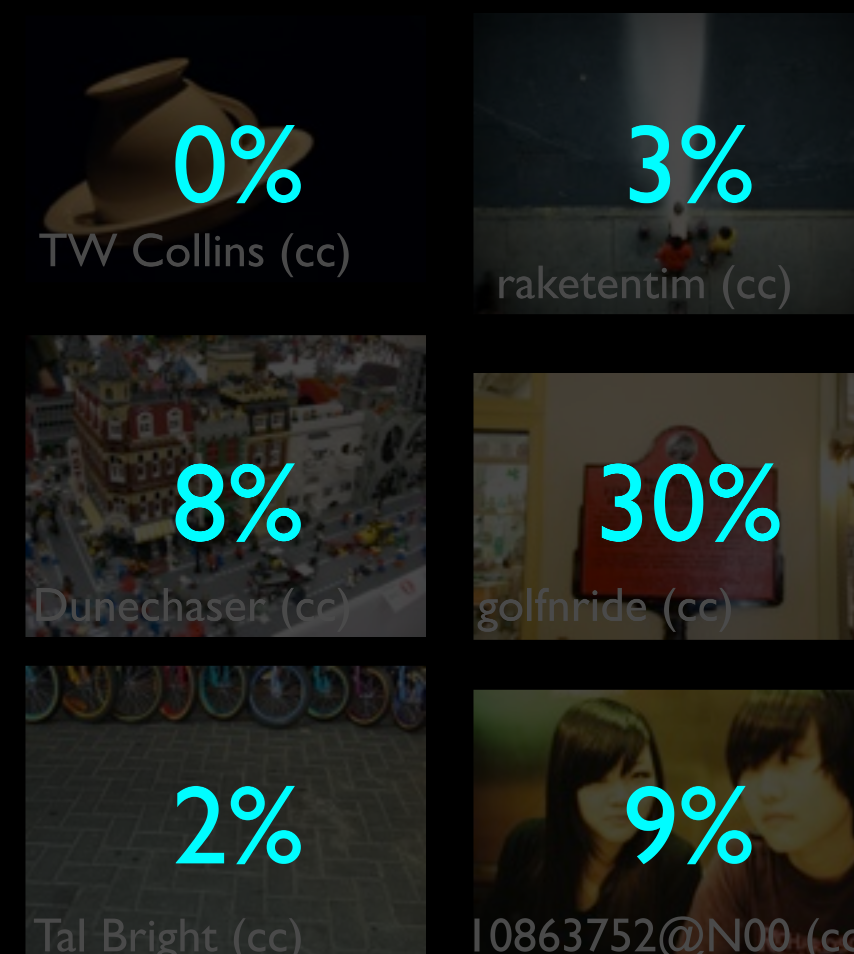
Input	Wilcoxon	Kolmogorow Smirnow	Chi-square



Statistical Framework

gold

$\overline{gold}$



percentage of yellow pixels

Statistical Framework

sorted list: 0% 2% 3% 5% 8% 9% 10% 30% 70% 90%

Statistical Framework

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

ranksum: $T = 4 + 7 + 9 + 10 = 30$

Statistical Framework

sorted list: 0% 2% 3% 5% 8% 9% 10% 30% 70% 90%

rank index: 1 2 3 4 5 6 7 8 9 10

ranksum: $T = 4 + 7 + 9 + 10 = 30$

Mann-Whitney-Wilcoxon ranksum test

$$\mu_T = \frac{n_w(n_w + n_{\overline{w}} + 1)}{2} \qquad \sigma_T^2 = \frac{n_w n_{\overline{w}}(n_w + n_{\overline{w}} + 1)}{12}$$

$n_w, n_{\overline{w}}$ cardinalities of both sets

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{30 - 22}{4.69} \approx 1.71$$

F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin, 1(6):80–83, 1945.

Statistical Framework

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

ranksum: $T = 4 + 7 + 9 + 10 = 30$

Mann-Whitney-Wilcoxon ranksum test

$$\mu_T = \frac{n_w(n_w + n_{\overline{w}} + 1)}{2} \qquad \sigma_T^2 = \frac{n_w n_{\overline{w}}(n_w + n_{\overline{w}} + 1)}{12}$$

$n_w, n_{\overline{w}}$ cardinalities of both sets

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{30 - 22}{4.69} \approx 1.71$$

$z > 0$ ➔ significantly more yellow pixels in *gold* images.

Computational Efficiency

gold

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

Computational Efficiency

gold

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

street

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

Computational Efficiency

gold

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

street

sorted list:	0%	2%	3%	5%	8%	9%	10%	30%	70%	90%
rank index:	1	2	3	4	5	6	7	8	9	10

- List is sorted only **once** for a **given characteristic.**
- Method easily scales to millions of images and thousands of keywords.

$$\sum z \neq 0$$

$m = 1, n = 1$ $\quad \mu_T = 1.5, \sigma_T = 0.25$ $\quad z = \frac{T - \mu_t}{\sigma_T}$

	black	gray	white
I_1	0.5	0.4	0.1
I_2	0.33	0.33	0.33
T	2	2	1
z	2	2	-2

	black	gray	white
I_1	0.5	0.3	0.2
I_2	0.33	0.33	0.33
T	2	1	1
z	2	-2	-2

light

original

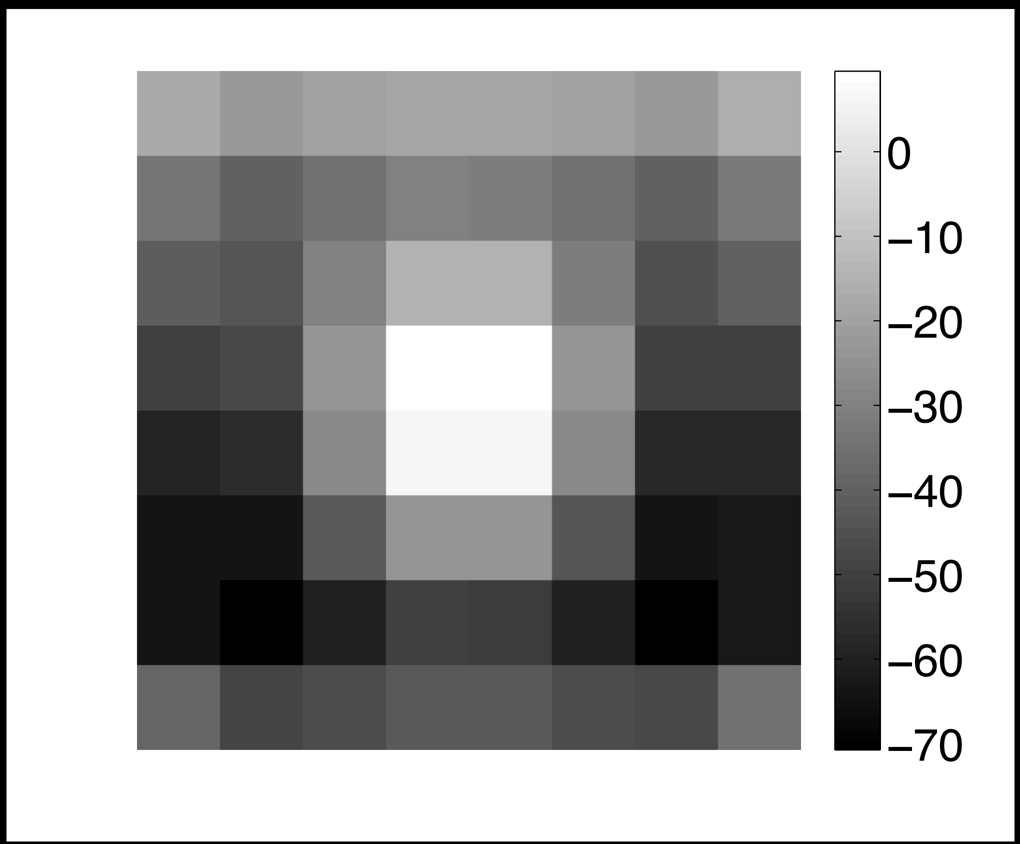
proposed

Approval rate

AMT: 19%
Artists: 77%

Other Characteristics

macro



- Spatial layout of high frequency content for keyword *macro*.
- Significantly less details along the image borders.

Significance distributions can be computed for any characteristic.

same input

different renderings

sand

dark

Keyword ➔ Image

- Not a classification task.

Keyword → Image

- Not a classification task.
- Instead: keyword's significance for an image characteristic:
  - Lightness
  - Color
  - Depth-of-field

Efficiency

- Wilcoxon ranksum test requires the values of both sets to be sorted.

Efficiency

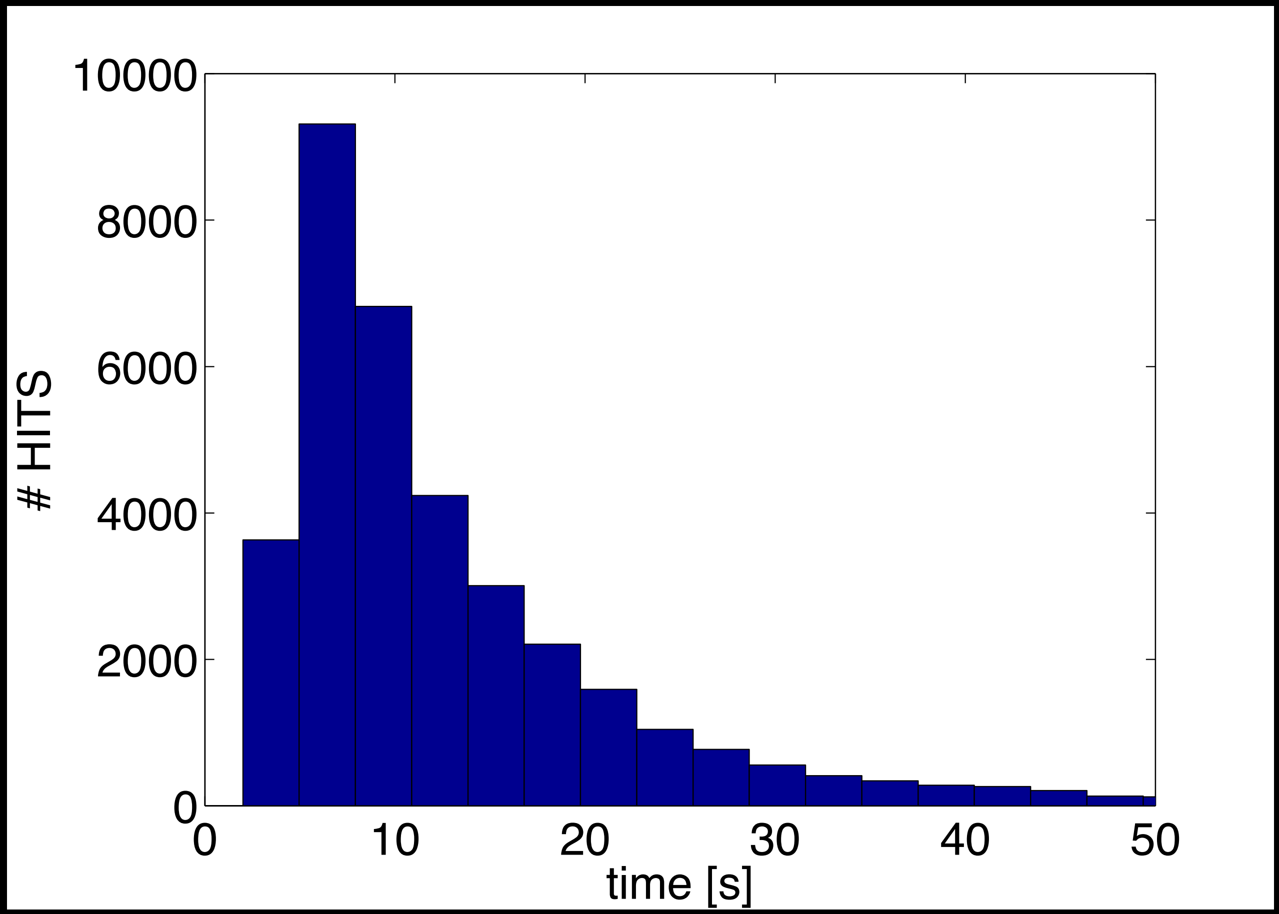
- Wilcoxon ranksum test requires the values of both sets to be sorted.
- But only once; the significance of an additional keyword is computed with a simple sum.

Efficiency

- Wilcoxon ranksum test requires the values of both sets to be sorted.
- But only once; the significance of an additional keyword is computed with a simple sum.
- The statistical framework easily scales to millions of images and thousands of keywords.

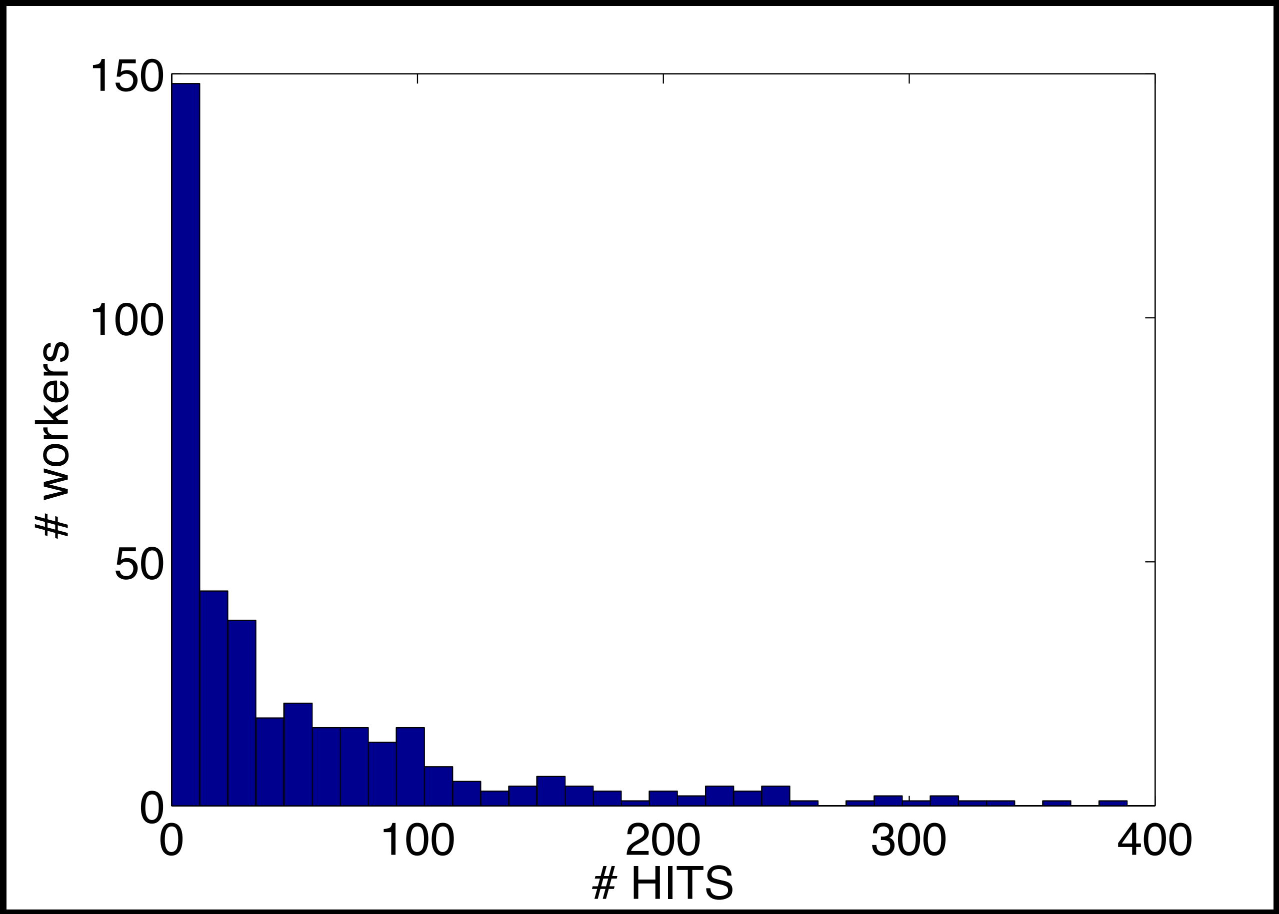
AMT Statistics

Time per HIT



$Q_{0.05} = 4s$
$Q_{0.5} = 10s$
$Q_{0.95} = 40s$

Number of HITs



$Q_{0.05} = 1$
$Q_{0.5} = 27$
$Q_{0.95} = 380$