What Makes the Visual Information Popular and Memorable?

Antonino Furnari
furnari@dmi.unict.it
www.dmi.unict.it/~furnari
Outline

• Part I:
  – Digital Images;
  – Image Features;
  – Exif;
    – Amazon Mechanical Turk;
• Part II: what makes an image memorable?
• Part III: what makes an image popular?
• Conclusion and open challenges.
Part I
Digital Images (Review)
Image Functions

An image is a 2D discrete function $f(x,y)$ measuring some property (e.g., intensity, colour) of a given point in the real scene.
Image Acquisition
Image Acquisition
Grayscale Images

A grayscale digital image is a table of discrete values \( I = f(x,y) \). Each value measures the intensity of the light reflected by a particular point in the real scene and belongs to the interval \([0, L-1]\), where \( L \) is the number of gray levels.
Color Images

\[ z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \]

- Component image 1 (Red)
- Component image 2 (Green)
- Component image 3 (Blue)
Color Images

Red + Green + Blue = RGB color
Color Spaces: RGB

Each color is identified by three values \((R, G, B)\) representing the amount of “red”, “green” and “blue” respectively.
Color Spaces: HSV

Each color is represented by three values (H,S,V) representing three properties of the color: its “hue”, its “saturation” and its “value” (or lightness).

Image Features

“a piece of information which is **relevant** for solving the **computational task** related to a certain application”

“features may be **specific structures** in the image such as points, edges or objects”

“the feature concept is **very general** and the choice of features in a particular system may be highly dependent on the specific problem at hand”

Feature Extraction

The features of an image are extracted through a process called **feature extraction** which can return:

- a **Boolean value** indicating whether the feature is present or not;
- some **numerical value** (scalar or vector) describing the extracted feature.
Global vs Local Features

Local Features encode a local property of the image: the presence of an edge in a given image region, the orientation the gradient vector at a particular point, the description of a given key-point, etc.

Global Features encode a global property of the image: the overall distribution of colors or edge orientations in the image, the presence of a given object in the image, the scene category of the image (e.g., indoor, outdoor), etc.

A global feature can be built out of many local features.
Low Level vs High Level Features

Abstraction
Semantics

High Level Features

Low Level Features
Edge/Gradient

\[ \nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \]
Color/Intensity Histogram

**Color histograms** are usually built considering the RGB or HSV color spaces. If a *grayscale* image is considered, we get an “intensity histogram”.

global/local feature
Keypoints/Corners

A **keypoint** is a distinctive point in the image which exhibits given characteristics. A good keypoint has some properties of invariance (e.g., to rotation, scale, change of viewpoint).

Examples of keypoint are the **corners**, which are defined as the intersection of two edges.
Scale Invariant Feature Transform (SIFT)

SIFT are a particular type of keypoints designed to be invariant to scale and rotation.
Keypoint detection/description

The **keypoint extraction process** usually involves two steps:

- The keypoint **detection**, where the keypoint are localized in the image;
- The keypoint **description**, where some feature (e.g., histograms of the gradient orientations) is extracted from the neighbourhood of the image point.
SIFT Descriptor

Image gradients → Keypoint descriptor → Numerical descriptor

local feature
Histograms of Oriented Gradient (HOG)

The HOG are based on local histograms of gradient orientations which are able to capture local structures still considering the global picture.
Texture / LBP

Fig. 2.1 Example of an input image, the corresponding LBP image and histogram

local feature
GIST

global feature
Local Self Similarities

Input image → Correlation surface → Image descriptor

Local/global feature
Bag of Visual Words

- An image can be treated as a document, and features extracted from the image are considered as the "visual words"...

**Image of an “object” category**

**Bag of visual words**

**Bag of (visual) Words:** an image is represented as an unordered collection of visual words

- D1: “face”
- D2: “bike”
- D3: "violin"
Convolutional Neural Networks
Low Level vs High Level Features

Abstraction
Semantics

High Level Features

Convolutional Networks
Bag of Visual Words
SIFT
GIST
HOG

Gradient Orientation Histograms
Intensity/Color Histograms
LBP

Low Level Features

Intensity/Color
Edge/Gradients
Exchangeable Image File Format (ExIF) is a standard that specifies the formats for images, sound, and ancillary tags used by digital cameras (including smartphones), scanners and other systems handling image and sound files recorded by digital cameras.

The specification uses the following existing file formats with the addition of specific metadata tags: JPEG DCT for compressed image files, TIFF for uncompressed image files, and RIFF WAV for audio files. It is not supported in JPEG 2000, PNG, or GIF.

This standard consists of the Exif image file specification and the Exif audio file specification.
ExIF Data

ExIF allows to integrate further information into the file. The information usually contained into a standard ExIF include:

- **Dimensions of the image**;
- **Date and Time of Acquisition**;
- **Features about acquisition**:
  - Exposure-time, Exposure Bias, F-Number, Aperture, ISO, Focal length, GPS coordinates etc;
- **Thumbnail preview** (a small picture which would be equal to the original picture).
ExIF and Social Media

More photos of La Reunion in my album: [link](http://www.flickr.com/photos/zeeyo/pictures/sets/72157632663452...)

© 2014 Yoann JEZEQUEL

---

Canon EOS 60D
EF-S15-85mm
f/3.5-5.6 IS USM

© Yoann JEZEQUEL

---

Flash (spento, non attivato)
Image Description - Storm on the sea
Make - Canon
X-Resolution - 240 dpi
Y-Resolution - 240 dpi
Software - Adobe Photoshop
Lightroom 5.3 (Windows)
Date and Time (Modified) - 2014:07:20 16:22:10
Artist - Yoann JEZEQUEL
Copyright - © Yoann JEZEQUEL
ISO Speed - 200
The Mechanical Turk

The Turk, also known as the Mechanical Turk or Automaton Chess Player, was a **fake chess-playing machine** constructed in the late 18th century.

From 1770 until its destruction by fire in 1854 it was exhibited by various owners as an **automaton**, though it was exposed in the early 1820s as an **elaborate hoax**.

wikipedia
Amazon Mechanical Turk is a crowdsourcing internet marketplace that enables individuals and businesses (known as Requesters) to coordinate the use of human intelligence to perform tasks that computers are currently unable to do.

The Requesters are able to post tasks known as HITs (Human Intelligence Tasks), such as choosing the best among several photographs of a storefront, writing product descriptions, or identifying performers on music CDs.

Workers can then browse among existing tasks and complete them for a monetary payment set by the Requester.
Amazon Mechanical Turk

All HITs
1-10 of 2614 Results

Select answers to questions
Requester: Aaron Adams
HIT Expired Date: Jan 12, 2015 (6 days 13 hours) Reward: $0.22
Time Allotted: 16 minutes
HITs Available: 10881

Description: Select responses that answer each question
Keywords: technology, internet, computing, os, programming
Qualifications Required:
- Total approved HITs is not less than 5000
- HIT approval rate (%) is not less than 90

Extract purchased items from a shopping receipt
Requester: Asaf Svil
HIT Expired Date: Jan 23, 2015 (6 days 23 hours) Reward: $0.39
Time Allotted: 2 hours
HITs Available: 18167

Description: Transcribe all the purchased items and total from a shopping receipt
Keywords: image, receipt, purchase, transcribe, extract, data, entry, transcription, text, easy, qualification, secure, job, basic, proof
Qualifications Required: None
Spearman's Rank Correlation

In statistics, **Spearman's** rank correlation coefficient, is a nonparametric measure of statistical dependence between two variables.

It assesses how well the relationship between two variables can be described using a **monotonic function**. If there are **no repeated data values**, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a **perfect monotone function of the other**.

[Graphs showing different correlation coefficients]
Part II

What makes an image memorable?
Image Memorability

- Is memorability an intrinsic quality of the images?
- How shall we define it?
- How shall we measure it?
- Which image content influences memorability?
- Can memorability be predicted basing on its visual content?

(Isola et al. 2011)
Which of these images are the most memorable?
Measuring the Image Memorability

The memorability is measured with a memory game on Amazon Mechanical Turks. The participants viewed a sequence of images, each of which was displayed for 1 seconds with a 1.4 seconds gap in between image presentations. The task was to press the spacebar whenever they saw an identical repeat of an image at any time in the sequence.

The sequence of images was composed of “targets” and “fillers”. The fillers provided spacing between the first and second repetition of a target. Moreover, responses on repeated fillers constituted a “vigilance task” that allowed to check that participants were attentive to the task.
Figure 2. Sample of the database used for the memory study. The images are sorted from more memorable (left) to less memorable (right).
Acquired Dataset

a) Most memorable images (86%)

b) Typical images (74%)

c) Least memorable images (34%)
Memorability Consistency

To evaluate consistency, the participants are split into two independent halves, and it is quantified how well image scores measured on the first half of the participants matched image scores measured on the second half of the participants.
Image Memorability

✔ Is memorability an intrinsic quality of the images?
✔ How shall we define it?
✔ How shall we measure it?
➜ Which image content influences memorability?
• Can memorability be predicted basing on its visual content?
Unusualness, Aesthetics, Memorability

How does unusualness and aesthetics influence memorability? Is memorability an “intuitive” concept?
Annotations

To study the influence of the semantic content of the images on their memorability, the dataset has been annotated by human users. The annotations include:

- Object categories;
- Scene categories;
- Attributes.
# Attributes

## Table 1: General attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial layout</strong></td>
<td>Enclosed space vs. Open space; Perspective view vs. Flat view; Empty space vs. Cluttered space; Mirror symmetry vs. No mirror symmetry (cf. [23])</td>
</tr>
<tr>
<td><strong>Aesthetics</strong></td>
<td>Post-card like? Buy this painting? Hang-on wall? Is aesthetic? Pleasant vs. Unpleasant; Unusual or strange vs. Routine or mundane; Boring vs. Striking colors; High quality (expert photography) vs. Poor quality photo; Attractive vs. Dull photo; Memorable vs. Not memorable; Sky present? Clear vs. Cloudy sky; Blue vs. Sunset sky; Picture of mainly one object vs. Whole scene; Single focus vs. Many foci; Zoomed-in vs. Zoomed-out; Top down view vs. Side view (cf. [7])</td>
</tr>
<tr>
<td><strong>Dynamics</strong></td>
<td>Action going on? Something moving in scene? Picture tells a story? About to happen? Lot going on? Dynamic scene? Static scene? Have a lot to say; Length of description</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>Famous place? Recognize place? Like to be present in scene? Many people go here?</td>
</tr>
<tr>
<td><strong>Contains a person?</strong></td>
<td></td>
</tr>
</tbody>
</table>

SMM – GMF – A.A. 2014/2015
# People Attributes

Table 2: Attributes describing people in image

<table>
<thead>
<tr>
<th>Visibility (per-person):</th>
<th>Face visible? Making eye-contact?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics (per-person):</td>
<td>Gender (male, female)? Age (child, teenager, adult, senior)? Race (Caucasian, SouthEast-Asian, East-Asian, African-American, Hispanic)?</td>
</tr>
<tr>
<td>Appearance (per-person):</td>
<td>Hair length (short, medium, long, bald)? Hair color (blonde, black, brown, red, grey)? Facial hair?</td>
</tr>
</tbody>
</table>
Example Attributes Annotations

(a) ↑attractive  (b) ↑funny  (c) ↑makes-sad  (d) ↑qual. photo  (e) ↑peaceful

(f) ↓attractive  (g) ↓funny  (h) ↓makes-sad  (i) ↓qual. photo  (j) ↓peaceful
Measuring the Importance of the Features

The “importance” of the considered features (annotations) is measured selecting a subset of them and learning a regressor able to predict the image memorability. The “performances” of the considered features are measured as the Spearman's rank correlation between predicted and ground truth memorabilities.

The features are selected choosing a “bit budget” in two ways:

- **Information-Theoretic**: a method to select the subset of features which maximizes the mutual information with memorability;

- **Predictive**: the features which allow the biggest boost in regression performances are chosen as “most predictive”.
Measuring the Importance of the Features (2)

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Perf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object annotations</td>
<td>0.494</td>
</tr>
<tr>
<td>Scene annotations</td>
<td>0.415</td>
</tr>
<tr>
<td>Attribute annotations</td>
<td>0.528</td>
</tr>
<tr>
<td>Objects + Scenes + Attributes</td>
<td>0.554</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information-theoretic</th>
<th>Predictive</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑ enclosed space</td>
<td>0.39</td>
</tr>
<tr>
<td>↑ face visible</td>
<td>0.37</td>
</tr>
<tr>
<td>↓ peaceful</td>
<td>-0.33</td>
</tr>
<tr>
<td>↓ sky present</td>
<td>-0.35</td>
</tr>
<tr>
<td>↑ enclosed space</td>
<td>0.39</td>
</tr>
<tr>
<td>↑ face visible</td>
<td>0.37</td>
</tr>
<tr>
<td>↑ tells a story</td>
<td>0.18</td>
</tr>
<tr>
<td>↑ recognize place</td>
<td>0.16</td>
</tr>
<tr>
<td>↓ peaceful</td>
<td>-0.33</td>
</tr>
</tbody>
</table>
Image Memorability

- Is memorability an intrinsic quality of the images?
- How shall we define it?
- How shall we measure it?
- Which image content influences memorability?
- Can memorability be predicted basing on its visual content?
Correlation with Simple Features

Some simple features are considered:

- **Color:**
  - Mean hue;
  - Mean saturation;
  - Mean value;

- **Intensity:**
  - Mean intensity;
  - Intensity variance;
  - Intensity skewness;

- **Objects statistics:**
  - Number of objects;
  - Mean class coverage;
  - Max class coverage;
Advanced Features

To analyse the importance of more sophisticated features, a **Support Vector Regressor** is trained using the considered features.

The performances are assessed using the **Spearman's rank correlation** and **relative sorting** as done to assess consistency among the participants.
Object Statistics

- **Object counts**: number of objects in the image;
- **Object areas**: pixel occupied by objects in the image;
- **Multiscale object areas**: concatenation of object areas per image and quadrant.
## Object Statistics (results)

<table>
<thead>
<tr>
<th></th>
<th>Object Counts</th>
<th>Object Areas</th>
<th>Multiscale Object Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 20</td>
<td>68%</td>
<td>67%</td>
<td>73%</td>
</tr>
<tr>
<td>Top 100</td>
<td>68%</td>
<td>68%</td>
<td>73%</td>
</tr>
<tr>
<td>Bottom 100</td>
<td>67%</td>
<td>64%</td>
<td>64%</td>
</tr>
<tr>
<td>Bottom 20</td>
<td>67%</td>
<td>63%</td>
<td>65%</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Object and Scene Semantics

- **Object label presence**: whether at least an object of a given class is present or not;
- **Labeled object counts**: number of objects per class;
- **Labeled object areas**: number of pixel occupied by objects of a specific class;
- **Labeled multiscale object areas**: object areas per image and quadrant;
- **Scene category**: scene category of the considered image;
- **Objects and scenes**: labeled multiscale object areas + scene category.
# Object and Scene Semantics (results)

<table>
<thead>
<tr>
<th></th>
<th>Object Label Presences</th>
<th>Labeled Object Counts</th>
<th>Labeled Object Areas</th>
<th>Labeled Multiscale Object Areas</th>
<th>Scene Category</th>
<th>Objects and Scenes</th>
<th>Other Humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 20</td>
<td>84%</td>
<td>82%</td>
<td>84%</td>
<td>84%</td>
<td>81%</td>
<td>85%</td>
<td>86%</td>
</tr>
<tr>
<td>Top 100</td>
<td>79%</td>
<td>79%</td>
<td>82%</td>
<td>82%</td>
<td>78%</td>
<td>82%</td>
<td>84%</td>
</tr>
<tr>
<td>Bottom 100</td>
<td>57%</td>
<td>57%</td>
<td>56%</td>
<td>56%</td>
<td>57%</td>
<td>55%</td>
<td>47%</td>
</tr>
<tr>
<td>Bottom 20</td>
<td>55%</td>
<td>54%</td>
<td>53%</td>
<td>52%</td>
<td>55%</td>
<td>53%</td>
<td>40%</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.43</td>
<td>0.44</td>
<td>0.47</td>
<td>0.48</td>
<td>0.37</td>
<td>0.50</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Prediction using Global Features

In order to build a predictor, some global features are chosen. Some of them have been found effective at scene recognition (GIST) as well as to predict the presence/absence of objects in images (SIFT, HOG, SSIM).

<table>
<thead>
<tr>
<th></th>
<th>Pixels</th>
<th>GIST</th>
<th>SIFT</th>
<th>SSIM</th>
<th>HOG 2x2</th>
<th>All Global Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 20</td>
<td>74%</td>
<td>82%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
</tr>
<tr>
<td>Top 100</td>
<td>72%</td>
<td>78%</td>
<td>79%</td>
<td>79%</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Bottom 100</td>
<td>61%</td>
<td>58%</td>
<td>57%</td>
<td>58%</td>
<td>57%</td>
<td>56%</td>
</tr>
<tr>
<td>Bottom 20</td>
<td>59%</td>
<td>57%</td>
<td>56%</td>
<td>56%</td>
<td>55%</td>
<td>54%</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.22</td>
<td>0.38</td>
<td>0.41</td>
<td>0.43</td>
<td>0.43</td>
<td>0.46</td>
</tr>
</tbody>
</table>
The Role of the Objects

[...]

More precisely, to quantify the contribution of an object \( i \) to an image, we take a prediction function, \( f \), that maps object features to memorability scores and calculate how its prediction \( m \) changes when we zero features associated with object \( i \) from the current image’s feature vector, \((a_1, \cdots, a_n)\). This gives us a score \( s_i \) for each object in a given image:

\[
\begin{align*}
m_1 &= f(a_1, \cdots, a_i, \cdots, a_n) \\
m_2 &= f(a_1, \cdots, 0, \cdots, a_n) \\
s_i &= m_1 - m_2
\end{align*}
\]
Figure 6. Objects sorted by their predicted impact on memorability. Next to each object name we report how much an image’s predicted memorability will change, on average, when the object is included in the image’s feature vector versus when it is not. For each object name, we also display two test set images that contain the object: on the left is the example image with the highest memorability score among all test set images that contain (over 4000 pixels of) the object. On the right is the example with the lowest score. Only objects that appear (cover over 4000 pixels) in at least 20 images in our training set are considered.
Predicting Memorability: Results
Part III

What makes an image popular?
Image Popularity

- Is popularity an intrinsic quality of the images?
- How shall we define it?
- How shall we measure it?
- Can it be predicted basing on its visual content?
- And using social cues?

(Khosla et al. 2014)
Aim

Hundreds of thousands of photographs are uploaded to the internet every minute through various social networking and photo sharing platforms. While some images get millions of views, others are completely ignored. Even from the same users, different photographs receive different number of views.

This begs the question: What makes a photograph popular? Can we predict the number of views a photograph will receive even before it is uploaded?

Two key components are investigated: the image content and social context.

(Khosla et al. 2014)
What is Image Popularity?

We focus on the number of views on Flickr as a medium to define the popularity of an image.

Visual media tends to receive views over some period of time. To normalize for this effect, the number of view is divided by the duration since the upload date of the given image.
Dataset

Using the **Flickr APIs**, a dataset of **2.3M** images is gathered. The images belong to **hundred of thousands** of different **users**.

Three different settings are considered:

- **One-per-user**: 930k images from 400k. About 2 images per user;

- **User-mix**: **100 random user** from the one-per-user dataset that had between 10k and 20k public photos shared on Flickr, resulting in a dataset of approximately **1.4M image**;

- **User-specific**: the user-mix dataset is split into 100 subsets, each containing images of a **specific user**.
Dataset

Figure 1: Sample images from our image popularity dataset. The popularity of the images is sorted from more popular (left) to less popular (right).
Image Popularity

• Is popularity an intrinsic quality of the images?
  ✔ How shall we define it?
  ✔ How shall we measure it?
• Can it be predicted basing on its visual content?
• And using social cues?

(Khosla et al. 2014)
Predicting Popularity

To predict image popularity, we first select some features and then build an SVR regressor.

The importance of considered the features is assessed measuring the performances of the regressor in terms of Spearman's rank correlation between predicted and ground truth popularity scores.
Color and Simple Image Features

Here all the images are used independently from the dataset settings.
The color space is grouped into 50 colors and each pixel is assigned with one of these colors. A SVR is trained on the 50-colors histograms in order to learn the importance of such colors.
Motivated by the studies on image popularity, some low level features are considered:

- **GIST** (to encode the scene category);
- **Color Histogram**;
- **Texture** (LBP);
- **Color Patches** (50 colors in a bag of words representation);
- **Gradient** (HOG);
- **Deep Learning** (based on CNN, to encode the presence of a given object).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Gist</th>
<th>Color histogram</th>
<th>Texture</th>
<th>Color patches</th>
<th>Gradient</th>
<th>Deep learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-per-user</td>
<td>0.07</td>
<td>0.12</td>
<td>0.20</td>
<td>0.23</td>
<td>0.26</td>
<td>0.28</td>
</tr>
<tr>
<td>User-mix</td>
<td>0.13</td>
<td>0.15</td>
<td>0.22</td>
<td>0.29</td>
<td>0.32</td>
<td>0.33</td>
</tr>
<tr>
<td>User-specific</td>
<td>0.16</td>
<td>0.23</td>
<td>0.32</td>
<td>0.36</td>
<td>0.34</td>
<td>0.26</td>
</tr>
</tbody>
</table>
High Level Features: Objects

An **object classifier** (based on Deep Learning) able to recognize 1000 object classes, is used to detect whether a given **object class** is present or not in the image.

The presence or not of a given object is used as a feature to learn the **SVR**.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Gist</th>
<th>Color histogram</th>
<th>Texture</th>
<th>Color patches</th>
<th>Gradient</th>
<th>Deep learning</th>
<th>Objects</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-per-user</td>
<td>0.07</td>
<td>0.12</td>
<td>0.20</td>
<td>0.23</td>
<td>0.26</td>
<td>0.28</td>
<td>0.23</td>
<td>0.31</td>
</tr>
<tr>
<td>User-mix</td>
<td>0.13</td>
<td>0.15</td>
<td>0.22</td>
<td>0.29</td>
<td>0.32</td>
<td>0.33</td>
<td>0.30</td>
<td>0.36</td>
</tr>
<tr>
<td>User-specific</td>
<td>0.16</td>
<td>0.23</td>
<td>0.32</td>
<td>0.36</td>
<td>0.34</td>
<td>0.26</td>
<td>0.33</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Objects Impact

Moreover, the learned SVR weights are indicative of the impact of a given object at predicting popularity. This leads to the definition of four classes:

• Strong positive impact objects;
• Medium positive impact objects;
• Low positive impact objects;
• Negative impact objects.
Negative Impact

spatula

plunger

laptop

golf cart

space heater
Low Positive Impact

wild boar  solar dish  horse cart

guacamole  catamaran
Medium Positive Impact

- cheetah
- giant panda
- basketball
- llama
- plow
- ladybug
Strong Positive Impact

miniskirt  maillot  bikini  cup

brassiere  perfume  revolver
Image Popularity

✔ Is popularity an intrinsic quality of the images?
✔ How shall we define it?
✔ How shall we measure it?
✔ Can it be predicted basing on its visual content?
➔ And using social cues?

(Khosla et al. 2014)
Social Cues

The image content has a significant correlation with popularity. However, the social context plays a significant role in the number of views an image will receive.

Thus, the following social features are considered per each user:

- **Mean views**: mean number of normalized views of all public images;
- **Photo count**: number of public photos uploaded by the user;
- **Contacts**: number of contacts of the given user;
- **Groups**: number of groups the user belong to;
- **Group members**: average number of members in the groups a given user belongs to;
- **Member duration**: the amount of time since the user joined Flickr;
- **Is pro**: whether the given user has a Pro Flickr account or not.

Moreover the following social features related to each image are considered:

- **Tags**: number of tags;
- **Title length**: length of the image title;
- **Description length**: length of the image description.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Content only</th>
<th>Social only</th>
<th>Content + Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-per-user</td>
<td>0.31</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td>User-mix</td>
<td>0.36</td>
<td>0.66</td>
<td>0.72</td>
</tr>
<tr>
<td>User-specific</td>
<td>0.40</td>
<td>0.21</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Results

predicted popularity

ground truth popularity
Conclusion

We have seen that, using Computer Vision techniques, it is possible to analyze the visual content of images uploaded on the social networks.

This information can be leveraged to make predictions about image memorability and popularity.

But the visual content of images could tell us much more and many challenges are still open.
Possible Theses

- Estimation of the popularity (likes/views) of food images;
- This is food;
- Food recommendation;
- Restaurant food logging;
- Menu ranking;
- Visual sentiment analysis;
- The social picture (understanding which part of a monument is most photographed);
- Selfie vs non-selfie;
- The wisdom of the crowd for automatic photo selection.