Content-Based Multimedia Retrieval
Content-Based Multimedia Retrieval

- What is content-based?
- General framework of multimedia Retrieval
- Content-based image/video retrieval
- Relevance Feedback
Content-based?

- There is an amazing growth in the amount of digital multimedia data in recent years.
- Lack of tools for classify and retrieve multimedia (image, audio and video ) content.
- There exists a huge semantic gap between low-level features and high-level semantic content.
- To let machine understand content is important and challenging.
Need of Image/Video Search

- Explosive growth of online image/video data, personal media, broadcast news videos, etc.
- 5 billion images on the Web, 31 million hours of TV programs each year
- Successful services like Youtube and Flickr
- Image/video search exciting opportunity
Example Search

- Text Query on Google: “Manhattan Cruise”

- Image content analysis may help refine results
**Image/Video Content**

- **Low-level visual content**
  - features
    - Colors
    - Shapes
    - Textures
    - Layout

- **Semantic contents**
  - features
    - high-level concepts such as objects and events
General System Framework of Multimedia Retrieval

Image, audio, video

Query 1

Query K

Features

Video 1

Video N

Features

Search Engine

Output List
General Video Retrieval System

- Query with relevance feedback

Query

Shot boundary detection

Feature generation

45217853

45217853

45217853

k-nn (boost, VSM)

Distances

Weighted sum of distances

ΣwD

Retrieved results

45217853

45217853

45217853

Relevance feedback

Key frames

Feature vectors

Test data
Content-based Video Retrieval

- Video indexing should be analogous to text document indexing
- To facilitate fast and accurate content access to video data, we should segment a video document into shots and scenes
- We should extract key frames or key sequences as index entries for scenes or stories.
Content-based Video Retrieval

- The process of attaching content based labels to video shots
- Essential for content-based classification and retrieval
- Using automatic analysis techniques
  - shot detection, video segmentation
  - key frame selection
  - object segmentation and recognition
  - visual/audio feature extraction
  - speech recognition, video text, VOCR
Content-based Video Retrieval

- Simple visual feature query
  - Retrieve video with key-frame: Color-R(80%), G(10%), B(10%)
- Feature combination query
  - Retrieve video with high motion upward(70%), Blue(30%)
- Query by example/instance
  - Retrieve video which is similar to example
- Localized feature query
  - Retrieve video with a running car toward right
- Object relationship query
  - Retrieve video with a girl watching the sun set
- Concept query (query by keyword)
  - Retrieve explosion, White Christmas
Content-based Video Classification

- Segment & classify videos into meaning categories
- Classify videos based on predefined topic
- Useful for browsing and searching by topic

Multimodal method
- Visual features
- Audio features
- Motion features
- Textual features

Domain-specific knowledge
Video Classification

- Machine learning methods
  - SVM
  - Boost
  - CNN
  - ...

Diagram:
- Video Input
- Visual
- Motion
- Audio
- Transcript
- Color
- Face
- Video Text
- Stare
- Sketch
- Texture
- Trajectory
- Camera motion
- Audio features
- Silence
- Speech
- Speech
- Noise
- Music
- Sy+M
- Sk+Is
- Spr Is
- Keyo categories
- Speech Recogn.
- VOCR
- Financial News
- Talk Show
- Sport News
- Commercial
Similarity Computation

Distance

- Euclidean distance
- Mahalanobis distance
- Cosine
- Histogram intersection
- Correlation
- Moment
- ...
Semantic Indexing & Querying

■ Limitation of QBE
  - Measuring similarity using only low-level features
  - Lack reflection of user’s perception
  - Difficult annotation of high level features

■ Syntactic to Semantic
  - Bridge the gap between low-level feature and semantic content
  - Semantic indexing, Query By Keyword (QBK)
Four Processes Involved by Video-content Analysis and Indexing

- Feature extraction
- Structure analysis
- Abstraction
- Indexing
Four Processes Involved by Video-content Analysis and Indexing

- Feature extraction
  - Video streams
  - Retrieval & browsing

- Features
  - Abstraction
    - Structure analysis
      - Metadata
    - Clustering & indexing

- Summary/skimmed video
Four Processes Involved by Video-content Analysis and Indexing

- Feature Extraction
  - The effectiveness of an indexing scheme depends on the effectiveness of attributes in content presentation.
  - An effective strategy in video-content analysis is to use attributes extractable from multimedia sources.
  - Much valuable information is also carried in other media components, such as text, audio, and speech that accompany the pictorial component.
Feature Extraction

- Color features
- Texture features
- Shape features
- Sketch features
- Audio features
- Camera motion features
- Object motion features
- Object-based features
- Face...
Four Processes Involved by Video-content Analysis and Indexing
Four Processes Involved by Video-content Analysis and Indexing

- **Structure analysis**
  - The process of extracting temporal structural information of video sequences or programs
  - Organizes video data according to their temporal structures and relations and thus build table of content
  - Stories -> scenes -> shots -> frames
Video Parsing/Structurization

---

Key-frame Extraction

Bottom-up

---

Scene

---

Plot

---

ACT

---

Movie

Top-down

---
Structure Analysis

Original Sequence

Extracted Keyframes
Four Processes Involved by Video-content Analysis and Indexing
Four Processes Involved by Video-content Analysis and Indexing

- Video abstraction
  - Video abstraction is the process of creating a presentation of visual information about a landscape or the structure of video, which should be much shorter than the original video.
  - Example: baseball game
  - Use an MPEG-7-compliant XML description format for the event segment
Relevance Feedback

- **Relevance Feedback**: involve the user in a loop, whereby images retrieved are used in further rounds of convergence onto correct returns

- **Relevance Feedback Approaches**
  - The usual situation: the user identifies images as good, bad, or don't care, and weighting systems are updated according to this user guidance.
  - Another approach is to move the query towards positively marked content.
  - An even more interesting idea is to move every data point in a disciplined way, by warping the space of feature points.
  - The basic advantage of putting the user into the loop by using relevance feedback is that the user need not provide a completely accurate initial query.
Quantifying Results

- Precision is the percentage of relevant documents retrieved compared to the number of all the documents retrieved.

\[
\text{Precision} = \frac{\text{Desired images returned}}{\text{All retrieved images}}
\]

- Recall is the percentage of relevant documents retrieved out of all relevant documents.

\[
\text{Recall} = \frac{\text{Desired images returned}}{\text{All desired images}}
\]

- These measures are affected by the database size and the amount of similar information in the database, and as well they do not consider fuzzy matching or search result ordering.
CBIR

“The process of retrieving images from a collection on the basis of features (such as colour, texture and shape) automatically extracted from the images themselves”
CBIR Simple Example

- For a given query…
  - Example Image
  - Rough Sketch
  - Explicit Description Criteria

... Return all ‘Similar’ Images
CBIR Query Types

COLOR

SKETCH

SHAPE

EXAMPLE

TEXTURE

More complex types exist yet above are most fundamental & most regularly used
On what basis are they similar?
- Color or content?
- Shape content?
- High level ideas (‘Masks’, ‘Gender’)?

Perception is always an issue

CBIR (Dis)Similarity?

Similarity is not so simple

Consider three images
However …

- Finding the right image is not always easy.
- There are many millions of digital images available on the Web.
- Many more images waiting to be digitised.
- Variable levels of metadata and associated content – if any exist at all …
- Diverse groups of users.
- Users who don’t always know either what they want or how to express it in words.
Why Research Image Data?

- Increasing use of digital images, particularly via the Internet, by professionals of all kinds.
- Inadequacy of current technology in handling images.
- Increasing interest in how people perceive, search for and use images.
- Exciting new applications opening up (e-commerce?).
Block Diagram of CBIR

1. **Image**
   - Features Extraction
     - Color
     - Shape
     - Color Sensation
     - Spatial Relation
   - Indexing & Filtering
   - Similarity Measure
     - Color
     - Shape
     - Color Sensation
     - Spatial Relation
   - Weight of Features
2. **Query Interface**
   - Query by Images
     - Query by Color
     - Query by Color Sensation
     - Query by Shape
     - Query by Spatial Relation
   - User Drawing
   - Query Server
3. **Server**
   - Indexing & Filtering
   - Similarity Measure
   - Weight of Features
4. **Client**
   - Query Interface
   - Query by Images
   - User Drawing

The diagram illustrates the process of Content-Based Image Retrieval (CBIR) system, where an image is the input, and the system processes it through various stages including features extraction, indexing, similarity measure, and querying to retrieve relevant images.
Query Specification

- Interfaces
  - Browsing and navigation
  - Specifying the conditions the objects of interest must satisfy, by means of queries

- Queries can be specified in two different ways
  - Using a specific query language
  - Query by example
    - Using actual data (object example)
Conditions on Multimedia Data

- **Query predicates**
  - **Attribute predicates**
    - Concern the attributes for which an exact value is supplied for each object
    - Exact-match retrieval
  - **Structural predicates**
    - Concern the structure of multimedia objects
    - Can be answered by metadata and information about the database schema
    - “Find all multimedia objects containing at least one image and a video clip”
Conditions on Multimedia Data

- Semantic predicates
  - Concern the semantic content of the required data, depending on the features that have been extracted and stored for each multimedia object
  - “Find all the red houses”
  - Exact match cannot be applied
Specify the degree of relevance of the retrieved objects

- Using some imprecise terms and predicates
  - Represent a set of possible acceptable values with respect to which the attribute or he features has to be matched
  - Normal, unacceptable, typical
- Particular proximity predicates
  - The relationship represented is based on the computation of a semantic distance between the query object and stored ones
  - Nearest object search
Assign each condition or term a given weight

- Specify the degree of precision by which a condition must be verified by an object
- “Find all the objects containing an image representing a screen (HIGH) and a keyboard (LOW)”

The corresponding query is executed by assigning some importance and preference values to each predicate and term
Issues Related to Feature Extraction

What are features of images ??

- Colour
  - Chromatic Histograms, dominant colours, moments,
- Shape
- Texture
For the Purpose of Color Based Image Retrieval

- Is the colour system device independent
- Is the colour system perceptual uniform
- Is the colour system linear
- Is the colour system intuitive
- Is the colour system robust against varying imaging conditions
  - Invariant to change in viewing direction
  - Invariant to change in object geometry
  - Invariant to change in direction of illumination
  - Invariant to change in intensity of illumination
  - Invariant to change in SPD of illumination
For the Purpose of Color Based Image Retrieval

- For the purpose of color based image retrieval, color systems composed according to the following criteria:
  - is the color system device independent?
  - perceptual uniform?
  - linear?
  - intuitive?
  - robust against varying imaging conditions:
    - invariant to a change in viewing direction
    - invariant to a change in object geometry
    - invariant to a change in intensity of the illumination
The histogram of image $I$ is defined as:

For a color $C_i$, $H_{ci}(I)$ represents the number of pixels of color $C_i$ in image $I$.

OR:

For any pixel in image $I$, $H_{ci}(I)$ represents the possibility of that pixel is in color $C_i$.

- Most commercial CBIR systems include color histogram as one of the features (e.g., QBIC of IBM).
- No space information.
Color Auto-correlogram

- Pick any pixel $p_1$ of color $C_i$ in the image $I$, at distance $k$ away from $p_1$ pick another pixel $p_2$, what is the probability that $p_2$ is also of color $C_i$?
The auto-correlogram of image $I$ for color $C_i$, distance $k$:

$\gamma^{(k)}_{C_i}(I) \equiv \Pr[|p_1 - p_2| = k, p_2 \in I_{C_i} | p_1 \in I_{C_i}]$

Integrate both color information and space information.
Implementations

- Pixel Distance Measures
  - Use D8 distance (also called chessboard distance):
    \[ D_8(p, q) = \max(|x_p - x_q|, |y_p - y_q|) \]
  - Choose distance \( k = 1, 3, 5, 7 \)
  - Computation complexity:
    - Histogram: \( O(n^2) \)
    - Correlogram: \( O(134 \times n^2) \)
Implementations

- Features Distance Measures:
  - $D(f(I_1) - f(I_2))$ is small $\Leftrightarrow I_1$ and $I_2$ are similar.
  - Example: $f(a) = 1000$, $f(a') = 1050$; $f(b) = 100$, $f(b') = 150$
  - For histogram:
    \[
    |I - I'|_h = \sum_{i \in [m]} \left| \frac{h_{C_i}(I) - h_{C_i}(I')}{1 + h_{C_i}(I) + h_{C_i}(I')} \right|
    \]
  - For correlogram:
    \[
    |I - I'|_\gamma \equiv \sum_{i \in [m], k \in [d]} \left| \frac{\gamma_{C_i}^{(k)}(I) - \gamma_{C_i}^{(k)}(I')}{1 + \gamma_{C_i}^{(k)}(I) + \gamma_{C_i}^{(k)}(I')} \right|
    \]

Which is smaller/more similar?
CBIR …

is about:

but is not about:

“blue”

“Botticelli”

“seashell”

“Venus”
SketchMate: Deep Hashing for Million-Scale Human Sketch Retrieval
Peng Xu\textsuperscript{1,2} \quad Yongye Huang\textsuperscript{1} \quad Tongtong Yuan\textsuperscript{1} \quad Kaiyue Pang\textsuperscript{2} \\

Yi-Zhe Song\textsuperscript{2} \quad Tao Xiang\textsuperscript{2} \quad Tim Hospedales\textsuperscript{3} \quad Zhanyu Ma\textsuperscript{1} \quad Jun Guo\textsuperscript{1} \\

\textsuperscript{1}Beijing University of Posts and Telecommunications \\
\textsuperscript{2}SketchX Lab, Queen Mary University of London \\
\textsuperscript{3}The University of Edinburgh
Outline

- Motivation
- Related work
- The proposed method
- Experimental results and discussions
- Conclusion
Motivation

Touch Screen

Draw Something
Motivation

- 50 million drawings
- 345 categories

Can a neural network learn to recognize doodling?

https://quickdraw.withgoogle.com/
Motivation

QuickDraw

TU-Berlin
Motivation

http://frauzufall.de/en/2017/google-quick-draw/
Motivation

Faces of Humanity

A new generated face every hour using “Quick, Draw!” game data from Google, millions of drawings from all around the world!

Each part of the face was drawn by a different person, at a different place, at a different time.

Birth: Apr 9, 2018 10:10 PM
Nationalities:

- left eye
  - Canada
- nose
  - Sweden
- right eye
  - United States of America
- mouth
  - United States of America
- circle
  - United States of America
- left ear
  - Czech Republic
- right ear
  - Thailand
Motivation

Random samples of circles from major language groups

<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>81%</td>
</tr>
<tr>
<td>Taiwan</td>
<td>56%</td>
</tr>
<tr>
<td>Egypt</td>
<td>39%</td>
</tr>
<tr>
<td>Russia</td>
<td>20%</td>
</tr>
<tr>
<td>United States</td>
<td>14%</td>
</tr>
</tbody>
</table>

South Africa  Russia
Korea         Brazil
United States  Germany


https://research.googleblog.com/2017/08/exploring-and-visualizing-open-global.html
Motivation

• Sketch Hashing Retrieval (SHR)

[1, 0, 0, 1, 1, 0, 1, ..., 0]
Outline

- Motivation
- Related Work
- The Proposed Method
- Experimental Results and Discussions
- Conclusion
Related Work

- Sketch Recognition
- Sketch-Based Image Retrieval (SBIR)
- Photo-Sketch Generation
- Sketch Generation
### Related Work

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sketch Amount</th>
<th>Category Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>TU-Berlin</td>
<td>20 000</td>
<td>250</td>
</tr>
<tr>
<td>Sketchy</td>
<td>75 000</td>
<td>125</td>
</tr>
<tr>
<td>QuickDraw-3.8M</td>
<td>3.8 million</td>
<td>345</td>
</tr>
</tbody>
</table>
Outline

- Motivation
- Related work
- The proposed method
- Experimental results and discussions
- Conclusion
The proposed method

- Two-branch CNN-RNN deep sketch hashing retrieval network
The proposed method

- CNN encoder takes in a raster pixel sketch and translates into a high-dimensional space.
The proposed method

- RNN inputs
The proposed method

- RNN encoder takes in a vector sketch and outputs its final time-step state
The proposed method

- Feature fused
The proposed method

- Loss function

\[ \mathcal{L}_{full} = \mathcal{L}_{cel} + \lambda_{scl} \mathcal{L}_{scl} + \lambda_{ql} \mathcal{L}_{ql}, \]

\[ \mathcal{L}_{cel} = \frac{1}{N} \sum_{n=1}^{N} - \log \frac{e^{W_{yn}^T f_n + \hat{b}_{yn}}}{\sum_{j=1}^{L} e^{W_{j}^T f_n + \hat{b}_{j}}}, \]

\[ \mathcal{L}_{scl} = \frac{1}{N} \sum_{n=1}^{N} \| f_n - c_{yn} \|^2_2, \]

\[ \mathcal{L}_{ql} = \frac{1}{N} \sum_{n=1}^{N} \| b_n - f_n \|^2_2, \text{ s.t. } b_n \in \{0, 1\}^D. \]
The proposed method

- Geometric interpretation of sketch feature layout obtained by different loss function

(a) cross entropy loss  
(b) cross entropy loss + common center loss  
(c) cross entropy loss + sketch center loss
The proposed method

- Statistic analysis for distances in the feature space

<table>
<thead>
<tr>
<th>len</th>
<th>dist</th>
<th>Our+CEL</th>
<th>Our+CEL+CL</th>
<th>Our+CEL+SCL</th>
<th>Our+CEL+SCL+QL (Full)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 bits</td>
<td>d₁</td>
<td>0.7501</td>
<td>0.5297</td>
<td>0.5078</td>
<td>0.5800</td>
</tr>
<tr>
<td></td>
<td>d₂</td>
<td>4.9764</td>
<td>3.2841</td>
<td>4.2581</td>
<td>4.8537</td>
</tr>
<tr>
<td></td>
<td>d₁/d₂</td>
<td>0.1665</td>
<td>0.1721</td>
<td>0.1257</td>
<td>0.1290</td>
</tr>
<tr>
<td></td>
<td>MAP</td>
<td>0.5969</td>
<td>0.5567</td>
<td>0.6016</td>
<td>0.6064</td>
</tr>
<tr>
<td>24 bits</td>
<td>d₁</td>
<td>1.2360</td>
<td>0.8285</td>
<td>0.6801</td>
<td>0.8568</td>
</tr>
<tr>
<td></td>
<td>d₂</td>
<td>6.1266</td>
<td>4.0388</td>
<td>5.0221</td>
<td>6.2243</td>
</tr>
<tr>
<td></td>
<td>d₁/d₂</td>
<td>0.2017</td>
<td>0.2051</td>
<td>0.1354</td>
<td>0.1377</td>
</tr>
<tr>
<td></td>
<td>MAP</td>
<td>0.6196</td>
<td>0.5856</td>
<td>0.6374</td>
<td>0.6388</td>
</tr>
<tr>
<td>32 bits</td>
<td>d₁</td>
<td>2.0066</td>
<td>1.5124</td>
<td>1.0792</td>
<td>1.2468</td>
</tr>
<tr>
<td></td>
<td>d₂</td>
<td>8.9190</td>
<td>7.3120</td>
<td>7.5340</td>
<td>8.6675</td>
</tr>
<tr>
<td></td>
<td>d₁/d₂</td>
<td>0.2250</td>
<td>0.2068</td>
<td>0.1432</td>
<td>0.1439</td>
</tr>
<tr>
<td></td>
<td>MAP</td>
<td>0.6412</td>
<td>0.5911</td>
<td>0.6473</td>
<td>0.6521</td>
</tr>
<tr>
<td>64 bits</td>
<td>d₁</td>
<td>4.7040</td>
<td>3.5828</td>
<td>1.6109</td>
<td>2.5231</td>
</tr>
<tr>
<td></td>
<td>d₂</td>
<td>15.4719</td>
<td>14.1112</td>
<td>11.6815</td>
<td>17.6179</td>
</tr>
<tr>
<td></td>
<td>d₁/d₂</td>
<td>0.3040</td>
<td>0.2539</td>
<td>0.1379</td>
<td>0.1432</td>
</tr>
<tr>
<td></td>
<td>MAP</td>
<td>0.6525</td>
<td>0.6136</td>
<td>0.6767</td>
<td>0.6791</td>
</tr>
</tbody>
</table>

d₁ denotes intra-class distance, d₂ denotes inter-class distance
Outline

- Motivation
- Related work
- The proposed method
- Experimental results and discussions
- Conclusion

### Experimental results and discussions

<table>
<thead>
<tr>
<th>No.</th>
<th>Model</th>
<th>Mean Average Precision</th>
<th>Precision @200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>16 bits</td>
<td>24 bits</td>
</tr>
<tr>
<td>1</td>
<td>DLBHC [13]</td>
<td>0.5453</td>
<td>0.5910</td>
</tr>
<tr>
<td>2</td>
<td>DSH-Supervised [14]</td>
<td>0.0512</td>
<td>0.0498</td>
</tr>
<tr>
<td>3</td>
<td>DSH-Sketch [15]</td>
<td>0.3855</td>
<td>0.4459</td>
</tr>
<tr>
<td>4</td>
<td>Our+CEL</td>
<td>0.5969</td>
<td>0.6196</td>
</tr>
<tr>
<td>5</td>
<td>Our+CEL+CL</td>
<td>0.5567</td>
<td>0.5856</td>
</tr>
<tr>
<td>6</td>
<td>Our+CEL+SCL</td>
<td>0.6016</td>
<td>0.6371</td>
</tr>
<tr>
<td>7</td>
<td>Our+CEL+SCL+QL (Full)</td>
<td>0.6064</td>
<td>0.6388</td>
</tr>
</tbody>
</table>

**Module 1:** Supervised Pre-Training on ImageNet

**Module 2:** Fine-tuning on Target Domain

---

ImageNet (~1.2M images)

Target domain dataset
## Experimental results and discussions

<table>
<thead>
<tr>
<th>No.</th>
<th>Model</th>
<th>Mean Average Precision</th>
<th>Precision @200</th>
<th>16 bits</th>
<th>24 bits</th>
<th>32 bits</th>
<th>64 bits</th>
<th>16 bits</th>
<th>24 bits</th>
<th>32 bits</th>
<th>64 bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DLBHC [13]</td>
<td>0.5453</td>
<td>0.5142</td>
<td>0.6109</td>
<td>0.6241</td>
<td>0.6169</td>
<td>0.6403</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DSH-Supervised [14]</td>
<td>0.0512</td>
<td>0.0510</td>
<td>0.0501</td>
<td>0.0531</td>
<td>0.0501</td>
<td>0.0454</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSH-Sketch [15]</td>
<td>0.3855</td>
<td>0.3486</td>
<td>0.4935</td>
<td>0.6065</td>
<td>0.4329</td>
<td>0.4823</td>
<td>0.6040</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Our+CEL</td>
<td>0.5969</td>
<td>0.5817</td>
<td>0.6412</td>
<td>0.6525</td>
<td>0.6292</td>
<td>0.6524</td>
<td>0.6730</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Our+CEL+CL</td>
<td>0.5567</td>
<td>0.5578</td>
<td>0.5911</td>
<td>0.6136</td>
<td>0.6038</td>
<td>0.6140</td>
<td>0.6412</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Our+CEL+SCL</td>
<td>0.6016</td>
<td>0.5928</td>
<td>0.6473</td>
<td>0.6767</td>
<td>0.6298</td>
<td>0.6543</td>
<td>0.6875</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Our+CEL+SCL+QL (Full)</td>
<td><strong>0.6064</strong></td>
<td><strong>0.5978</strong></td>
<td><strong>0.6521</strong></td>
<td><strong>0.6791</strong></td>
<td><strong>0.6324</strong></td>
<td><strong>0.6603</strong></td>
<td><strong>0.6882</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experimental results and discussions

<table>
<thead>
<tr>
<th>No.</th>
<th>Model</th>
<th>Mean Average Precision</th>
<th>Precision @200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>16 bits</td>
<td>24 bits</td>
</tr>
<tr>
<td>1</td>
<td>DLBHC [13]</td>
<td>0.5453</td>
<td>0.5910</td>
</tr>
<tr>
<td>2</td>
<td>DSH-Supervised [14]</td>
<td>0.0512</td>
<td>0.0498</td>
</tr>
<tr>
<td>3</td>
<td>DSH-Sketch [15]</td>
<td>0.3855</td>
<td>0.4459</td>
</tr>
<tr>
<td>4</td>
<td>Our+CEL</td>
<td>0.5969</td>
<td>0.6196</td>
</tr>
<tr>
<td>5</td>
<td>Our+CEL+CL</td>
<td>0.5567</td>
<td>0.5856</td>
</tr>
<tr>
<td>6</td>
<td>Our+CEL+SCL</td>
<td>0.6016</td>
<td>0.6371</td>
</tr>
<tr>
<td>7</td>
<td>Our+CEL+SCL+QL (Full)</td>
<td><strong>0.6064</strong></td>
<td><strong>0.6388</strong></td>
</tr>
</tbody>
</table>

Experimental results and discussions

- Comparison with State-of-the-Art

<table>
<thead>
<tr>
<th>Query</th>
<th>Our Full Model</th>
<th>DLBHC</th>
<th>DSH-Sketch</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Query" /></td>
<td><img src="image" alt="Our Full Model" /></td>
<td><img src="image" alt="DLBHC" /></td>
<td><img src="image" alt="DSH-Sketch" /></td>
</tr>
<tr>
<td>precision = 0.8889</td>
<td>precision = 0.7778</td>
<td>precision = 0.7500</td>
<td></td>
</tr>
</tbody>
</table>
Outline

- Related work
- Motivation
- The proposed method
- Experimental results and discussions
  ✓ Conclusion
Conclusion

- Deep sketch hashing retrieval network
- A novel hashing loss, sketch center loss
- Superior performances on 3.8M sketch dataset

**Paper Link:**

**SketchX website:**
http://sketchx.eecs.qmul.ac.uk/downloads/
Thank you!