

## ▼ 5.INTERPRETATION AND SIMILARITY

### (Biểu diễn ngữ nghĩa và sự tương đồng ngữ nghĩa)

Khi thông tin từ hình ảnh được **captured** (ghi lại) trong một tập hợp các đặc trưng (**feature set**),

có hai cách để hiểu ngữ nghĩa chúng: Một cách là **Nội suy** cố gắng giải thích từ đặc trưng (**feature set**) của ảnh, Cách còn lại gần với các phương pháp **machine learning** là so sánh các đặc trưng (**feature set**) với các phần tử (**elements**) trong một bộ dữ liệu nhất định (đã thu thập),

trên cơ sở của một hàm so sánh sự tương đồng **similarity function**, xem hình 10

### ▼ 5.1 Semantic Interpretation (Biểu diễn ngữ nghĩa)

Trong truy xuất dựa trên nội dung **content-based retrieval**, rất hữu ích để đẩy việc giải thích ngữ

nghĩa của các đặc trưng **features** xuất phát từ hình ảnh càng nhiều càng tốt.

Các đặc trưng ngữ nghĩa (**Semantic features**) nhằm mã hóa các diễn giải (ngữ nghĩa) của hình ảnh (có thể liên quan đến ứng dụng)

**Semantic features aim at encoding interpretations of the image which may be relevant to the application.**

Tất nhiên, những diễn giải như vậy là một tập hợp con của những diễn giải có thể có của một hình ảnh. Cuối cùng, hãy xem xét một vector đặc trưng  $F$  xuất phát từ hình ảnh  $i$ . Đối với các giải

thích ngữ nghĩa đã cho  $z$  từ tập hợp tất cả các giải thích  $Z$ , một giai đoạn học tập dẫn đến xác suất có điều kiện:

$$P = P(z|F) \quad \text{xác suất của ngữ nghĩa } z \text{ khi biết vector đặc trưng } F$$

Một đặc trưng ngữ nghĩa mạnh có diễn giải  $z_j$  sẽ tạo ra một  $P(z|F) = \delta(z - z_j)$  Nếu đặc trưng này không mang ngữ nghĩa, nó sẽ tạo ra một phân phối  $P(z|F) = P(z)$  độc lập với giá trị

của đặc trưng. Trong thực tế, nhiều loại đặc trưng sẽ tạo ra phân phối xác suất không phải là xung khắc cũng không độc lập với giá trị đặc trưng. Điều này có nghĩa là giá trị đặc trưng sẽ giải

thích hình ảnh, nhưng không xác định nó hoàn toàn.

Với định nghĩa **ngữ nghĩa yếu** (*weak semantics*), chúng tôi thu thập các phương pháp cố gắng kết hợp các đặc trưng trong một số giải thích có ý nghĩa về mặt ngữ nghĩa. Ngữ nghĩa yếu nhằm

mục đích mã hóa, theo cách đơn giản và gần đúng, một tập hợp con của các diễn giải có thể có của một hình ảnh được quan tâm trong một ứng dụng nhất định. Ví dụ, hệ thống Truy xuất hình ảnh bởi ngữ nghĩa màu, Trong Hệ thống đa phương tiện. Sử dụng các đặc trưng màu xuất phát từ lý thuyết màu sắc của **Itten** (Hệ màu HSV) để mã hóa ngữ nghĩa liên quan đến độ tương phản và hài hòa màu sắc trong ứng dụng nghệ thuật.

Trong hệ thống MAVIS2 (**Lưu trữ và truy xuất cho cơ sở dữ liệu Media**), dữ liệu được xem xét ở bốn cấp độ ngữ nghĩa, được thể hiện trong bốn lớp được gọi là **raw media** (dữ liệu thô), **selection layer**, **selection expression layer**, **conceptual layers**. Mỗi lớp mã hóa thông tin ở mức độ **symbolic** (đặc trưng) ngày càng cao (gần với con người). Các **Agents** (tác nhân) được đào tạo để tạo liên kết giữa các **đặc trưng**, **feature signatures** chữ ký đặc trưng ở **selection layer**, **interrelated signatures** (Chữ ký liên quan) ở lớp **selection expression layer** và **conceptual layers** (được biểu thị dưới dạng nhãn văn bản) ở **conceptual layers**. Ngoài các kết nối dọc, hai lớp trên cùng có các kết nối nội bộ đo lường sự tương đồng giữa các khái niệm ở cấp độ ngữ nghĩa đó và góp phần xác định sự tương đồng giữa các yếu tố ở cấp độ ngữ nghĩa thấp hơn.

## ▼ 5.2 Similarity between Features (tương đồng đặc trưng)

Một con đường khác để hiểu ý nghĩa cho một bộ đặc trưng được quan sát, là so sánh một cặp quan sát khác bằng một **similarity function** (hàm tương tự). Trong khi tìm kiếm hình ảnh truy vấn

$i_q(x)$  trong số các **element** (thành phần) của tập dữ liệu hình ảnh  $i_d(x)$ , **knowledge of the domain**

(kiến thức về miền) sẽ được thể hiện bằng cách xây dựng một phép đo tương tự  $S_{q,d}$  giữa các hình ảnh  $q$  và  $d$  trên cơ sở một số tập đặc trưng **feature set**. Thước đo tương tự phụ thuộc vào

loại đặc tính, xem Hình 10.

Độ tương tự của hai vectơ đặc trưng  $F$ , đặc trưng **accumulative** tích lũy hoặc đối tượng giống nhau, được đưa ra bởi:

$$S_{q,d} = s(F_q, F_d)$$

Ở mức sử dụng tốt nhất, phương pháp **similarity** (tương tự) có thể được thao tác để thể hiện các

nội dung ngữ nghĩa khác nhau; hình ảnh sau đó được nhóm lại bằng sự tương đồng bằng cách hình ảnh gần giống nhau về mục đích sử dụng hoặc mục đích của ứng dụng. Đáng ngạc nhiên là

có rất ít nghiên cứu dành riêng để mô tả các phương pháp tương tự. Tuy nhiên một vài ý tưởng,

đã xuất hiện. Một giả định phổ biến là sự giống nhau giữa hai vectơ đặc trưng  $F$  có thể được biểu thị bằng biểu thức sau:

$$s(F_q, F_d) = g \circ d(F_q, F_d)$$

Trong đó  $g$  là hàm dương, đơn điệu không tăng và  $d$  là hàm khoảng cách trên  $F$ . Giả định này phù hợp với một lớp các mô hình tâm lý về nhận thức tương tự của con người như hệ thống **Pattern Analysis and Machine Intelligence** (Phân tích mẫu và trí thông minh máy). Nếu không gian

đặc trưng là không gian vectơ,  $d$  thường là khoảng cách **Euclidean** đơn giản, mặc dù có dấu hiệu cho thấy các phương pháp khoảng cách phức tạp hơn có thể tốt hơn. **Similarity Model** (Mô hình

tương tự) này rất phù hợp cho truy vấn sớm bởi các hệ thống ví dụ trong đó các hình ảnh được sắp xếp theo độ tương tự với một ví dụ (câu truy vấn).

Một quan điểm khác xem sự tương đồng là một khái niệm xác suất cơ bản. Quan điểm này bắt

nguồn từ tài liệu **Toward a Unified Theory of Similarity and Recognition** (Hướng tới một lý thuyết

thống nhất về sự tương đồng và nhận diện), và **content-based retrieval** (truy xuất dựa trên nội dung đã được đề xuất, ví dụ, trong **Beyond Eigenfaces: Probabilistic Matching for Face Recognition**

(xác suất matching cho nhận diện khuôn mặt). Một hình thức chung của một biện pháp tương tự như vậy sẽ là:

$$s(F_q, F_d) = f(P(\hat{F}_q \approx F_d))$$

Trong đó  $\approx$  có nghĩa là hai đặc trưng mô tả hình ảnh của cùng một **Class** và  $\hat{F}_q = \bar{F}_q + v$  trong đó  $\bar{F}_q$  là xác suất thực và  $v$  nhiễu do điều kiện đo lường và **sensory**.

Đo khoảng cách giữa các **histograms** là một nhánh nghiên cứu tích cực kể từ những năm đầu giai đoạn này. Trong đó **histograms** có thể được xem là một tập hợp các đặc trưng được sắp xếp:

$$s(F^q, F^d) = g \circ d(F^q, F^d)$$

Trong truy xuất dựa trên nội dung, **histograms** chủ yếu được sử dụng cùng với các đặc trưng màu sắc,

nhưng không sử dụng các thông tin kết cấu hoặc thuộc tính hình học cục bộ. Swain và Ballard

trong bài báo [169] đề xuất sử dụng **intersection distance**(khoảng cách giao)  $d_{\cap}(F^q, F^d) = \sum_{i=1}^n \min(F_j^q, F_j^d)$  trong đó  $F_q, F_d$  là 2 histogram chứa  $n$  bin màu. Họ cũng chứng minh rằng

nếu tất cả các hình ảnh có cùng số pixel ví dụ  $\sum_j F_j$  là giống nhau cho tất cả các hình ảnh, sau đó

khoảng cách này có các thuộc tính thứ tự giống như khoảng cách L1. Trong bài báo [62], một cách

tiếp cận khác được mô tả như sau. Khoảng cách giữa hai biểu đồ được xác định ở dạng vector là:

$$d_{\Sigma}(F^q, F^d) = \sqrt{(F^q - F^d)^t \Sigma (F^q - F^d)}$$

Trong đó ma trận  $\Sigma$  biểu thị sự giống nhau giữa các **bin** thứ  $j$  và thứ  $k$ . Điều này có lợi thế là xem xét sự tương đồng giữa các giá trị trong không gian đặc trưng, tức là kết hợp số liệu của không gian đặc trưng vào thước đo tương tự.

Các hàm khoảng cách thường được sử dụng khác cho biểu đồ màu bao gồm khoảng cách **Minkowski** :

$$d_r(F^q, F^d) = \left[ \sum_{i=1}^n |F_j^q - F_j^d|^r \right]^{\frac{1}{r}}$$

Các biện pháp này không tính đến sự giống nhau giữa các **bin** khác nhau, nhưng có liên quan **histogram**. Trong [166], điều này được quan sát thấy rằng điều này có thể dẫn đến sai lầm.

Bài viết đề xuất việc sử dụng **cumulative histogram** (tích lũy) có dạng  $\tilde{F}(m) = \sum_{k=0}^m F_k^q$  sau khi sắp xếp các bin theo tham số  $j$ .

So sánh giữa các **cumulative histogram** (biểu đồ tích lũy) thay vì **histogram** đơn giản cho thấy

rằng trước đây có xu hướng phạt ít hơn cho những thay đổi trong việc gán **bin** do nhiều. Một cách

khác, cũng được khám phá trong bài báo, là mô tả histogram bằng mean, Variance, Skewness, trong đó 3 là một phát hiện thực nghiệm. Trong [166], **histogram** được áp dụng cho hình ảnh màu bằng cách thể hiện màu sắc trong hệ thống **HSV** và tính toán từng kênh một cách riêng biệt, dẫn đến chín tham số, ba khoảng khắc cho mỗi trong ba kênh màu. Một thước đo khoảng cách

gần đây cho biểu đồ màu được tìm thấy trong [4], trong đó **histogram** màu sắc và **histogram** bão hòa được hình thành riêng biệt với các ưu điểm của việc lưu vào bộ nhớ và khả năng loại trừ màu khỏi truy vấn. Tham chiếu so sánh màu sắc trên cơ sở khoảng cách góc trong không gian RGB.

Biện pháp tự nhiên để so sánh các tập hợp các đặc trưng tích lũy (**accumulative features**) được

xếp là số liệu thống kê kiểm tra không theo quy chuẩn. Chúng có thể được áp dụng cho các phân phối của các hệ số biến đổi để xác định khả năng hai mẫu xuất phát từ cùng một phân phối [35], [131]. Chúng cũng có thể được áp dụng để so sánh sự bằng nhau của hai **histograms** và tất cả các biến thể của chúng.

### ▼ 5.3 Similarity of Object Silhouettes (tương đồng hình dạng)

Trong [183], một đánh giá tốt được đưa ra các phương pháp để so sánh các hình dạng trực tiếp

sau khi phân đoạn thành một tập hợp các điểm đối tượng  $t(x)$ :

$$S_{q,d} = s(t_q(x), t_d(x)) \quad (15)$$

mà không có một mô tả trung gian về các đặc trưng của hình dạng (shape).

Để so sánh hình dạng (**shape**), các tác giả phân biệt giữa biến đổi, khoảnh khắc **moments**, khớp

biến dạng (**deformation matching**), khớp không gian tỷ lệ và đo lường độ không giống nhau. Khó khăn cho việc khớp hình dạng dựa trên các biến đổi toàn cục là sự không thể giải thích được

của kết quả và **brittleness** cho các độ lệch nhỏ. Khoảnh khắc, cụ thể là sự kết hợp bất biến của

chúng, thường được sử dụng trong truy xuất [89]. Việc khớp một truy vấn và một đối tượng trong

tập dữ liệu có thể được thực hiện dọc theo tập hợp các hình dạng **eigen** được sắp xếp [150] hoặc

với khớp khớp đàn hồi [36], [11]. Kết hợp không gian tỷ lệ dựa trên việc đơn giản hóa dần dần biên cạnh bằng cách làm mịn [116]. Bằng cách so sánh chữ ký của các đường chéo không bị hủy

của độ cong, hai hình dạng được khớp với nhau theo tỷ lệ và xoay bất biến. Một chất tương tự rời rạc có thể được tìm thấy trong [94], trong đó các điểm được loại bỏ khỏi biên cạnh số hóa trên

cơ sở các quy tắc có động lực nhận thức. Kết quả trên cơ sở dữ liệu 2.000 hình ảnh được báo cáo

là hoạt động tốt hơn hầu hết các phương pháp được liệt kê ở trên.

Khi dựa trên một số liệu, các biện pháp khác nhau sẽ đưa ra một phạm vi sai lệch có trật tự phù hợp với cách giải thích có thể dự đoán được. Trong [183], một phân tích được đưa ra cho **Hausdorff** và các số liệu liên quan giữa hai hình dạng về độ mạnh và độ phức tạp tính toán.

Chỉ

số **Hausdorff** được định hướng được định nghĩa là khoảng cách tối đa giữa một điểm trên

đối tượng truy vấn  $q$  và đối tác gần nhất của nó trên  $d$ . Số liệu một phần của **Hausdorff**, được định nghĩa là mức tối đa thứ  $k$  thay vì mức tối đa tuyệt đối, được sử dụng trong [63] để truy xuất biến đổi **affine**.

## 5.4 Similarity of Structural Features (tương đồng cấu trúc đặc trưng)

Kết quả của một mô tả cấu trúc là một tập hợp các giá trị đặc trưng  $H$  được sắp xếp phân cấp, xem Hình 9. Trong phần này, chúng ta xem xét sự giống nhau giữa hai mô tả cấu trúc hoặc bố cục.

### structural or layout descriptions

$$S_{q,d} = s(H_q, H_d)$$

Nhiều kỹ thuật khác nhau đã được nghiên cứu cho sự giống nhau của cấu trúc đặc trưng. Trong

[191], [74], một **framework Bayes** được phát triển để **matching** với các đồ thị quan hệ **attributed**

bằng cách rời rạc **relaxation**. Điều này được áp dụng cho các **line patterns** từ các bức ảnh chụp từ trên không.

Một số liệu để so sánh hai sắp xếp **tôpô** của các bộ phận được đặt tên, áp dụng cho hình ảnh y tế,

được nói đến ở trong trong [172]. Khoảng cách được lấy từ số bước chỉnh sửa cần thiết để 2 ảnh giống nhau về **Voronoidiagrams** của hai hình ảnh.

Trong [23], các  $2D$  — *strings* mô tả mối quan hệ không gian giữa các đối tượng đã được thảo luận

và sau đó được xem xét lại trong [198]. Từ các mối quan hệ cấu trúc liên kết như vậy của các vùng

hình ảnh, trong [71] một chỉ mục 2D được xây dựng trong các cây chuỗi ký hiệu ( **symbol strings**),

mỗi chuỗi biểu thị hình chiếu của một vùng trên trục tọa độ. Khoảng cách giữa  $H_q$  và  $H_d$  là số lượng

thao tác chỉnh sửa được cân nhắc cần thiết để chuyển đổi cây này thành cây khác. Trong [153],

một đồ thị được hình thành từ hình ảnh trên cơ sở đối xứng khi nó xuất hiện từ trục trung gian.

Độ tương tự được đánh giá theo hai giai đoạn thông qua khớp dựa trên đồ thị, sau đó là khớp biến dạng năng lượng **energy-deformation matching**.

Trong [53], các cây được sắp xếp phân cấp được so sánh với mục đích truy xuất bằng cách viết lại chúng thành chuỗi. Một thước đo tương tự dựa trên khoảng cách thiết lập điểm tương đồng giữa các node lá tương ứng trong cây. Ở các node nội của cây, tổng số điểm tương đồng của các nhánh tương ứng được lấy làm thước đo cho độ tương tự của cây (phụ). Từ một thử nghiệm kích thước nhỏ, kết luận rằng các bộ đặc trưng được sắp xếp phân cấp có hiệu quả hơn các bộ đặc trưng đơn giản, với các tính toán nhanh được cho các bộ dữ liệu lớn hơn.

Trong [163], hình ảnh được chuyển thành các vùng đồng nhất để truy xuất dựa trên bố cục màu. Các vùng được quét, thường là 5 lần quét dọc cách đều nhau và được chuyển đổi thành một chuỗi các ký hiệu được lấy từ một từ điển **visual**(trực quan). Các chuỗi được tóm tắt thành **histograms** vùng,  $F_{(i,j)}$ , cho biết số lần một **symbol** đứng trước một **symbol** khác trong một lần quét. Trong quá trình truy vấn, sự tương tự của  $q$  với  $d$  được tính bởi  $\sum_i (F^q, F^d)^{-1} \sum_j (F^q, F^d)$ , nghĩa là sự tương ứng giữa các **element** của biểu đồ **histograms** theo thứ tự từng vùng.

## ▼ 5.5 Similarity of Salient Features (tương đồng đặc trưng nổi bật)

Các đặc trưng nổi bật được sử dụng để ghi lại thông tin trong ảnh ở một số điểm giới hạn. Sự giống nhau giữa các hình ảnh sau đó có thể được kiểm tra theo nhiều cách khác nhau. Đầu tiên, màu sắc, kết cấu hoặc đặc điểm hình dạng cục bộ (local) có thể được sử dụng để xác định các điểm nổi bật của dữ liệu giống hệt với các điểm nổi bật của truy vấn.

$$S_{q,d} = g \circ d(F_q, F_d) \quad (17)$$

trong đó  $F_q$  và  $F_d$  là các vectơ đặc trưng của các thuộc tính nổi bật và  $g$  là một hàm đơn điệu tùy chọn. Một thước đo tương tự giữa các giá trị đặc trưng được đo của các điểm màu do phân đoạn yếu bao gồm khoảng cách **Mahalanobis** giữa vectơ đặc trưng bao gồm màu sắc, kết cấu, vị trí, diện tích, độ lệch tâm và hướng của hai hình elip [20]. Nếu các đặc trưng của hình **elip** được thu thập trong một vectơ  $F$ , khoảng cách giữa  $q$  và  $d$  được tính bởi.

$$d_{q,d} = \left[ (F_q - F_d)^T \Sigma^{-1} (F_q - F_d) \right]^{\frac{1}{2}}$$

trong đó  $\Sigma$  là một ma trận trọng số đường chéo được thiết lập bởi người dùng. Sự giống nhau giữa hai điểm màu được định nghĩa là:

$$S_{q,d} = \exp(-d_{q,d}/2)$$

Ở vị trí thứ hai, người ta có thể lưu trữ tất cả các điểm nổi bật từ một hình ảnh trong **histogram** dựa trên một số đặc điểm, chẳng hạn như màu sắc ở bên trong so với màu sắc bên ngoài. Sự giống nhau sau đó dựa trên sự hiện diện theo nhóm của các điểm tương tự [57].

$$S_{q,d} = g \circ d(F^q, F^d) \quad (18)$$

trong đó  $F_q$  và  $F_d$  là **histograms** chỉ cho thấy sự hiện diện của các điểm nổi bật. Số liệu  $d(F^q, F^d)$

hiện đang nhằm đến việc đo lường sự hiện diện của cùng một tập hợp các điểm nổi bật. So sánh

**histograms** thừa thớt từ lâu đã được sử dụng trong truy xuất văn bản, trong đó mô hình không gian vectơ [143] ngụ ý đăng ký trong  $N$  chiều **histogram**  $F$  với nhiều chiều khác nhau trong từ điển,

thường là 10.000. Trong một không gian vectơ nhị phân, mỗi *bit* biểu thị cho từ đó có mặt hay không trong văn bản. Một văn bản là một điểm trong không gian nhiều chiều này. Sự khác nhau

giữa văn bản  $d$  trong tệp dữ liệu và truy vấn  $q$  rút ngắn khoảng cách giao nhau đã thảo luận ở trên

khoảng cách được tính như sau:

$$d_{\cap}(F^q, F^d) = \sum_{1 \dots i \dots n} \cap_i F^d(\dots i \dots) F^q(\dots i \dots)$$

Tổng khoảng cách của tất cả các chiều **dimensions**. Chiến lược tương tự được sử dụng khi so sánh các đặc trưng nổi bật bắt nguồn từ các hình ảnh khác nhau. Giao điểm là thích hợp khi cả  $q$  và  $d$  có thể bị che khuất một phần trong ảnh hoặc lộn xộn với nền. Khi  $q$  không bị lộn xộn hoặc bị chặn nhưng  $d$  vẫn có thể bị lộn xộn, giao điểm nên được thay thế bằng  $< -operation$ .

Mô hình đã được sử dụng trong truy xuất hình ảnh trong [158], trong khi vẫn giữ quyền truy cập

vào vị trí của chúng trong hình ảnh bằng cách chiếu ngược [169]. Theo sự phát triển của mô hình

không gian vectơ trong truy xuất văn bản, trọng số trên mỗi chiều có thể ủng hộ sự xuất hiện của



một số đặc trưng nổi bật hơn một đặc trưng khác. Xem thêm [69] để so sánh với biểu đồ tương quan.

Cách thứ ba cho sự tương đồng của các điểm nổi bật là chỉ tập trung vào các mối quan hệ không gian giữa các điểm nổi bật trong tập  $P_q$  và  $P_d$ .

$$S_{q,d} = g \circ d(P_q, P_d) \quad (19)$$

Trong các phương pháp dựa trên từng điểm để so sánh hình dạng, độ tương tự hình dạng (shape)

được nghiên cứu trong [83], trong đó các điểm cong tối đa trên đường viền và độ dài giữa chúng

được sử dụng để mô tả vật thể. Để tránh các tính toán mở rộng, người ta có thể tính toán các tập

hợp điểm bất biến, được gọi là **crossratio**. Do đặc tính bất biến của chúng, các biện pháp này có

xu hướng chỉ có sức mạnh phân biệt những dấu đặc biệt giữa các đối tượng khác nhau. Một phiên

bản gần đây hơn cho sự giống nhau của các tập hợp điểm không tên được tìm thấy trong băm hình học **geometric hashing** [192], trong đó mỗi bộ ba kéo dài một cơ sở cho các điểm còn lại của đối tượng. Một đối tượng không xác định được so sánh trên mỗi bộ ba để xem liệu có đủ các

điểm nằm tương tự được tìm thấy hay không. Băm hình học, mặc dù hấp dẫn trong khái niệm của nó,

Chi phí cao về mặt tính toán để được sử dụng trên các tập dữ liệu rất lớn của truy xuất hình ảnh

do tính ẩn danh của các điểm. Độ tương tự của hai tập điểm  $P_q$  và  $P_d$  được đưa ra trong ma trận

dòng được thảo luận trong [188]. Một khoảng cách được đưa ra cho sự tương tự bất biến

$$D^2(P_q, P_d) = 1 - \frac{\|P_q P_d^T\|^2 + 2 \det(P_q P_d^T)}{\|P_d\|^2 \|P_q\|^2}$$

cho các phép biến đổi affine:

$$D^2(P_q, P_d) = 2 - \text{tr}(P_q^+ P_d + P_q^+ P_d)$$

Trong đó  $P_d^+$  và  $P_q^+$  lần lượt là giả nghịch đảo (**pseudoinverse**) của  $P_d$  và  $P_q$ .

## ▼ 5.6 Similarity at the Semantic Level (tương đồng ngữ nghĩa)

Trong [70], hệ thống phân cấp trừu tượng dựa trên kiến thức được sử dụng để truy cập dữ liệu hình ảnh dựa trên ngữ cảnh và hồ sơ người dùng, được tạo tự động từ phân tích cụm cơ sở dữ liệu. Cũng trong [24], mục tiêu là tạo ra một không gian khái niệm rất lớn lấy cảm hứng từ tìm kiếm dựa trên từ điển đồng nghĩa từ cộng đồng truy xuất thông tin. Trong [115], một mô tả ngôn ngữ về phẩm chất hình ảnh kết cấu được đưa ra và sắp xếp theo thứ bậc về tầm quan trọng nhận thức trên cơ sở thử nghiệm tâm lý sâu rộng. Một khái niệm tổng quát hơn về sự tương tự là cần thiết cho phản hồi liên quan, trong đó sự tương đồng liên quan đến một tập hợp hình ảnh là bắt buộc. Cuối cùng, trong [45], các mối quan hệ phức tạp hơn được trình bày giữa các hàm tương tự và khoảng cách xác định một số đo có trọng số của hai điểm tương đồng đơn giản hơn.

$$S(s, S_1, S_2) = w_1 \exp(-d(S_1, s)) + w_2 \exp(-d(S_2, s))$$

Mục đích của biện pháp lưỡng chiết là tìm tất cả các vùng tương tự với hai điểm truy vấn đã chỉ định, một ý tưởng khái quát cho các truy vấn tương tự được đưa ra nhiều ví dụ. Cách tiếp cận có thể được mở rộng với định nghĩa về đại số hoàn chỉnh các biện pháp tương tự với các toán tử thành phần phù hợp [45], [38]. Sau đó, có thể định nghĩa các toán tử tương ứng với sự phân tách, kết hợp và phủ định các biện pháp tương tự, giống như các cơ sở dữ liệu truyền thống. Đại số rất hữu ích cho người dùng để thao tác trực tiếp sự giống nhau như một phương tiện để thể hiện các đặc tính trong các giá trị tính năng cụ thể.

## ▼ 5.7 Learning an Interpretation (học từ một diễn giải)

Khi các bộ dữ liệu phát triển lớn và khả năng xử lý có sẵn phù hợp với sự tăng trưởng đó, cơ hội sẽ nảy sinh để học hỏi kinh nghiệm. Thay vì thiết kế, thực hiện và thử nghiệm một thuật toán để phát hiện các đặc điểm trực quan cho từng thuật ngữ ngữ nghĩa khác nhau, có thể tìm hiểu ngữ nghĩa của các đối tượng từ sự xuất hiện của chúng.

Để xem xét về nhận dạng mẫu thống kê, xem [80]. Trong [182], một loạt các kỹ thuật được thảo luận coi việc thu hồi là một vấn đề phân loại.

Một cách tiếp cận là phân tích thành phần chính qua một chồng hình ảnh được lấy từ cùng một lớp  $z$  của các đối tượng. Điều này có thể được thực hiện trong không gian đặc trưng [118] hoặc ở cấp độ của toàn bộ hình ảnh, ví dụ, các mặt trong [113]. Phân tích mang lại một tập hợp các hình ảnh "eigenface", ghi lại các đặc điểm chung của khuôn mặt mà không cần mô hình hình học.

Các cách hiệu quả để học hỏi từ dữ liệu được dán nhãn một phần gần đây đã được giới thiệu trong [194], [139], cả hai đều sử dụng nguyên tắc tải nạp [181]. Điều này tiết kiệm nỗ lực ghi nhãn toàn bộ tập dữ liệu, không khả thi và không đáng tin cậy khi nó phát triển lớn.

Trong [186], công việc sơ bộ được báo cáo theo hướng tự động phát hiện các danh mục trên các tập dữ liệu hoàn toàn không được gắn nhãn. Chúng đại diện cho các đối tượng như các

chòm sao xác suất của các tính năng. Các phần cứng nhắc định kỳ được chọn tự động bằng cách tối đa hóa kỳ vọng.

Trong [176], một số lượng lớn các tính năng được tính toán trước được xem xét, trong đó một tập hợp con nhỏ được chọn bằng cách tăng [80] để tìm hiểu lớp hình ảnh.

Một kỹ thuật thú vị để thu hẹp khoảng cách giữa mô tả văn bản và hình ảnh để khai thác thông tin ở cấp độ tài liệu được mượn từ truy xuất thông tin, được gọi là lập chỉ mục ngữ nghĩa tiềm ẩn [151], [199]. Đầu tiên, một kho văn bản được hình thành từ các tài liệu (trong trường hợp này là hình ảnh có chú thích) từ đó các tính năng được tính toán. Sau đó, bằng cách phân tách giá trị số ít, từ điển bao gồm các chú thích có tương quan với các tính năng thu được từ các hình ảnh. Việc tìm kiếm là cho các mối tương quan ẩn của các tính năng và chú thích.

## ▼ 5.8 Discussion on Interpretation and Similarity

(Thảo luận về giải thích và tương đồng) Bất cứ khi nào hình ảnh cho phép giải thích rõ ràng, hệ thống dựa trên nội dung lý tưởng nên sử dụng thông tin đó. Một diễn giải ngữ nghĩa mạnh xảy ra khi một dấu hiệu có thể được xác định tích cực trong hình ảnh. Điều này hiếm khi xảy ra do sự đa dạng lớn của các dấu hiệu trong một lớp hình ảnh rộng lớn và sự to lớn của nhiệm vụ xác định thuật toán phát hiện đáng tin cậy cho mỗi trong số chúng. Ngữ nghĩa yếu dựa vào phân loại không chính xác gây ra bởi các biện pháp tương tự, tốt nhất là trực tuyến bằng tương tác. Việc phân loại có thể đồng ý với các khái niệm ngữ nghĩa của người dùng, nhưng nói chung, thỏa thuận này không hoàn hảo. Do đó, việc sử dụng ngữ nghĩa yếu thường được kết hợp với khả năng đưa ngữ nghĩa của người dùng theo nhu cầu của mình bằng cách giải thích. Ngữ nghĩa có thể điều chỉnh có thể sẽ nhận được nhiều sự chú ý hơn trong tương lai, đặc biệt là khi các bộ dữ liệu phát triển lớn.

Sự giống nhau là một cách giải thích của hình ảnh dựa trên sự khác biệt với hình ảnh khác. Đối với mỗi loại tính năng, cần có một biện pháp tương tự khác nhau. Đối với sự tương đồng giữa các bộ tính năng, đặc biệt chú ý đến việc thiết lập sự tương đồng giữa các biểu đồ do hiệu quả tính toán và hiệu quả truy xuất của chúng.

Sự giống nhau của hình dạng đã nhận được sự chú ý đáng kể trong bối cảnh truy xuất dựa trên đối tượng. Nói chung, các lược đồ khớp hình dạng toàn cầu bị phá vỡ khi có sự tắc nghẽn hoặc lộn xộn trong cảnh. Hầu hết các phương pháp so sánh hình dạng toàn cầu đều ngầm định yêu cầu quan điểm trực diện dựa trên nền đủ rõ ràng để đạt được phân đoạn đủ chính xác. Với sự bao gồm gần đây của các điểm mạnh về mặt nhận thức trong hình dạng của các vật thể, một bước tiến quan trọng đã được thực hiện.

Sự giống nhau của các mô tả theo thứ tự đáng được chú ý vì đây là một cơ chế để tránh các vấn đề với phân đoạn trong khi duy trì một số mối quan hệ có ý nghĩa về mặt ngữ nghĩa trong

hình ảnh. Một phần của khó khăn ở đây là cung cấp sự phù hợp của các nhiễu loạn một phần. Học các mô hình tính toán cho ngữ nghĩa là một cách tiếp cận thú vị và tương đối mới. Nó thu hút sự chú ý nhanh chóng khi các bộ dữ liệu và sức mạnh máy móc phát triển lớn. Việc học mở ra khả năng diễn giải hình ảnh mà không cần thiết kế và thử nghiệm máy dò cho mỗi khái niệm mới. Một cách tiếp cận như vậy là học tập dựa trên ngoại hình về các đặc điểm chung của ngăn xếp hình ảnh từ cùng một lớp. Học tập dựa trên ngoại hình phù hợp cho các lĩnh vực hẹp. Để thành công của phương pháp học tập, có một sự đánh đổi giữa việc chuẩn hóa các đối tượng trong tập dữ liệu và kích thước của tập dữ liệu. Dữ liệu càng được chuẩn hóa thì càng cần ít dữ liệu, nhưng mặt khác, kết quả sẽ càng ít được áp dụng rộng rãi. Các cách tiếp cận thú vị để rút ra các lớp ngữ nghĩa từ chú thích hoặc bộ dữ liệu được gắn nhãn một phần hoặc không được gắn nhãn đã được trình bày gần đây, xem ở trên.

## 6.Interaction