

Báo cáo: Ứng dụng Template Matching trong Nhận dạng và Đếm đối tượng

Hoàng Đăng Khoa

Mục lục

1	Giới thiệu	1
2	Phương pháp tiếp cận	2
2.1	Template Matching	2
2.2	Non-Maximum Suppression	3
3	Kết quả	3
3.1	Công cụ và dữ liệu	3
3.2	Kết quả	4

1 Giới thiệu

Bài báo cáo này trình bày về quá trình sử dụng thuật toán khớp mẫu (template matching) để giải quyết hai tác vụ quan trọng trong xử lý ảnh: nhận diện đối tượng (Object Finding) và đếm số lượng đối tượng (Object Counting).

- **Nhận diện đối tượng (Object Finding):** Tác vụ này yêu cầu xác định vị trí của một hoặc nhiều đối tượng đã được định nghĩa trước trong một bức ảnh. Mẫu đối tượng được cung cấp trước và hệ thống phải tìm kiếm vị trí của nó trong ảnh đầu vào.



Hình 1: Ví dụ bài toán Nhận diện đối tượng

- **Đếm số lượng đối tượng (Object Counting):** Tác vụ này yêu cầu đếm số lượng xuất hiện của một loại đối tượng cụ thể trong bức ảnh, chẳng hạn như đếm số con mèo. Để thực hiện, cần lựa chọn một mẫu đại diện cho tất cả các đối tượng thuộc loại này.

Thuật toán template matching là một phương pháp hiệu quả để giải quyết cả hai tác vụ. Dối với bài toán nhận diện đối tượng, mẫu đối tượng (template) được cung cấp sẵn. Tuy nhiên đối với bài toán đếm số lượng, việc lựa chọn một mẫu chuẩn cho toàn bộ các đối tượng cần đếm là yếu tố quan trọng để đạt được độ chính xác cao.

2 Phương pháp tiếp cận

Trong phần này, chúng tôi trình bày chi tiết phương pháp tiếp cận để giải quyết bài toán nhận diện và đếm đối tượng, bao gồm hai kỹ thuật chính: thuật toán khớp mẫu (Template Matching) và Non-Maximum Suppression (NMS).

2.1 Template Matching

Thuật toán khớp mẫu (Template Matching) là phương pháp sử dụng để tìm kiếm một mẫu (template) trong một hình ảnh lớn. Chúng tôi sử dụng *Normalized Cross-Correlation* (NCC) để đánh giá mức độ tương đồng giữa mẫu và các vùng của hình ảnh.

Giả sử T là mẫu cần tìm và I là bức ảnh đầu vào, kích thước của T là $w \times h$. Khi đó độ tương quan giữa mẫu T và các điểm ảnh trong vùng $I(x : x + w, y : y + h)$, ký hiệu là $NCC(x, y)$, sẽ được tính bằng công thức sau:

$$NCC(x, y) = \frac{\sum_{i=0}^{w-1} \sum_{j=0}^{h-1} [I(x+i, y+j) - \mu_I] [T(i, j) - \mu_T]}{\sqrt{\sum_{i=0}^{w-1} \sum_{j=0}^{h-1} [I(x+i, y+j) - \mu_I]^2 \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} [T(i, j) - \mu_T]^2}}$$

Trong đó:

- μ_I là giá trị trung bình của các điểm ảnh trong vùng $I(x : x + w, y : y + h)$ của bức ảnh I .
- μ_T là giá trị trung bình của mẫu T .
- $I(x+i, y+j)$ là giá trị của điểm ảnh tại vị trí $(x+i, y+j)$ trong ảnh I .
- $T(i, j)$ là giá trị của điểm ảnh tại vị trí (i, j) trong mẫu T .

Giá trị của NCC nằm trong khoảng từ -1 đến 1 . Một giá trị gần 1 cho thấy sự tương đồng cao giữa mẫu và vùng ảnh, trong khi giá trị gần -1 chỉ ra sự tương phản.

Ngoài ra, để đảm bảo phép tính Normalized Cross-Correlation chỉ được thực hiện trên các vùng có ý nghĩa (ví dụ như không tính toán trên các vùng nền trắng hoặc không chứa thông tin hữu ích), chúng tôi áp dụng thêm bước masking. Bước này cho phép chỉ thực hiện tính toán trên các pixel mà giá trị của mặt nạ (mask) là 1 . Khi đó công thức NCC trở thành như sau:

$$NCC(x, y) = \frac{\sum_{i=0}^{w-1} \sum_{j=0}^{h-1} M(i, j) [I(x+i, y+j) - \mu_I] [T(i, j) - \mu_T]}{\sqrt{\sum_{i=0}^{w-1} \sum_{j=0}^{h-1} M(i, j) [I(x+i, y+j) - \mu_I]^2 \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} M(i, j) [T(i, j) - \mu_T]^2}}$$

với $M(i, j)$ là giá trị của ma trận mask tại vị trí (i, j) . Với $M(i, j) = 1$ chỉ những pixel nằm trong vùng cần tính NCC, và $M(i, j) = 0$ chỉ những pixel bị loại bỏ khỏi phép tính.

2.2 Non-Maximum Suppression

Sau khi áp dụng thuật toán khớp mẫu, có thể tồn tại nhiều vùng trong ảnh có điểm số NCC cao trùng lặp với nhau. Do đó, tôi sử dụng kỹ thuật Non-Maximum Suppression (NMS) để loại bỏ các vùng trùng lặp, chỉ giữ lại các điểm có giá trị NCC cao nhất tại mỗi vùng.

Quá trình NMS được thực hiện như sau: 1. Xác định các vùng có giá trị NCC lớn hơn một ngưỡng τ . 2. Sắp xếp các vùng theo thứ tự giảm dần của điểm NCC. 3. Duyệt qua danh sách các vùng, với mỗi vùng được chọn ta loại bỏ các vùng có IOU với vùng đó vượt quá một ngưỡng cho trước (Intersection over Union). Công thức tính IoU giữa hai vùng A và B là:

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

4. Lặp lại bước 3 cho đến khi không còn vùng nào cần loại bỏ

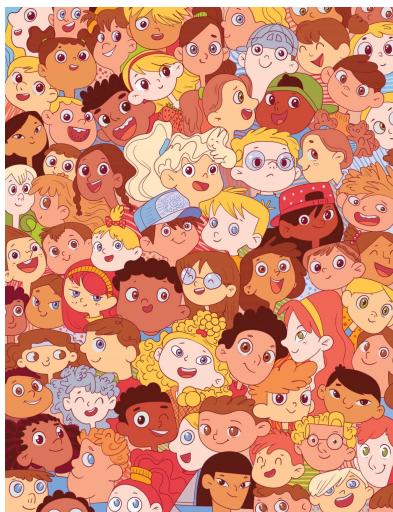
Quá trình này giúp đảm bảo rằng chỉ có một đối tượng được giữ lại tại mỗi vùng trong ảnh, giúp tăng độ chính xác trong việc nhận diện và đếm đối tượng.

3 Kết quả

3.1 Công cụ và dữ liệu

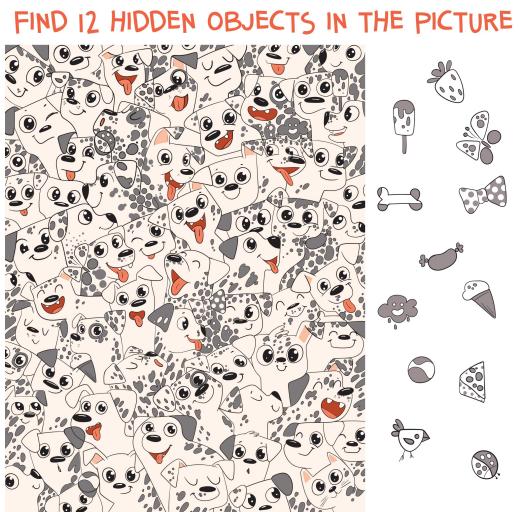
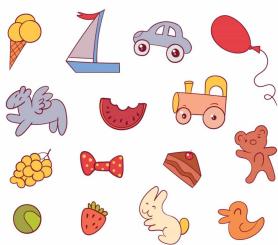
Trong thí nghiệm này, chúng tôi áp dụng thuật toán Normalized Cross-Correlation với hàm ‘cv2.matchTemplate’ trong thư viện OpenCV, cùng với tham số cv2.TM_CCOEFF_NORMED để đánh giá độ tương đồng giữa mẫu và ảnh đầu vào. Để loại bỏ các phát hiện trùng lặp, chúng tôi sử dụng kỹ thuật Non-Maximum Suppression (NMS) với giá trị threshold cho IOU là 0.05.

Các thí nghiệm được thực hiện trên hai bài toán nhận diện đối tượng (Object Finding) và đếm số lượng đối tượng (Object Counting). Cụ thể, chúng tôi thử nghiệm thuật toán NCC và NMS trên hai ví dụ cho bài toán nhận diện đối tượng (Hình 5) và hai ví dụ cho bài toán đếm số lượng đối tượng (Hình 3).



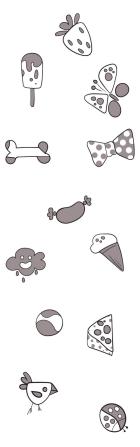
(a) Ví dụ 1

FIND
15
HIDDEN
OBJECTS
IN THE
PICTURE



(b) Ví dụ 2

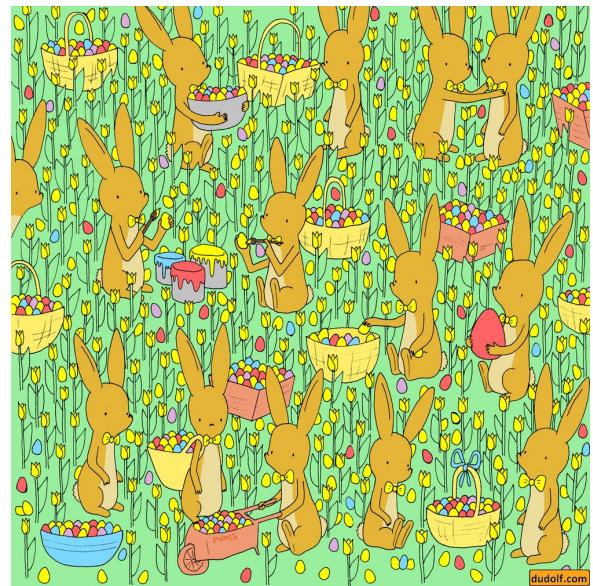
FIND 12 HIDDEN OBJECTS IN THE PICTURE



Hình 2: Hai ví dụ minh họa bài toán Nhận diện đối tượng



(a) Ví dụ 1: Dếm số lượng đôi ủng

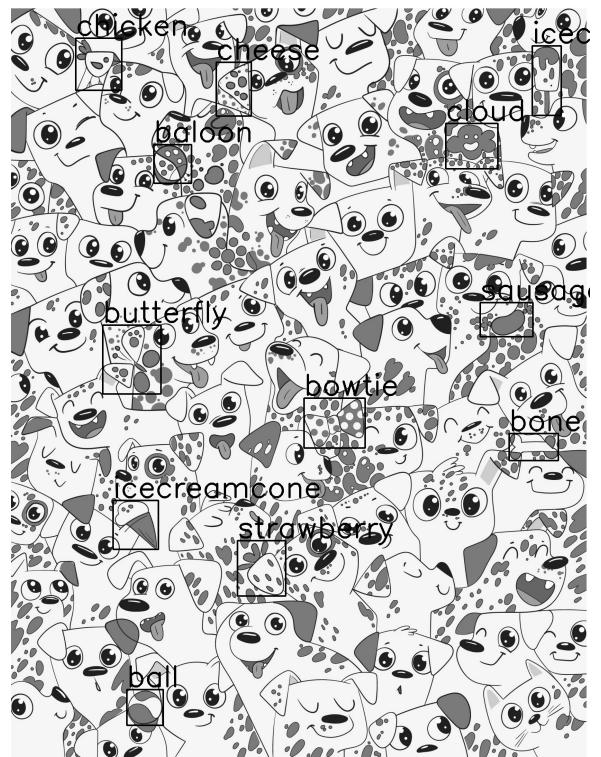
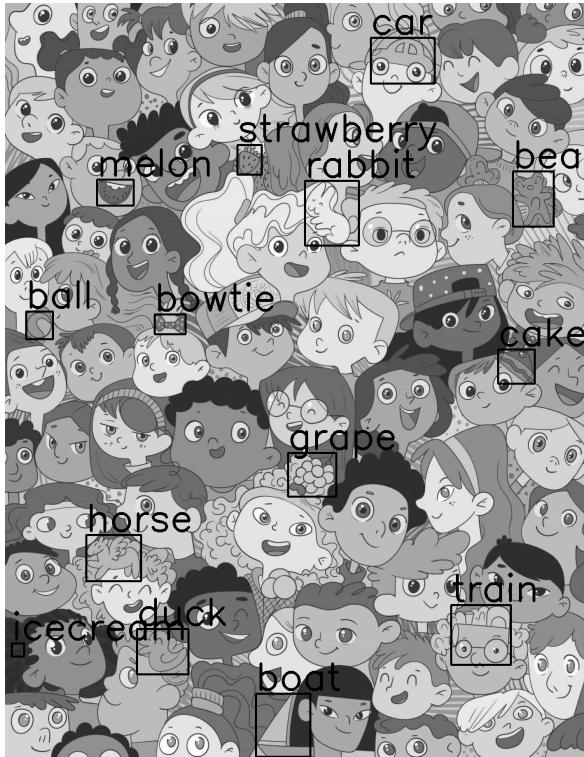


(b) Ví dụ 2: Dếm số lượng con thỏ

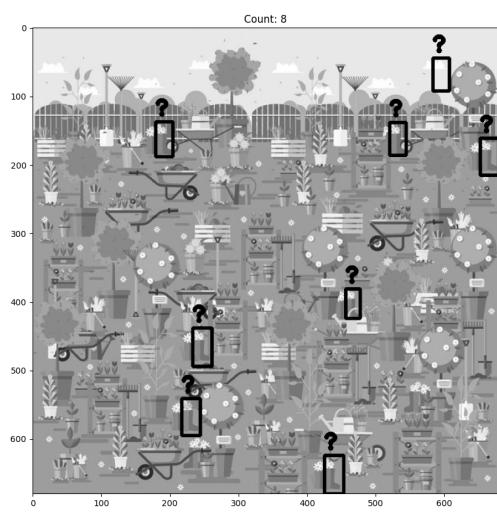
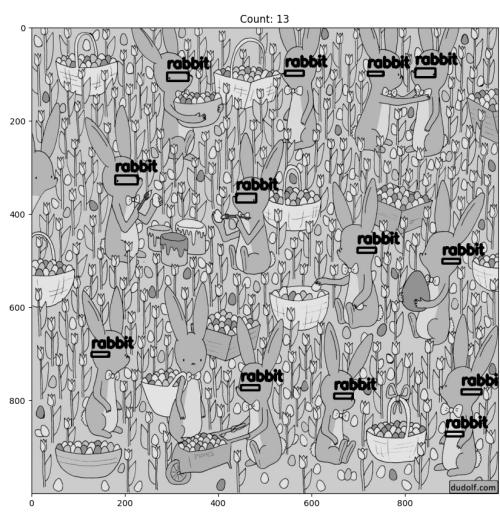
Hình 3: Hai ví dụ minh họa bài toán Dếm số lượng đối tượng

Có thể thấy rằng đối với bài toán Nhận diện đối tượng các mẫu cho trước và hình ảnh tương ứng trong hình có kích cỡ khác nhau. Chính vì vậy thuật toán được lặp lại với các scale khác nhau của mẫu để giúp tăng độ chính xác trong quá trình nhận diện.

3.2 Kết quả



Hình 4: Kết quả trên hai ví dụ bài toán Nhận diện đối tượng



Hình 5: Kết quả trên hai ví dụ bài toán Tìm kiếm đối tượng