TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT

KHOA ĐÀO TẠO CHẤT LƯỢNG CAO



**Đề Tài Cuối Kỳ**

Matching neural paths: transfer from recognition tocorrespondence search

SV thực hiện Nguyễn Đình Quyền - 15110115

GV hướng dẫn Nguyễn Thiên Bảo

TP.HCM, 2018

**Mục Lục**

Trang

1. Introduction 5

2. Notation 5

3. Correspondence via path matching 6

4. Linear-time backward algorithm 6

5. Choice of neuron matching function m and operators ⊕ 7

6. Experiments 14

6.1 Experimental setup 14

6.2 Baselines 15

6.3 KITTI 2012 15

6.4 Ablation study on KITTI 2012 15

6.5 KITTI 2015 16

6.6 Style transfer experiment on KITTI 2015 16

7. Conclusion 29

**Matching neural paths: transfer from recognition to**

**correspondence search**

1. **Giới thiệu:**

Tìm sự tương ứng giữa các bộ phận giữa các đối tượng là một vấn đề tồn tại lâu dài trong học máy. Mức độ tương ứng được thiết lập có thể xuống thấp như pixel cho hình ảnh hoặc mili giây dấu thời gian cho tín hiệu âm thanh.Thông thường, rất mơ hồ khi khớp ở mức thấp như vậy: pixel hoặc dấu thời gian chỉ không chứa đủ thông tin để phân biệt đối xử và nhiều thông tin sai lệch sẽ làm theo. Một đại diện ngữ nghĩa phân cấp có thể giúp giải quyết sự mơ hồ: chúng ta có thể chọn trận đấu cấp thấp cũng phù hợp ở cấp độ cao hơn. Ví dụ, một chiếc xe hơi chứa một bánh xe có chứa một bu lông. Nếu chúng ta muốn kiểm tra xem bu-lông này có khớp với bu-lông trong một góc nhìn khác của xe hơi, chúng ta nên kiểm tra xem bánh xe và xe có phù hợp không.

Một đại diện ngữ nghĩa phân cấp có thể có thể được tính toán bằng một nơ ron tích chập mạng. Các tính năng trong một mạng như vậy được cấu thành theo cách phân cấp: cấp thấp hơn Các tính năng được sử dụng để tính toán các tính năng cao cấp của nhiều nếp cuộn áp dụng, poolings Max-và chức năng kích hoạt phi tuyến tính trên chúng. Tuy nhiên, đào tạo một mạng lưới thần kinh tích chập như vậy để dự đoán tương ứng trực tiếp có thể không phải là một lựa chọn trong một số lĩnh vực nơi mà các tương ứng chân thực là khó khăn và tốn kém để có được. Điều này đặt ra câu hỏi về khả năng mở rộng của các phương pháp như vậy và thúc đẩy tìm kiếm các phương pháp không cần đào tạo dữ liệu tương ứng.

Để giải quyết vấn đề dữ liệu đào tạo, chúng tôi có thể chuyển kiến ​​thức từ miền nguồn các nhãn được trình bày cho miền đích nơi không có nhãn hoặc một vài dữ liệu được dán nhãn. Nhất hình thức chuyển giao phổ biến là từ các nhiệm vụ phân loại. Lời hứa của nó là hai lần. Đầu tiên, phân loại nhãn là một trong những dễ dàng nhất để có được vì nó là một nhiệm vụ tự nhiên đối với con người. Điều này cho phép tạo ra rất lớn bộ dữ liệu nhận dạng nhưImagenet . Thứ hai, các tính năng từ cấp thấp đến trung cấp đã được hiển thị để chuyển tốt cho một loạt các nhiệm vụ.

Mặc dù đã có một tiến bộ lớn trong việc chuyển từ phân loại sang phát hiện phân đoạn và các nhiệm vụ lý luận ngữ nghĩa khác như dự đoán độ sâu hình ảnh đơn, việc chuyển sang tìm kiếm tương ứng đã bị giới hạn .

Một giải pháp chung cho việc chuyển không giám sát từ nhận dạng sang tìm kiếm tương ứng ở mức thấp nhất (pixel, dấu thời gian mili giây âm thanh). Cách tiếp cận là khớp các đường kích hoạt đến từ một mạng nơ ron tích chập, được áp dụng trên hai đối tượng để khớp. Chính xác hơn, để thiết lập kết hợp khớp ở mức thấp nhất, yêu cầu các tính năng phải khớp ở tất cả các cấp phân cấp tính năng tích chập khác nhau. Những tính năng cấp độ khác nhau tạo thành đường dẫn. Một đường dẫn như vậy sẽ bao gồm các kích hoạt thần kinh có thể truy cập từ tính năng cấp thấp nhất đến tính năng cấp cao nhất trong cấu trúc liên kết mạng (nói cách khác, tính năng cấp thấp nhất nằm trong trường tiếp nhận ở cấp cao nhất). Vì mọi tính năng cấp thấp nhất thuộc về nhiều đường dẫn, sẽ bỏ phiếu dựa trên tất cả các đường dẫn đó.

Mặc dù tổng số đường dẫn như vậy là theo cấp số nhân của số lớp và do đó không thể tính toán một cách dễ dàng, chứng minh rằng việc bỏ phiếu có thể xảy ra trong thời gian đa thức trong một lần lùi qua mạng. Thuật toán dựa trên lập trình động và tương tự như đường chuyền ngược để tính toán độ dốc trong mạng lưới thần kinh.

Xác thực theo kinh nghiệm được thực hiện trên nhiệm vụ tương ứng âm thanh nổi trên hai bộ dữ liệu: KITTI 2012 và KITTI 2015 . Chúng tôi định lượng cho thấy rằng phương pháp có khả năng cạnh tranh giữa các phương pháp không yêu cầu dữ liệu miền mục tiêu được gắn nhãn. Chúng tôi cũng cho thấy một cách định lượng rằng ngay cả những thay đổi mạnh mẽ trong cấu trúc cấp thấp có thể được xử lý hợp lý bằng phương pháp do tính mạnh mẽ của hệ thống phân cấp nhận dạng: chúng tôi áp dụng các kiểu chuyển khác nhau cho các hình ảnh tương ứng trong KITTI 2015 và vẫn tìm thấy các thư tương ứng.

1. **Kí hiệu:**

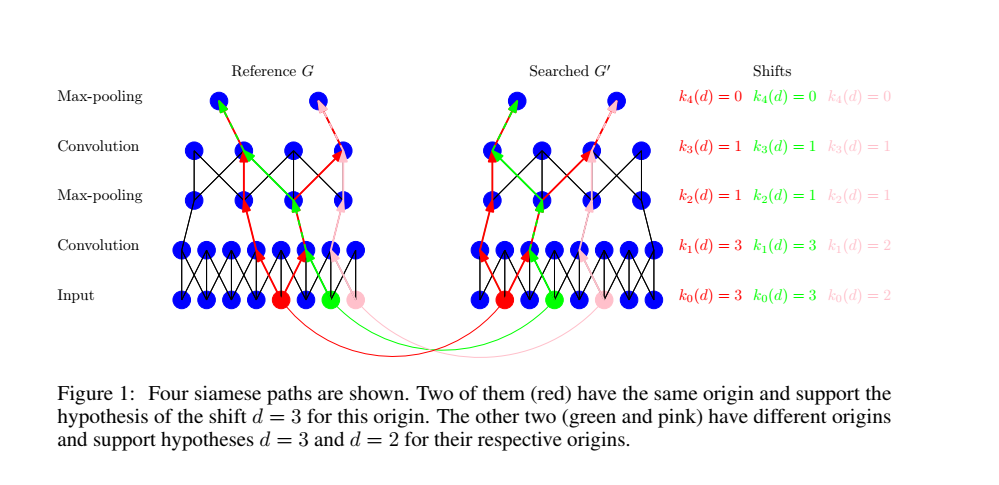
Pháp của thường áp dụng cho các trường hợp dữ liệu đầu vào có bố cục cấu trúc liên kết lưới đa chiều. Chúng ta sẽ giả sử các đối tượng đầu vào o là từ tập hợp các lưới B chiều R B và chạy các mạng thần kinh tích chập trên các lưới đó. Các kích hoạt trên mỗi lớp từ các mạng đó sẽ được chứa trong tập hợp các lưới hai chiều (B 1) Ψ ⊂ RB 1. Cả dữ liệu đầu vào và kích hoạt sẽ được lập chỉ mục bởi vectơ hai chiều (B 1) x = (x , y, .., c) ∈ NB 1, trong đó x là chỉ mục cột, y là chỉ mục hàng, v.v. và c ∈ {1 ,. . . , C} là chỉ mục kênh (giả sử C = 1 cho dữ liệu đầu vào, đây là giả định không hạn chế ).

Tìm kiếm sự tương ứng giữa các lưới đó, do đó mục tiêu của chúng tôi sẽ là ước tính các thay đổi d ∈ D ⊂ Z B 1 cho tất cả các yếu tố trong lưới. Sự lựa chọn của bộ dịch chuyển D phụ thuộc vào nhiệm vụ. Ví dụ: đối với âm thanh B = 1 và chỉ có thể xem xét các ca 1D. Đối với hình ảnh, B = 2 và D có thể là một tập hợp các ca 1D (thường được gọi là tác vụ âm thanh nổi) hoặc một tập hợp các ca 2D (thường được gọi là tác vụ dòng quang).

Trong công việc này, sẽ xử lý các kiến trúc mạng nơ ron tích chập, bao gồm các cấu trúc, nhóm tối đa và các hàm kích hoạt phi tuyến tính (một ví dụ về kiến trúc như vậy là một mạng VGG , nếu chúng ta bỏ qua softmax mà chúng ta sẽ không sử dụng cho việc chuyển nhượng). Chúng tôi giả định rằng mọi lớp chập được theo sau bởi một hàm kích hoạt phi tuyến tính trong bài báo và sẽ không chỉ định rõ ràng các chức năng đó.

Biểu đồ tính toán của các kiến trúc này là một biểu đồ chu kỳ có hướng G = {A, E}, trong đó A = {a1 ,. . . , a | A |} là một tập hợp các nút, tương ứng với kích hoạt nơ-ron (| A | biểu thị kích thước của tập hợp này) và E = {e1 ,. . . , e | E |} là một tập hợp các cung, tương ứng với các phụ thuộc tính toán (| E | biểu thị kích thước của tập hợp này). Mỗi cung được biểu diễn dưới dạng một tuple (ai, aj), trong đó ai là đầu vào (gốc), aj là đầu ra (điểm cuối). Tập hợp nút bao gồm các lớp tách rời A = SL `= 0 A`. Các cung chỉ được phép đi từ lớp trước sang lớp kế tiếp.

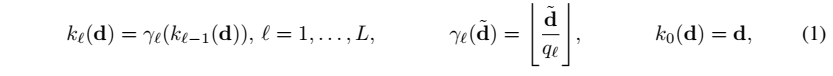
Chúng ta sẽ sử dụng ký hiệu A` (x) cho nút trong lớp `-th ở vị trí x; trong (x`) cho tập hợp nguồn gốc x` 1 của cung, nhập lớp `tại vị trí x` của đối tượng tham chiếu; x` 1 ∈ out (x`) cho tập hợp các điểm cuối của cung, thoát khỏi lớp `tại vị trí x` của đối tượng tham chiếu. Đặt f` F = {maxpool, conv} là toán tử toán học tương ứng với tính toán chuyển tiếp trong lớp `là một ← f` (trong (a)), a ∈ A` (với một sự lạm dụng ký hiệu nhẹ, chúng ta sử dụng một cho cả các nút trong biểu đồ tính toán và các giá trị kích hoạt được tính toán trong các nút đó).



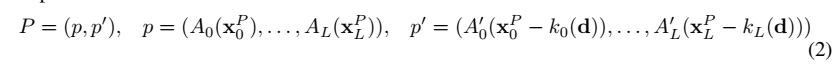
1. **Sự tương ứng thông qua khớp đường dẫn:**

Xem xét hai đối tượng, tham chiếu o và tìm kiếm o 0 ∈, mà muốn tìm các tương ứng. Sau khi áp dụng CNN trên chúng, chúng ta nhận được biểu đồ G và G0 kích hoạt. Mục tiêu là thiết lập sự tương ứng giữa các lớp dữ liệu đầu vào A0 và A0 0. Nghĩa là, mọi ô A0 (x) trong đối tượng tham chiếu o ∈ có một sự thay đổi nhất định d ∈ D trong đối tượng tìm kiếm o 0 và muốn ước tính d.

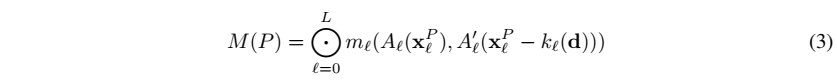
Ở đây có ý tưởng nền tảng của phương pháp của chúng tôi: chúng tôi thiết lập khớp A0 (x) với A0 0 (x - d) cho một ca d nếu có một cặp đường dẫn song song ,bắt nguồn từ các nút đó và kết thúc ở các lớp cuối AL, A0 L, khớp nhau. Cặp đường dẫn này phải có cùng một sự thay đổi không gian đối với nhau ở tất cả các lớp, cho đến việc lấy mẫu con và đi qua các kênh tính năng tương tự đối với nhau. Tính đến việc lấy mẫu con bằng các hàm trên mỗi lớp.



Trong đó k` (d) là cách dịch chuyển của lớp 0 d biến đổi ở lớp `, q` là hệ số phụ mẫu không gian lớp thứ` (lưu ý rằng làm tròn và chia trên các vectơ được thực hiện theo phần tử). Sau đó, một đường dẫn P có thể được biểu diễn dưới dạng.

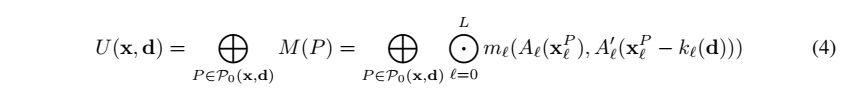


Trong đó x P 0 = x và x P `biểu thị vị trí tại đó đường dẫn P cắt lớp` của đồ thị kích hoạt tham chiếu. Các đường dẫn như vậy được minh họa trong Hình 1. Logic rất đơn giản: khớp trong đường dẫn có nghĩa là hệ thống phân cấp nhận dạng phát hiện các tính năng giống nhau ở các mức độ nhận thức khác nhau với cùng một dịch chuyển (lên đến mẫu phụ) đối với vị trí ước tính hiện tại x, cho phép dự đoán tự tin về trận đấu. Có thể thiết lập một đường dẫn siamese là phù hợp với Trực tuyến bằng cách tính toán chức năng khớp (cao nếu nó khớp, thấp nếu không).



Vì muốn ước tính sự thay đổi cho một nút A0 (x), chúng ta sẽ xem xét tất cả các ca làm việc có thể và bỏ phiếu cho mỗi nút. Chúng ta hãy biểu thị một tập hợp các đường dẫn siamese, bắt đầu từ A` (x) và A0 `(x - d) và kết thúc ở lớp cuối cùng, là P` (x, d).

Với mỗi ca d ∈ D, giới thiệu U (x, d) là khả năng ghi nhật ký của sự kiện mà d là ca đúng, tức là A0 (x) khớp với A0 0 (x - d). Để thu thập bằng chứng từ tất cả các đường dẫn có thể, chúng ta Tổng hợp các chức năng phù hợp cho tất cả các đường dẫn riêng lẻ, dẫn đến.



trong đó toán tử tổng hợp sẽ được thảo luận sau.

Phân phối U (x, d) có thể được sử dụng để thu được giải pháp là d ∗ (x) = arg maxd∈DU (x, d) hoặc để xử lý sau phân phối với bất kỳ loại tối ưu hóa làm mịn không gian nào và sau đó thực hiện lại giải pháp chi phí tốt nhất.

1. **Thuật toán ngược thời gian tuyến tính:**
2. **Lựa chọn chức năng khớp nơ ron m và toán tử**
3. **Thí nghiệm**
4. **Kết luận**

Trong công việc này, đã trình bày một phương pháp để chuyển từ nhận dạng sang tìm kiếm tương ứng ở mức thấp nhất. Vì vậy, chúng tôi sử dụng lại các đường dẫn kích hoạt từ các mạng thần kinh tích chập sâu và đề xuất một thuật toán đa thức hiệu quả để tổng hợp số lượng các đường dẫn như vậy. Các kết quả thực nghiệm về tác vụ khớp âm thanh nổi cho thấy phương pháp của có khả năng cạnh tranh giữa các phương thức không sử dụng dữ liệu được gắn nhãn từ miền đích. Sẽ rất thú vị khi áp dụng kỹ thuật này vào âm thanh, điều này sẽ trở nên khả thi khi một mô hình tích chập sâu chất lượng cao có thể truy cập được cho công chúng .