SKETCHZOO: TRUY VẤN HÌNH ẢNH ĐỘNG VẬT TỪ NÉT VẼ TAY SỬ DỤNG SIAMESE NETWORK

Nguyễn Tuấn Quang - 230101017

Tóm tắt

- Lớp: CS2205.CH190
- Link Github của nhóm: https://github.com/TuanQuang5720/CS2205
- Link YouTube video: https://www.youtube.com/watch?v=TOcyUHNI4Us
- Họ và tên: Nguyễn Tuấn Quang - 230101017

Giới thiệu

- Ngày nay, người dùng có xu hướng tìm kiếm nội dung trực quan bằng các cách tự nhiên hơn như vẽ phác thảo thay vì gõ từ khóa. Tuy nhiên, việc truy xuất hình ảnh từ bản vẽ tay vẫn còn là thách thức do sự khác biệt lớn giữa sketch và ảnh thật.
- **SketchZoo** ra đời nhằm giải quyết vấn đề này. Hệ thống cho phép người dùng **tìm kiếm ảnh thật chỉ từ một bản vẽ tay đơn giản**, ứng dụng mô hình **Siamese Network** kết hợp giữa **contrastive learning** và **triplet learning** để học đặc trưng và đo lường độ tương đồng giữa ảnh và sketch.
- Không như các mô hình generative đòi hỏi tài nguyên cao, SketchZoo được thiết kế hướng đến **độ chính xác cao**, hiệu quả, và có thể triển khai trên các thiết bị tài nguyên thấp như mobile device.
- **Input:** Ảnh sketch do người dùng vẽ tay (đen trắng, nét đơn giản)
- **Output:** Top-N ảnh thật trong cơ sở dữ liệu có nội dung gần giống nhất với sketch (theo lớp/đối tượng)

Mục tiêu

Nghiên cứu và xây dựng mô hình học sâu truy xuất ảnh từ phác thảo tay, sử dụng kiến trúc mạng Siamese kết hợp hai chiến lược học: Triplet Loss và Contrastive Loss.

Thực nghiệm và so sánh hiệu quả giữa các phương pháp học và tiền xử lý dữ liệu thông qua các độ đo: Accuracy, Precision@k, Recall@k, Mean Average Precision (mAP):

- So sánh hiệu quả giữa hai hàm mất mát: Triplet Loss và Contrastive Loss
- Đánh giá tác động của kỹ thuật data augmentation cho cả ảnh thật và ảnh phác thảo
- Thực nghiệm với nhiều kiến trúc backbone: ResNet18, ResNet32, ResNet50, ResNet101
- Kiểm nghiệm mô hình trên tập dữ liệu Sketchy và bộ dữ liệu thực tế lưu trữ trong MongoDB

Phát triển ứng dụng SketchZoo, cho phép người dùng nhập phác thảo tay đơn giản và truy xuất các hình ảnh thật tương đồng trong cơ sở dữ liệu.

Giải pháp góp phần hỗ trợ các hệ thống sáng tạo nội dung hình ảnh, công cụ giáo dục trực quan và tăng cường tương tác người - máy.

Nội dung và Phương pháp

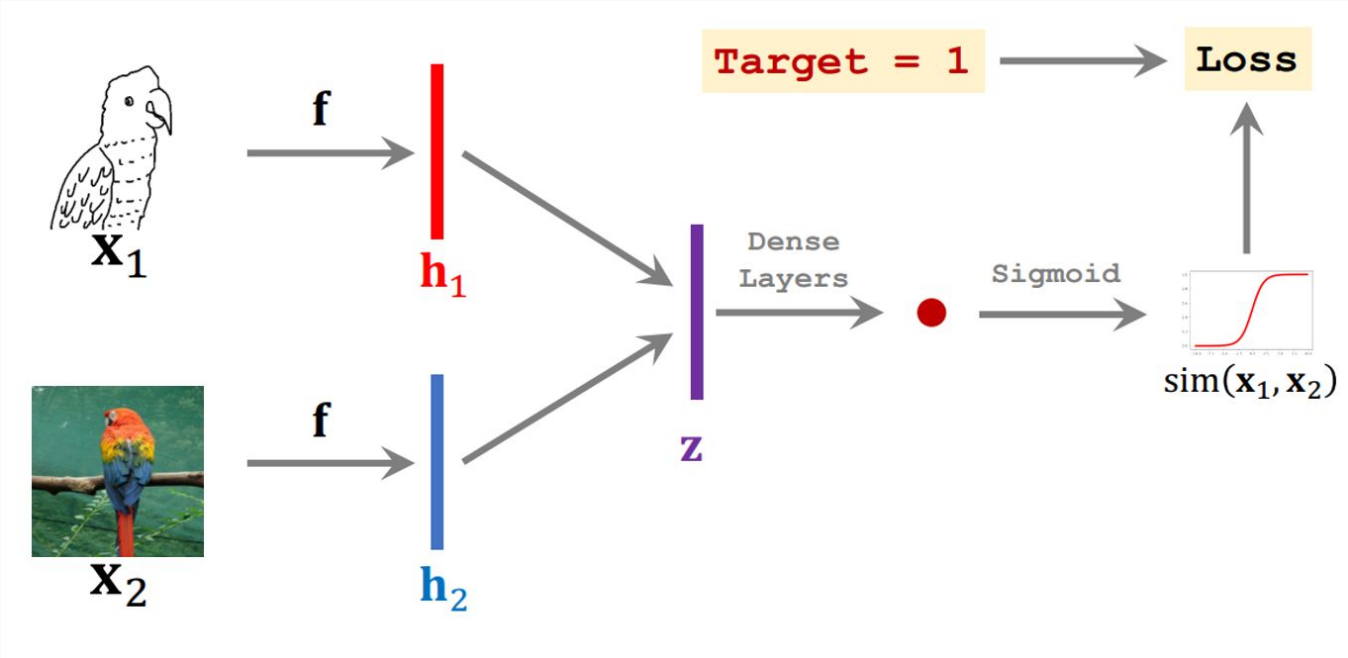
- Nội dung
  - Nghiên cứu và phát triển hệ thống truy xuất ảnh từ phác thảo tay bằng mô hình học sâu, sử dụng kiến trúc mạng Siamese với hai cơ chế học: Contrastive Loss và Triplet Loss.
  - Tìm hiểu nguyên lý hoạt động và hiệu quả học biểu diễn qua không gian embedding của các cặp (sketch, photo).
  - Xây dựng và hiệu chỉnh bộ dữ liệu từ Sketchy Dataset và tập ảnh thực tế lưu trữ trên MongoDB, đảm bảo dữ liệu đa dạng lớp và cân bằng mẫu giữa sketch và photo nhằm tránh overfitting.
  - Tiền xử lý ảnh và phác thảo bằng các kỹ thuật augmentation như xoay, lật, jitter, affine transform,... để tăng độ tổng quát cho mô hình.
  - Triển khai mô hình huấn luyện dựa trên nhiều backbone CNN: ResNet18, ResNet32, ResNet50, ResNet101 để đánh giá ảnh hưởng kiến trúc đến chất lượng biểu diễn.

Nội dung và Phương pháp

- Phương pháp
  - Thực nghiệm song song hai chiến lược học:
    - Contrastive Loss (mục tiêu kéo gần vector embedding của các cặp giống nhau và đẩy xa các cặp khác nhau)
    - Triplet Loss với chiến lược chọn mẫu khác nhau: random, semi-hard, hard negative
  - Huấn luyện mô hình Siamese sử dụng ảnh thật và ảnh phác thảo để học ra không gian biểu diễn chung, phục vụ cho việc tính khoảng cách và truy xuất ảnh. Đánh giá hiệu quả mô hình qua các độ đo:
    - Precision@k, Recall@k, Accuracy, mAP (mean Average Precision)
    - Tốc độ suy luận (inference time) để đánh giá khả năng triển khai thực tế
  - Xây dựng ứng dụng SketchZoo, cho phép người dùng nhập một ảnh phác thảo đầu vào, hệ thống sẽ truy xuất và hiển thị top-n ảnh thật giống nhất từ cơ sở dữ liệu.

Nội dung và Phương pháp

- Contrastive learning



Kết quả dự kiến

- So sánh hiệu quả giữa các mô hình thông qua các độ đo:
  - Precision@k, Recall@k, mAP: đánh giá khả năng truy xuất chính xác
  - Embedding Distance (L2/Cosine): đánh giá độ gần về mặt biểu diễn
  - Inference Time: đo tốc độ suy luận phục vụ triển khai thực tế
- Hệ thống có khả năng truy xuất chính xác top-n ảnh thật giống nhất với sketch đầu vào, hoạt động tốt với dữ liệu chưa từng thấy.
- Ứng dụng hỗ trợ người dùng nhập ảnh phác thảo và nhận về kết quả tương ứng, phù hợp cho các bài toán tìm kiếm bằng sketch và hỗ trợ sáng tạo hình ảnh.

Tài liệu tham khảo

[1]. Ong, Eng-Jon, Sameed Husain, and Miroslaw Bober. "Siamese network of deep fisher-vector descriptors for image retrieval." *arXiv preprint arXiv:1702.00338* (2017).

[2]. Melekhov, Iaroslav, Juho Kannala, and Esa Rahtu. "Siamese network features for image matching." *2016 23rd international conference on pattern recognition (ICPR)*. IEEE, 2016.

[3]. Qi, Yonggang, et al. "Sketch-based image retrieval via siamese convolutional neural network." *2016 IEEE international conference on image processing (ICIP)*. IEEE, 2016.

[4]. Sun, Qianru, et al. "Meta-transfer learning for few-shot learning." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.

[5]. Golfe, Alejandro, et al. "Enhancing image retrieval performance with generative models in siamese networks." *IEEE Journal of Biomedical and Health Informatics* (2025).