

SKETCH-BASED IMAGE RETRIEVAL

Group:

21522057 - Hồ Thị Khánh Hiền

21522028 - Nguyễn Như Hà

Instructor: PhD. Mai Tiến Dũng

GROUP MEMBERS



21522028

Nguyễn Như Hà



21522057

Hồ Thị Khánh Hiền

TABLE OF CONTENT

I. Tổng quan

1. Giới thiệu
2. Lý do
3. Phát biểu bài toán

II. Phương pháp

1. Resnet50
2. VGG16
3. Triplet Margin Loss
4. Nearest Neighbors

III. Dataset

IV. Thực nghiệm

V. Đánh giá

1. Độ đo
2. Kết quả

VI. Demo

SKETCH- BASED IMAGE RETRIEVAL

I. TỔNG QUAN

1. Giới thiệu

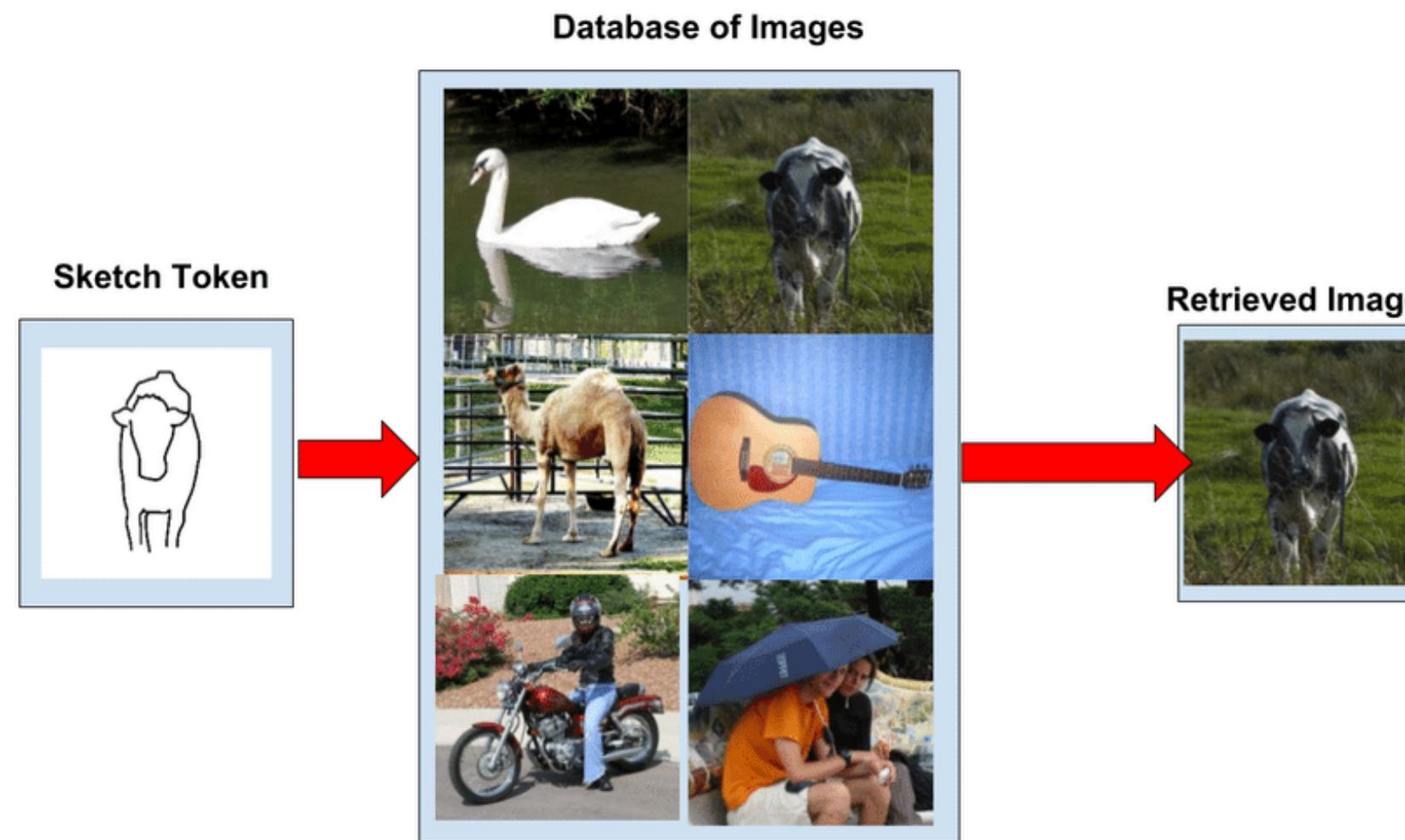
2. Lý do

3. Phát biểu bài toán

I. TỔNG QUAN:

1. Giới thiệu:

- Là bài toán tìm kiếm các hình ảnh từ một cơ sở dữ liệu dựa trên một hình chứa nét vẽ (sketch) do người dùng vẽ.



I. TỔNG QUAN:

2. Lý do chọn đề tài:

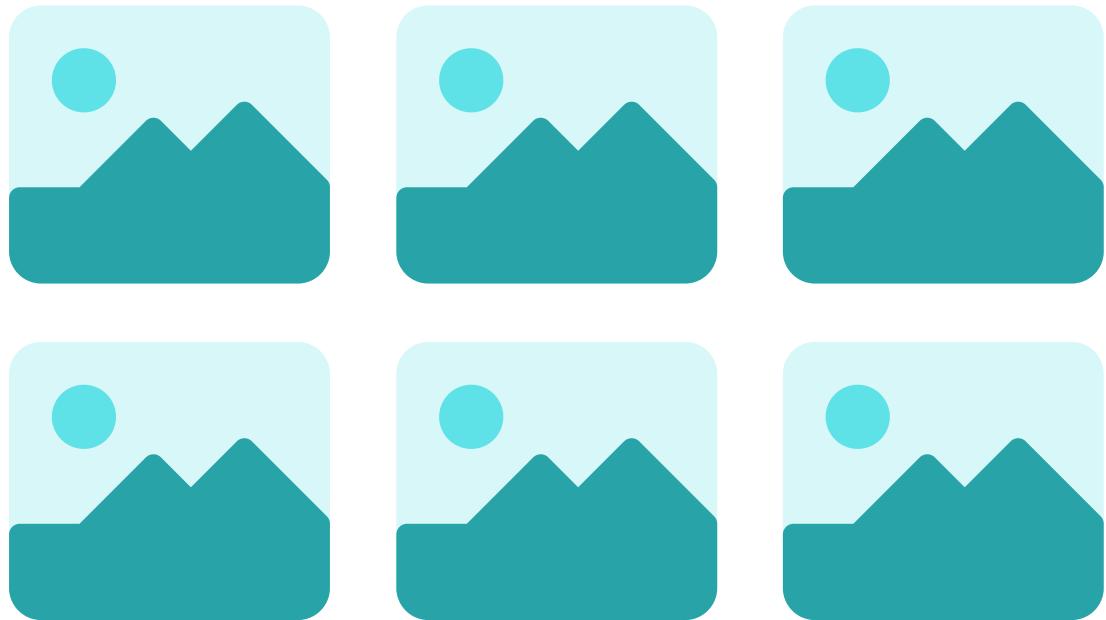
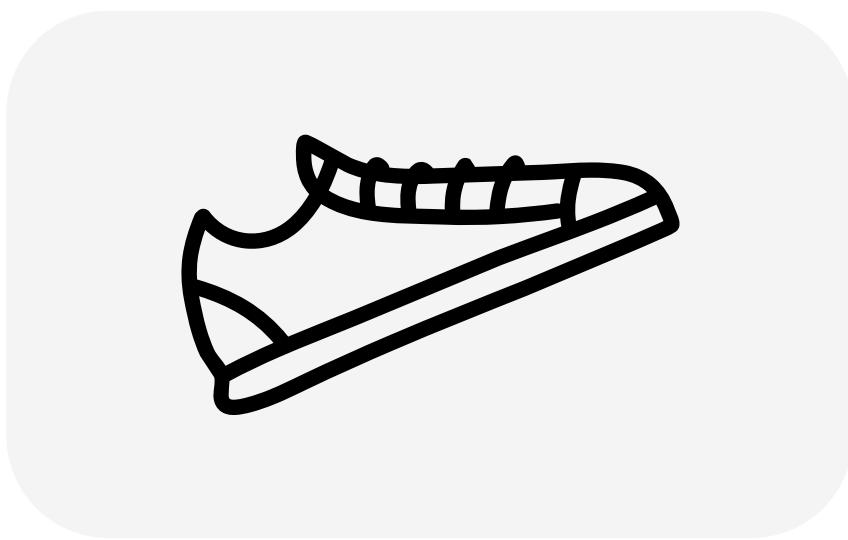
Tạo ra một công cụ giúp tìm kiếm các hình ảnh phù hợp bằng các bản vẽ. Điều này đặc biệt hữu ích:

- Tạo ra sự thuận tiện
- Khắc phục hạn chế ngôn ngữ



I. TỔNG QUAN:

3. Phát hiểu bài toán:



• Input:

- 1 bản vẽ được người dùng cung cấp bằng cách tải lên hoặc vẽ bằng công cụ kỹ thuật số.
- Bộ dữ liệu bao gồm các hình ảnh thực và ảnh phác thảo tương ứng với hình ảnh đó.

• Output:

- Tập hợp các hình ảnh từ bộ dữ liệu mà được đánh giá là phù hợp nhất với ảnh phác thảo truy vấn.

SKETCH- BASED IMAGE RETRIEVAL

II. PHƯƠNG PHÁP

1. Resnet50

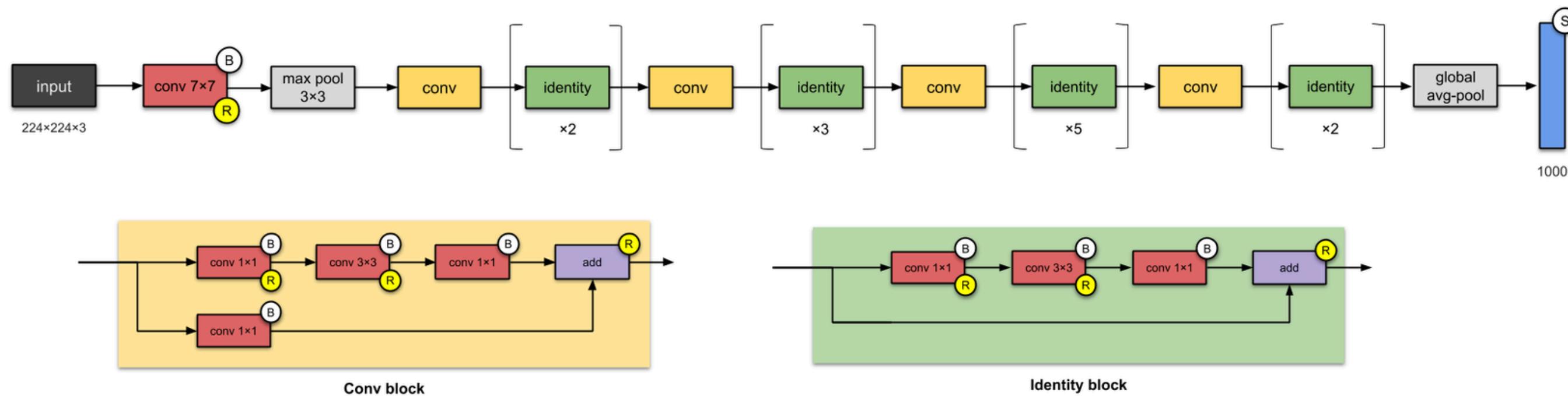
2. VGG16

3. Triplet Margin Loss

4. Nearest Neighbors

II. PHƯƠNG PHÁP:

1. ResNet50



- Được phát triển bởi các nhà nghiên cứu tại Microsoft Research Asia, năm 2015.
- ResNet-50 bao gồm 50 lớp (48 convolutional layers, 1 MaxPool layer, và 1 average pool layer).
- Kiến trúc của mạng ResNet50 bao gồm 2 khối đặc trưng là khối tích chập (Conv Block) và khối xác định (Identity Block).

SKETCH- BASED IMAGE RETRIEVAL

II. PHƯƠNG PHÁP

1. Resnet50

2. VGG16

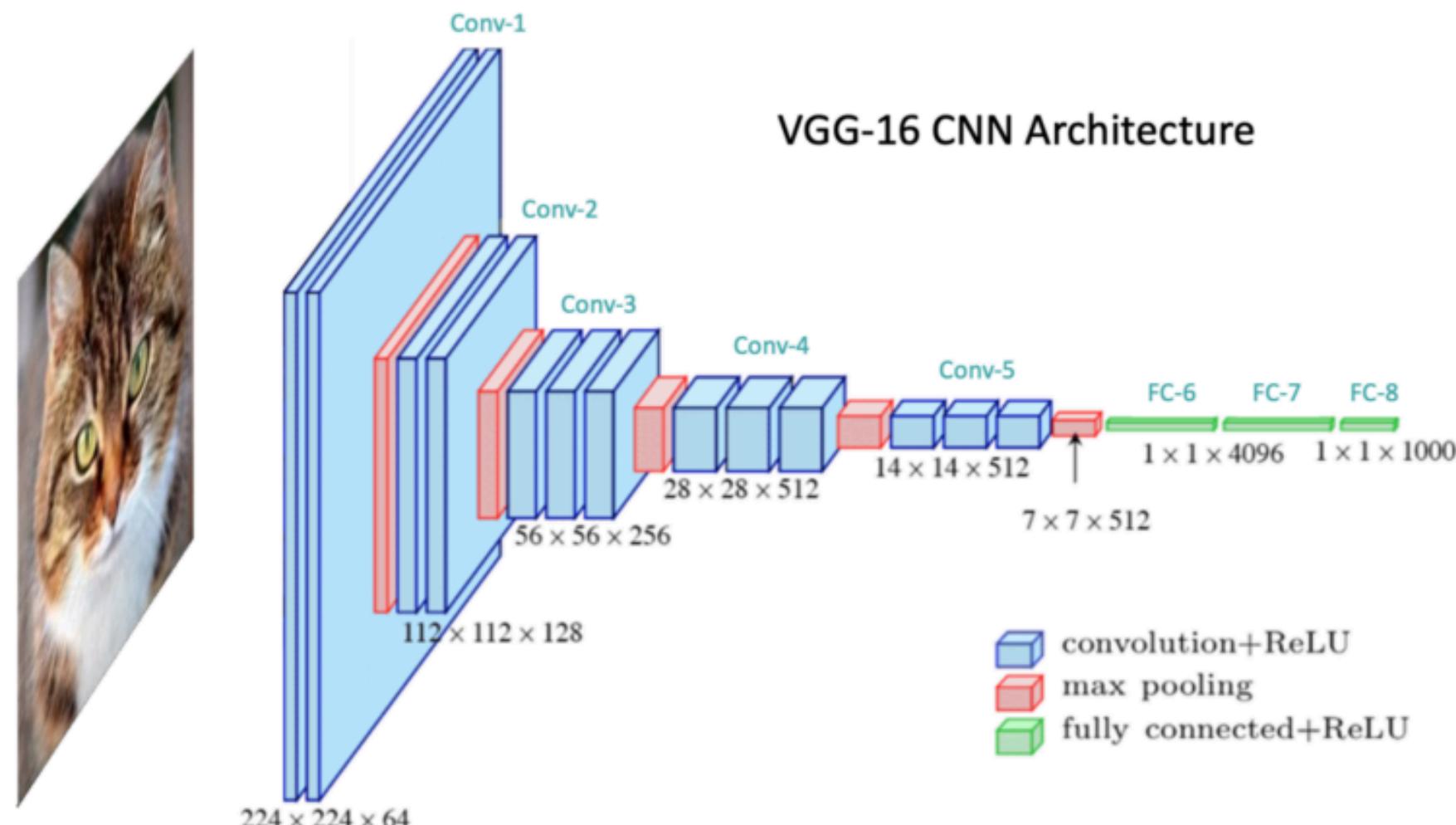
3. Triplet Margin Loss

4. Nearest Neighbors

II. PHƯƠNG PHÁP:

2. VGG16

- Là mạng CNN được đề xuất bởi K. Simonyan và A. Zisserman, University of Oxford.
- Gồm 16 lớp (13 lớp convolutional và 3 lớp fully connected).



SKETCH- BASED IMAGE RETRIEVAL

II. PHƯƠNG PHÁP

1. Resnet50

2. VGG16

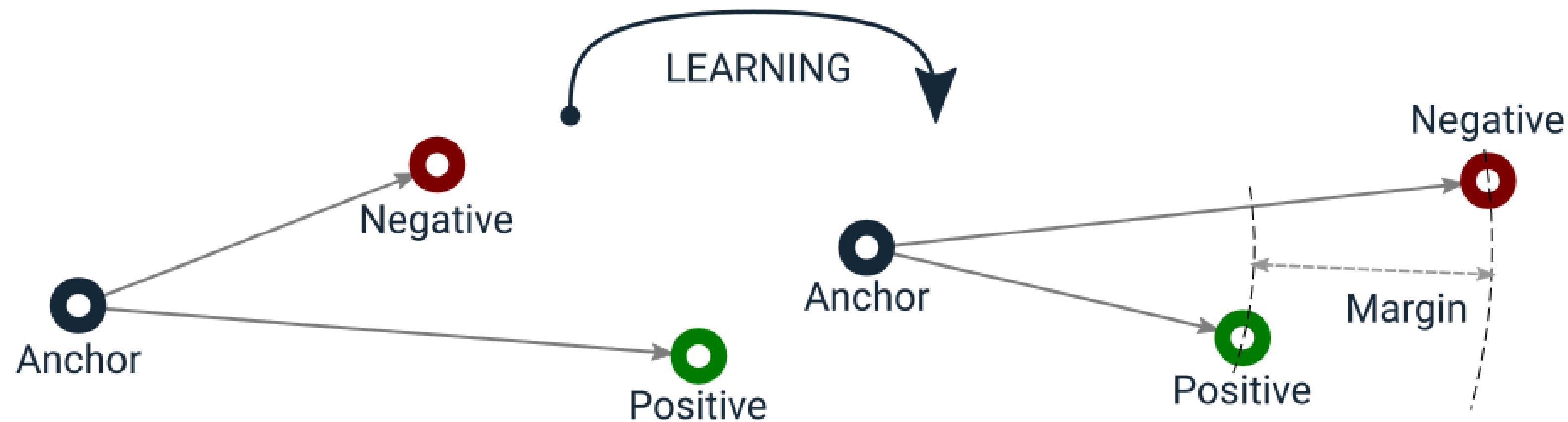
3. Triplet Margin Loss

4. Nearest Neighbors

II. PHƯƠNG PHÁP:

3. Triplet Margin Loss

- Được giới thiệu lần đầu tiên trong **FaceNet: A Unified Embedding for Face Recognition and Clustering** năm 2015.
- Một trong những hàm mất mát phổ biến nhất cho việc học có giám sát.



SKETCH- BASED IMAGE RETRIEVAL

II. PHƯƠNG PHÁP

1. Resnet50

2. VGG16

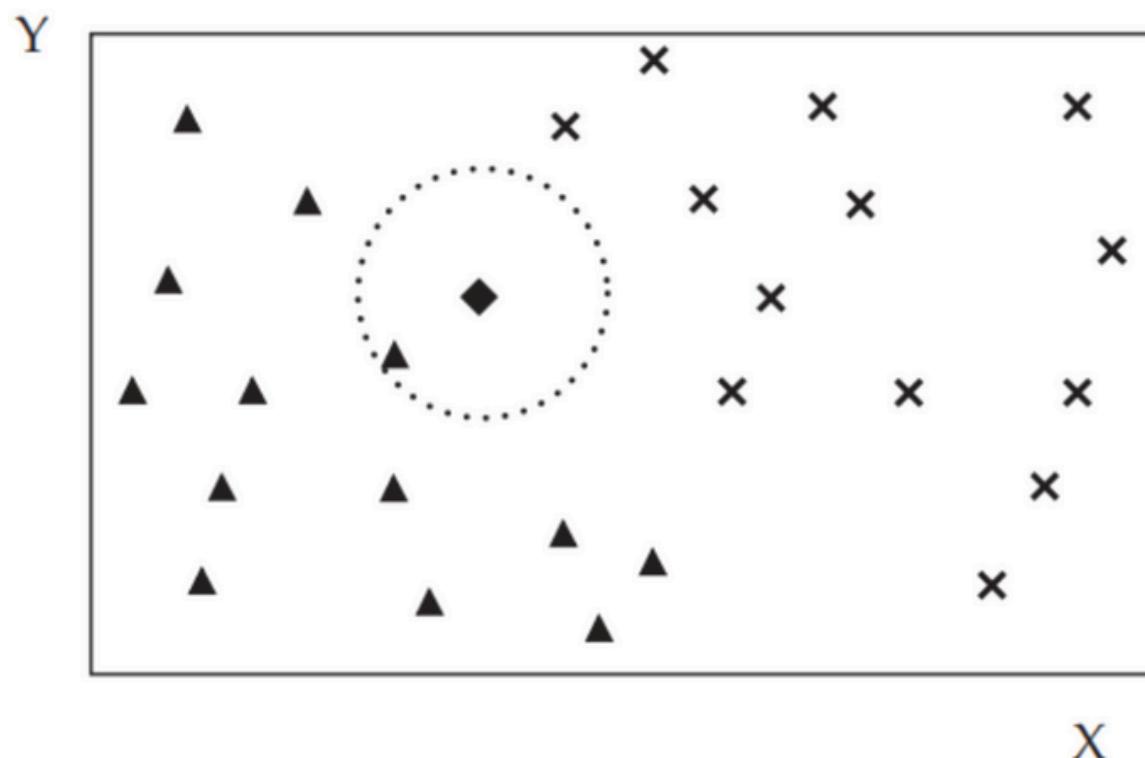
3. Triplet Margin Loss

4. Nearest Neighbors

II. PHƯƠNG PHÁP:

4. Nearest Neighbors

- Mô tả lần đầu tiên vào đầu những năm 1950.
- Thuật toán vẽ tất cả các vectơ trong không gian đa chiều và tìm các điểm lân cận gần nhất với điểm query.



- Nhóm đã dùng thuật toán vét cạn (brute force) và tính khoảng cách euclidean giữa các điểm.

SKETCH- BASED IMAGE RETRIEVAL

III. DATASET

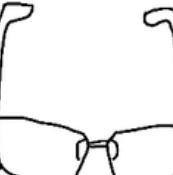
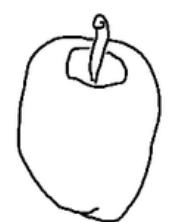
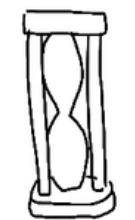
III. DATASET:

The Sketchy Database

- Vẽ phác thảo dựa trên 125 lớp gồm các ảnh vật thể khác nhau.
- 75,471 phác thảo của 12,500 đối tượng.
- Train:Test = 8:2



III. DATASET:



SKETCH- BASED IMAGE RETRIEVAL

IV. THỰC NGHIỆM

1. Preprocessing

2. Tạo triplets

3. Training

IV. THỰC NGHIỆM:

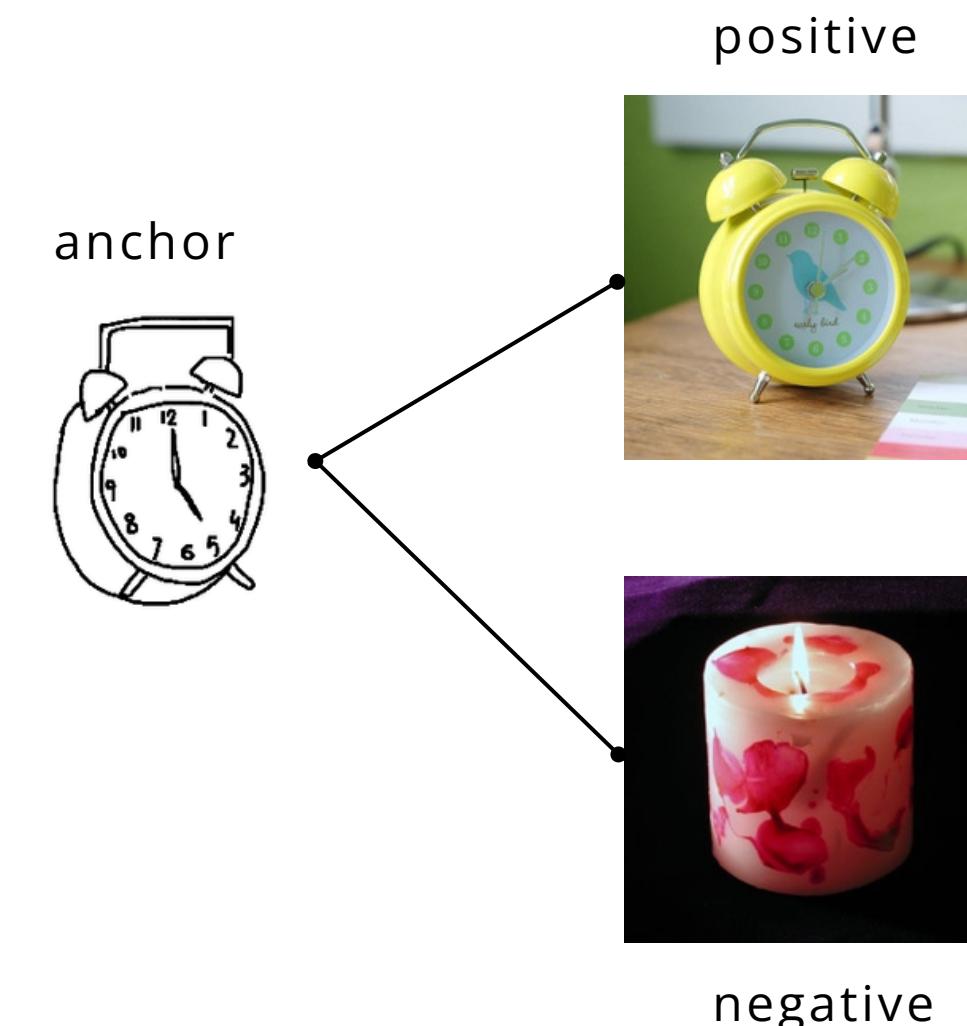
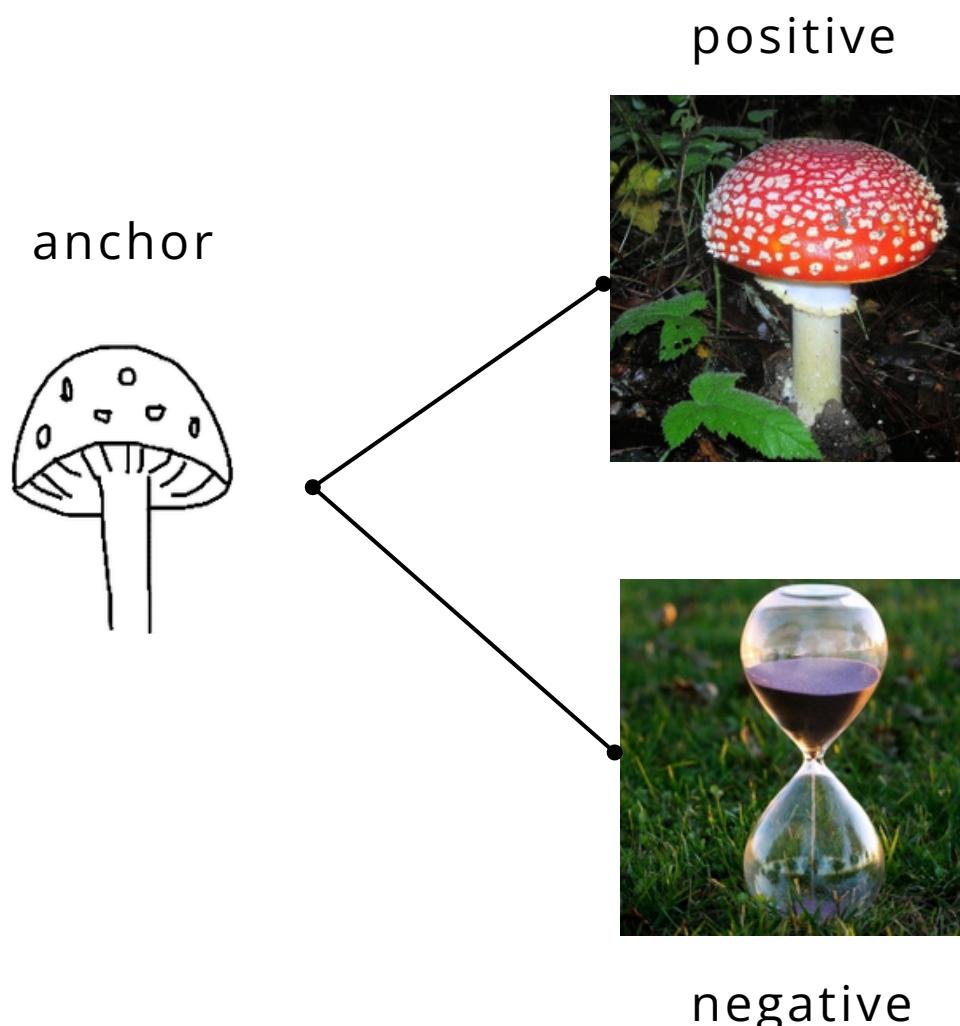
1. Pre-processing:

```
self.transform = transforms.Compose([
    transforms.Resize(224),
    transforms.CenterCrop(224),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
                         std=[0.229, 0.224, 0.225])
])
```

IV. THỰC NGHIỆM:

2. Tạo triplets:

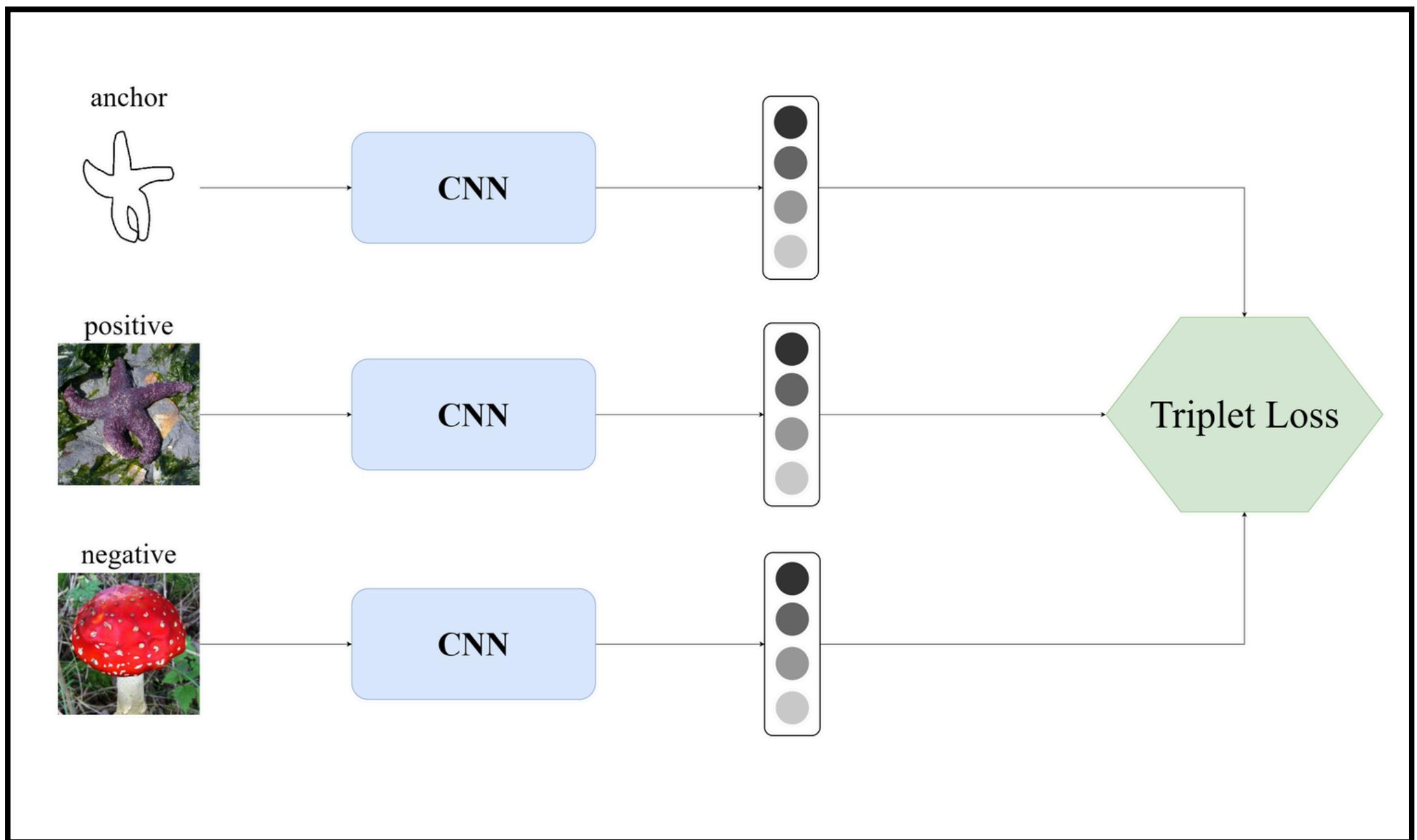
- Triplets bao gồm 1 ảnh phác thảo (anchor), 1 ảnh thực tương ứng với ảnh phác thảo (positive) và 1 ảnh thực thuộc 1 class khác so với ảnh phác thảo (negative).



IV. THỰC NGHIỆM:

3. Training:

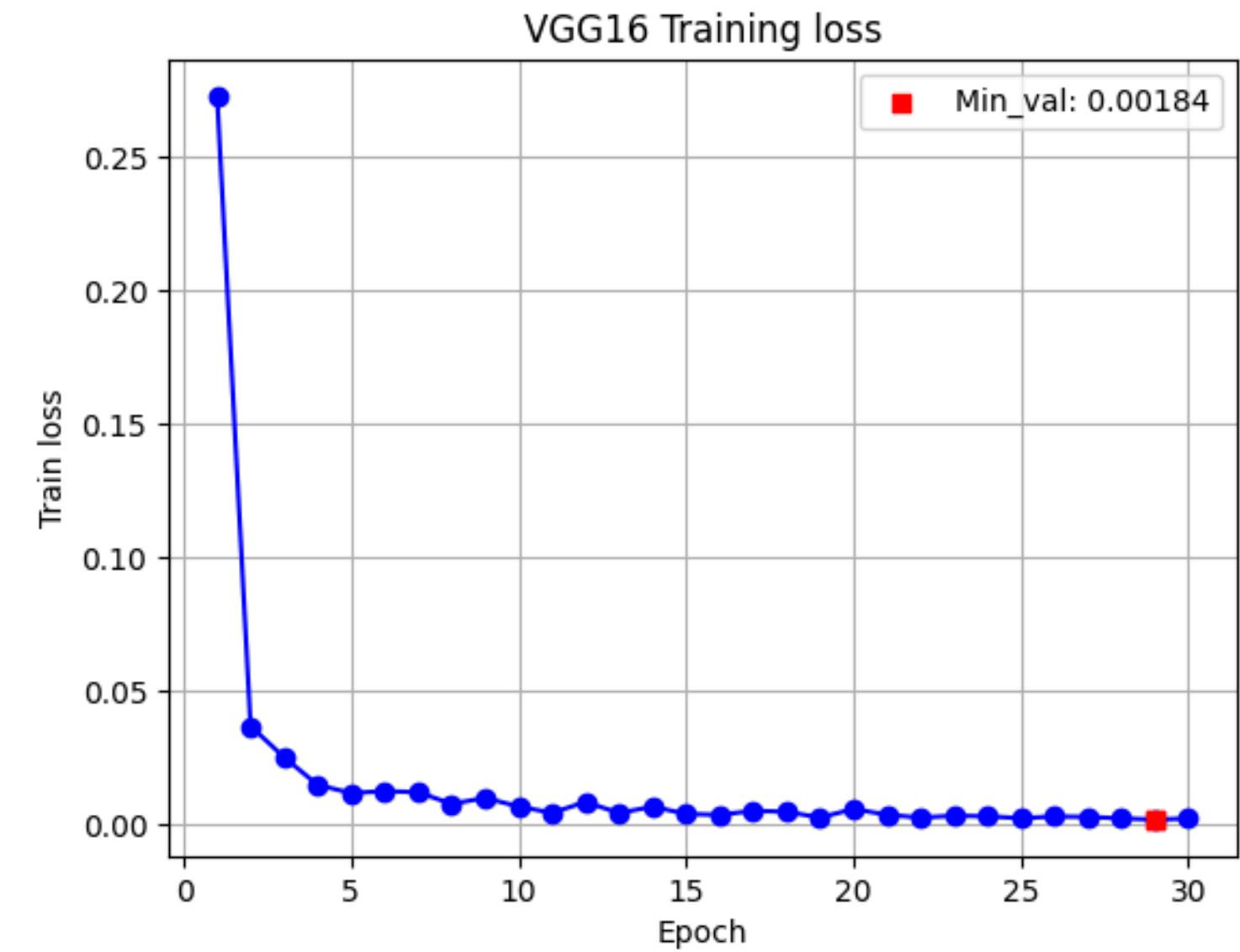
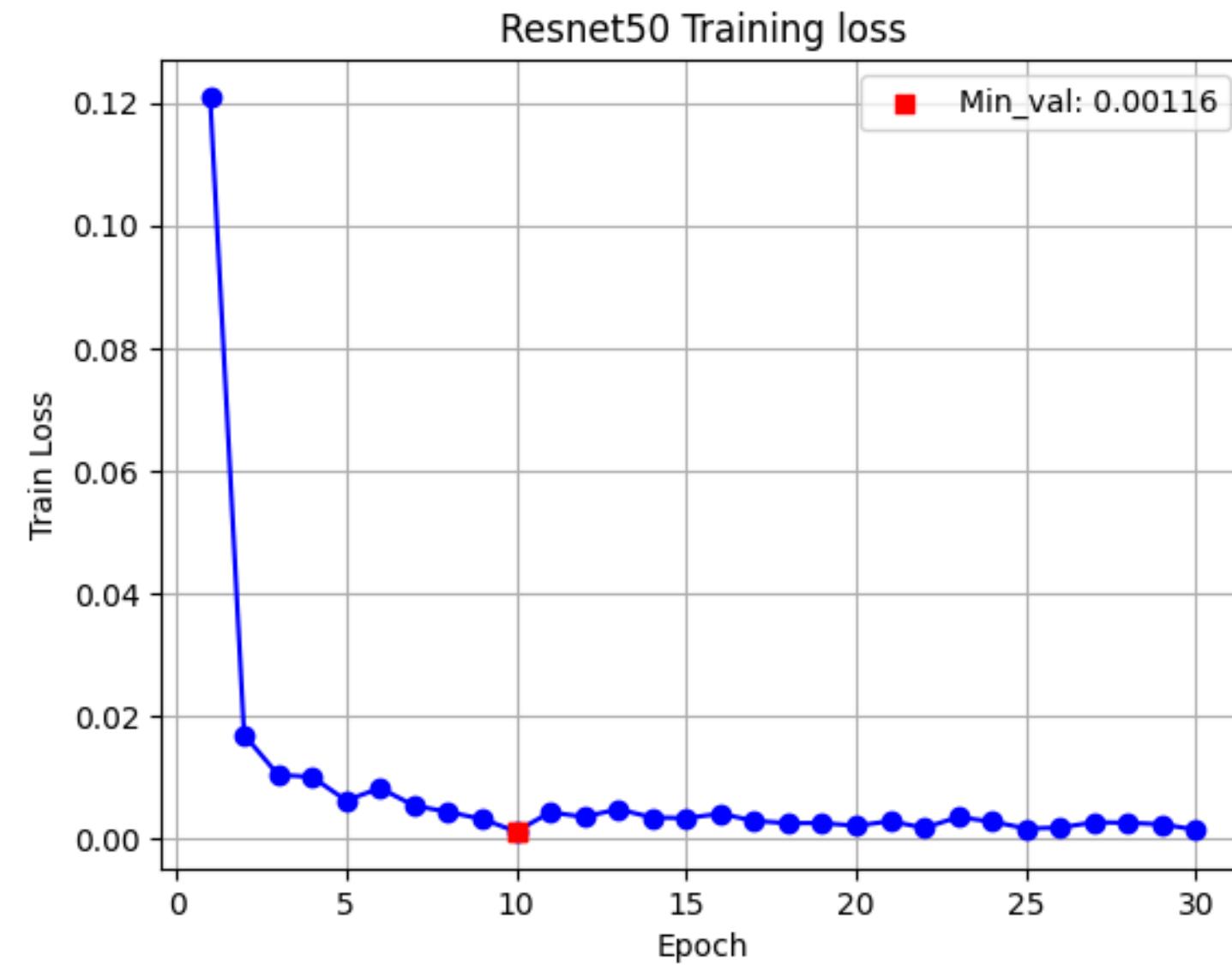
- *epoch* = 30
- *learning_rate* = 1e-3
- *batch_size* = 32
- *Triplet margin loss*
($m=1$, $p=2$)



IV. THỰC NGHIỆM:

3. Training:

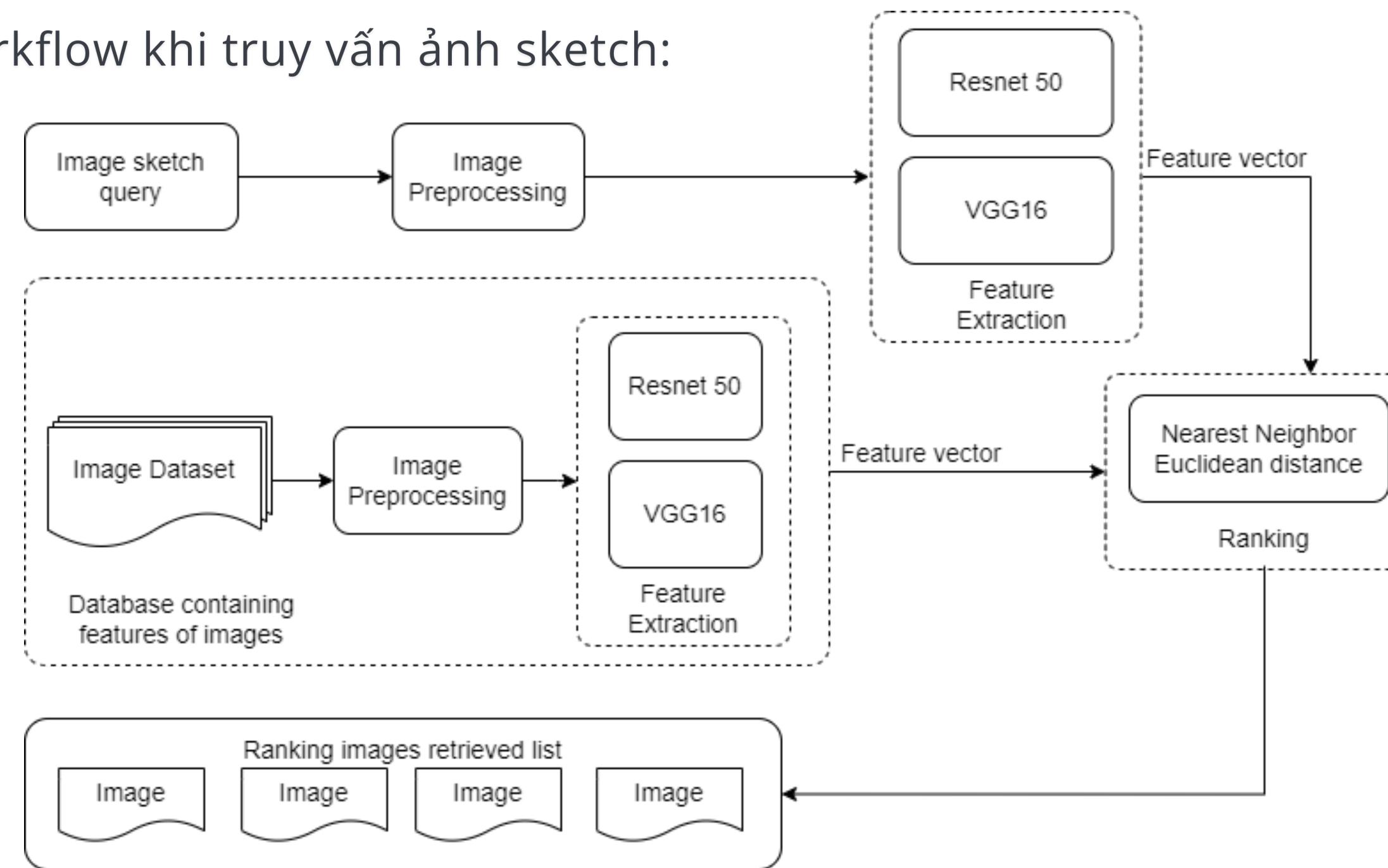
- Loss trong quá trình train:



IV. THỰC NGHIỆM:

3. Training:

- Workflow khi truy vấn ảnh sketch:



SKETCH- BASED IMAGE RETRIEVAL

V. ĐÁNH GIÁ

1. Độ đo

2. Kết quả

V. ĐÁNH GIÁ:

1. Độ đo:

- Precision

$$\text{Precision} = \frac{\text{NumberOfRelevantItemsRetrieved}}{\text{TotalNumberOfItemsRetrieved}}$$

- Average Precision (AP)

$$\text{AveP} = \frac{\sum_{k=1}^n P(k) * \text{rel}(k)}{\text{TotalNumberOfRelevantItems}}$$

- mean Average Precision (MAP)

$$\text{mAP} = \frac{\sum_{q=1}^Q \text{AveP}(q)}{Q}$$

SKETCH- BASED IMAGE RETRIEVAL

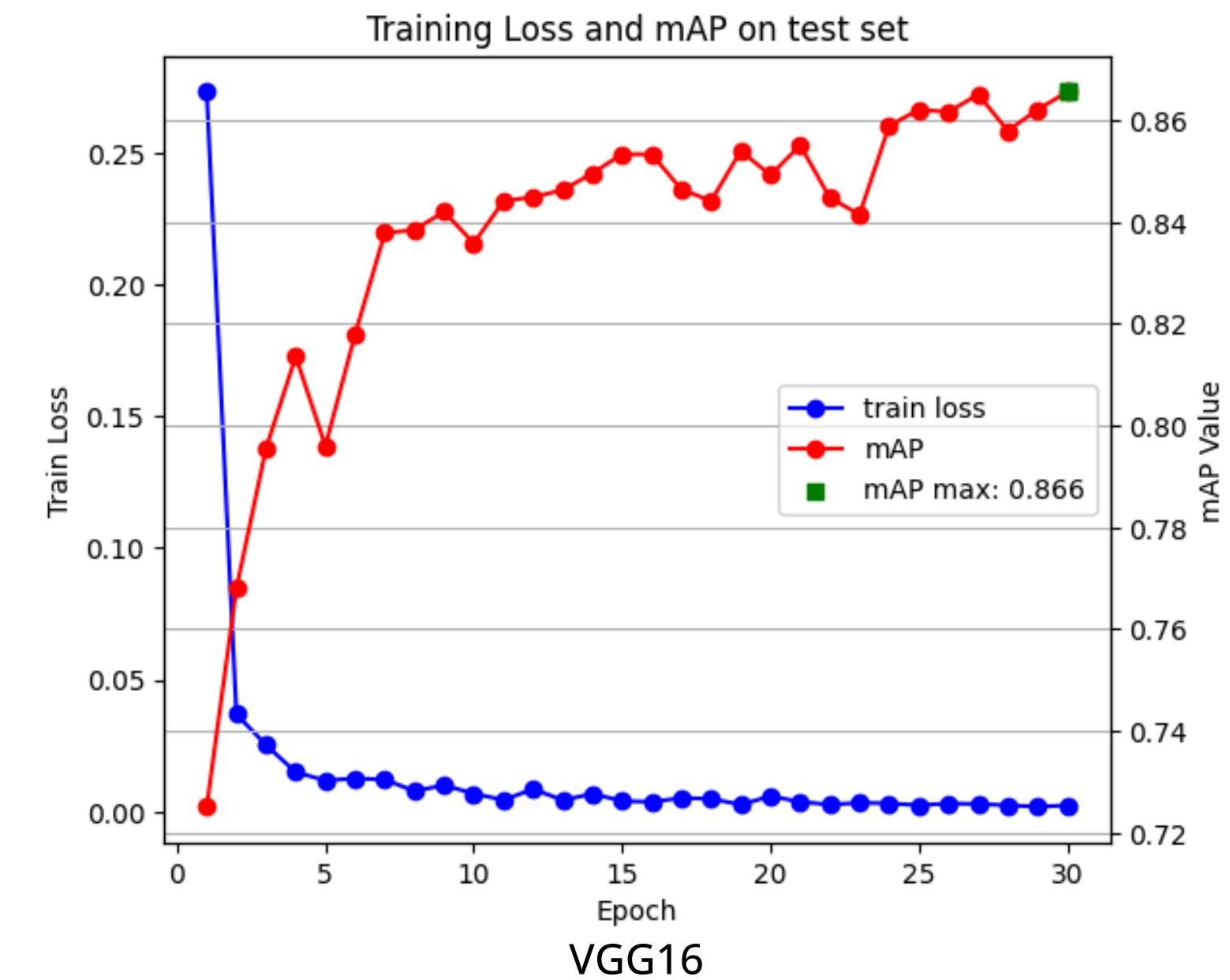
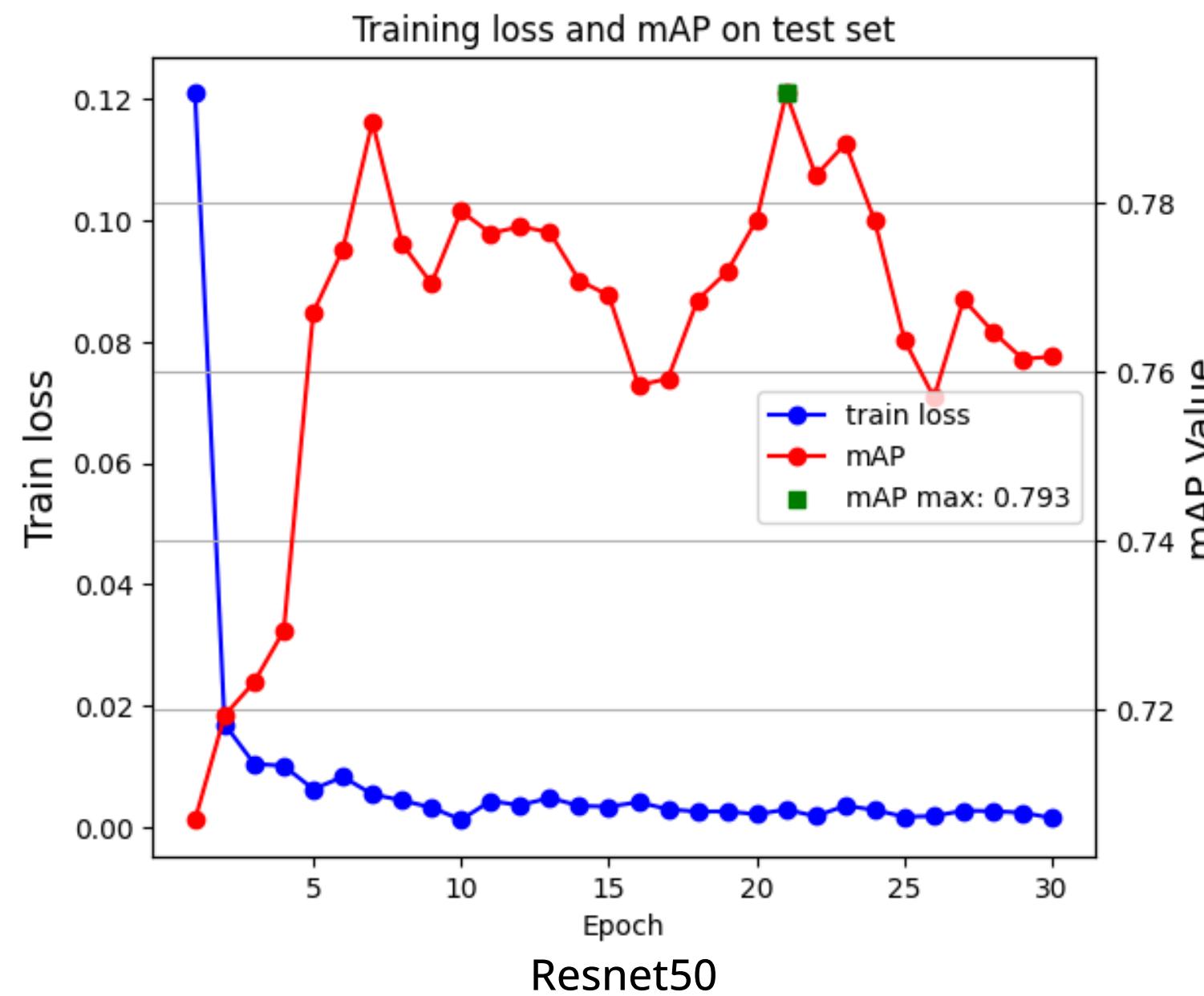
V. ĐÁNH GIÁ

1. Độ đo

2. Kết quả

V. ĐÁNH GIÁ:

2. Kết quả:



SKETCH- BASED IMAGE RETRIEVAL

VI. DEMO



THANK YOU

Sketch-Based Image Retrieval



21522028@gm.uit.edu.vn

21522057@gm.uit.edu.vn



References

- ResNet-50: The Basics and a Quick Tutorial. Datagen
<https://datagen.tech/guides/computer-vision/resnet-50/>
<https://phamdinhkhanh.github.io/2020/05/31/CNNHistory.html#46-resnet-50-2015>
- VGG16 architecture:
[Understanding VGG16: Concepts, Architecture, and Performance \(datagen.tech\)](#).
<https://ichi.pro/vi/vgg16-la-gi-gioi-thieu-ve-vgg16-267001881294357>
- Triplet margin loss:
[TripletMarginLoss — PyTorch 2.3 documentation](#)
[Triplet Loss: Intro, Implementation, Use Cases \(v7labs.com\)](#)
- Wikipedia. Evaluation measures (information retrieval) (20/12/2023)
[https://en.wikipedia.org/wiki/Evaluation_measures_\(information_retrieval\)](https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval))
- Dataset: [The Sketchy Database \(gatech.edu\)](#) <https://sketchy.eye.gatech.edu/>