

## Project Overview

This project focuses on training and evaluating a **Siamese Neural Network (SNN)** for **one-shot facial recognition**. The model is tested on data from unseen classes, evaluating its generalization capabilities in classifying from just a single example.

### Problem: Making predictions using limited data during training

- CNNs are powerful for image classification due to their ability to extract hierarchical features through layered transformations.
- Weakness:** they require **many labeled examples per class** to perform well.
- Collecting such data is often impractical in dynamic or low-resource settings.
- One-shot learning** addresses this: classifies using only one **example per class**.
- Compares similarities** between inputs rather than learning a set of fixed classes.

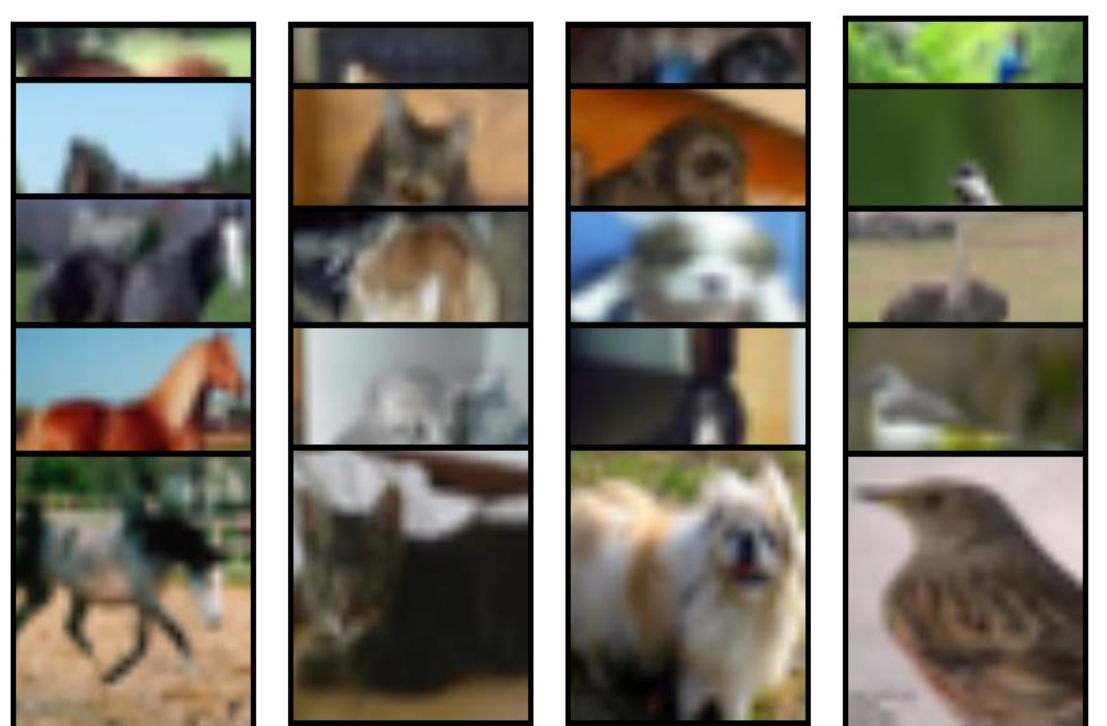
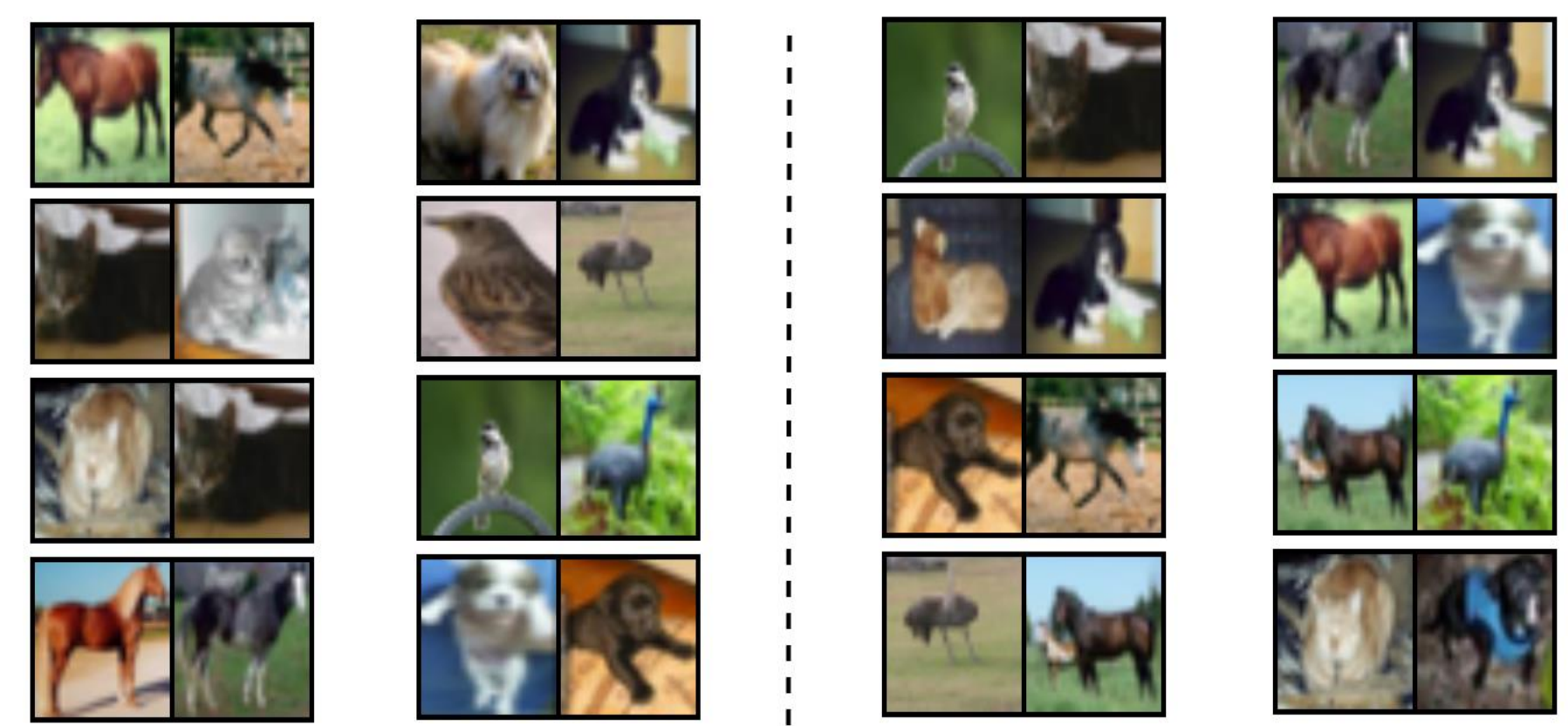

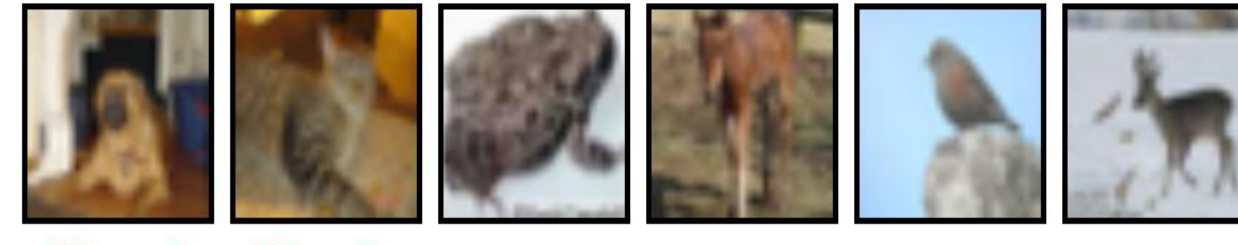
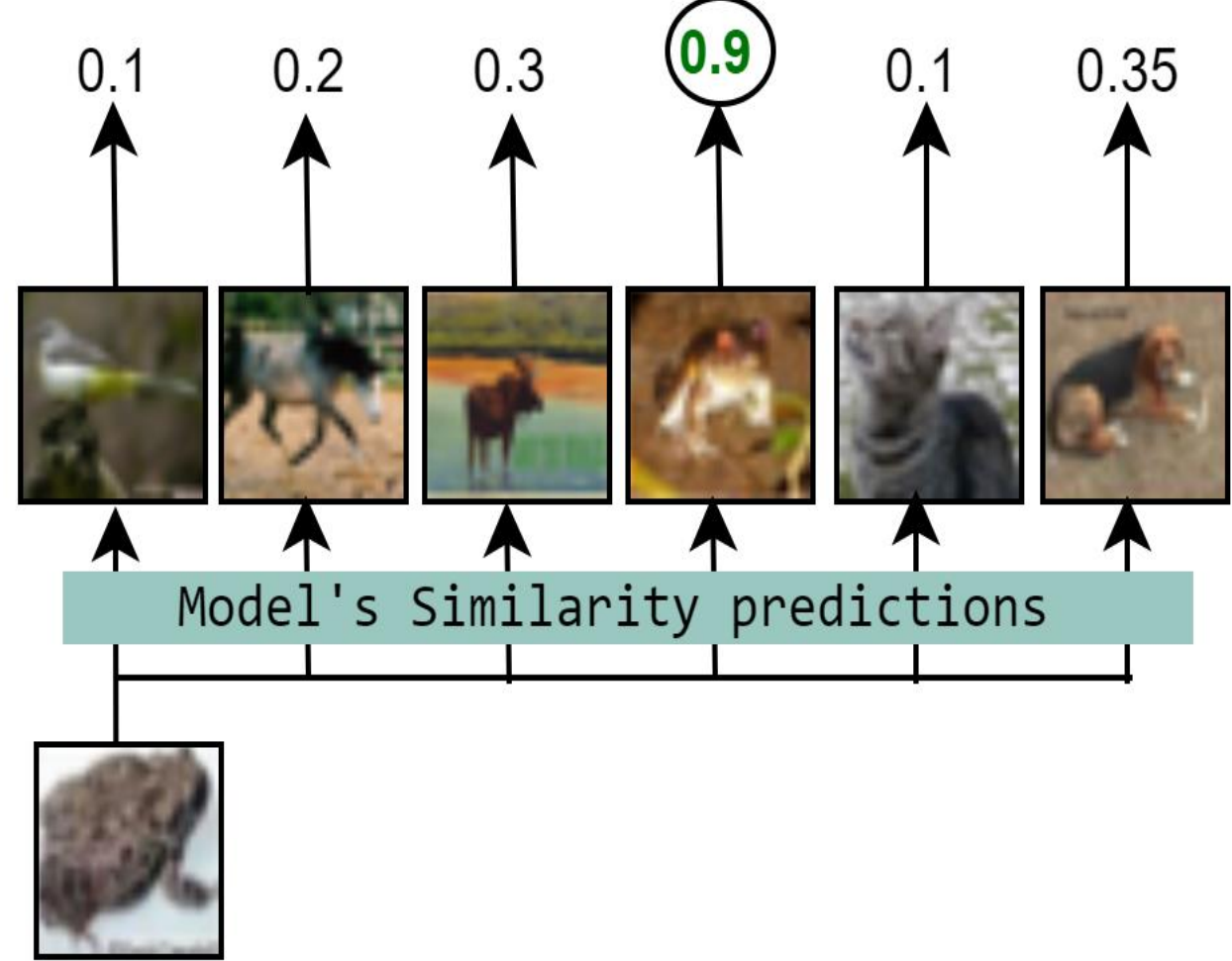
Classification with CNNs	One-shot Classification
<p><b>Training set</b></p>  <p>"Horse" "Cat" "Dog" "Bird"</p>	<p><b>Training set</b></p>  <p>Similar Different</p> <p>We organize the available data into <b>similar</b> and <b>different</b> pairs, and train a model to extract features from images and predict their similarities.</p>
<p><b>Queries</b></p>  <p>Predictions: "Dog" "Cat" ?</p>  <p>Predictions: "Dog" "Cat" "Frog" "Horse" "Bird" "Deer"</p> <p>The input's features are extracted and compared to the features of each image in the <b>support set</b>, and the <b>most similar class is chosen as the prediction</b>. In this example, the model learned general features that set different animals apart, so it can use the support set as a way to classify unseen data.</p> <p>No frog or deer images used during training, yet the model can classify them. =&gt; allows us to introduce new classes by simply updating the support set.</p>	<p><b>Queries</b></p> <p>The input's features are extracted and compared to the features of each image in the <b>support set</b>, and the <b>most similar class is chosen as the prediction</b>. In this example, the model learned general features that set different animals apart, so it can use the support set as a way to classify unseen data.</p> <p>No frog or deer images used during training, yet the model can classify them. =&gt; allows us to introduce new classes by simply updating the support set.</p>
<p>- Can accurately recognize classes, but only if they are included during training.</p> <p>- To predict new or unseen classes: large amounts of data collection is required; the model must be retrained every time the target classes change.</p> <p>=&gt; <b>less flexible</b> in environments where the class list <b>changes often</b> or labeled <b>data is scarce</b>.</p>	<p><b>Support set:</b> a single labeled example from each target class. =&gt; represents the set of possible predictions (easily modifiable; requires no fine-tuning)</p> <p><b>Classification Process</b></p> <p>Classification: "Frog"</p>  <p>Model's Similarity predictions</p>
CNNs need <b>multiple</b> labeled examples per class and <b>struggle with unseen</b> categories.	One-shot learning uses a single example per class, generalizing better in low-data scenarios.

Figure 1: Making predictions with supervised learning vs one-shot learning

## Siamese Neural Networks (SNN)

- consists of **two identical feature extraction subnets** (convolutional NN) that share the same weights. Each subnet processes one pair-image and outputs feature embeddings.
- The **absolute distance** between the two embedding vectors is the **output (1 or 0 for same or different images; Fig.3)**.

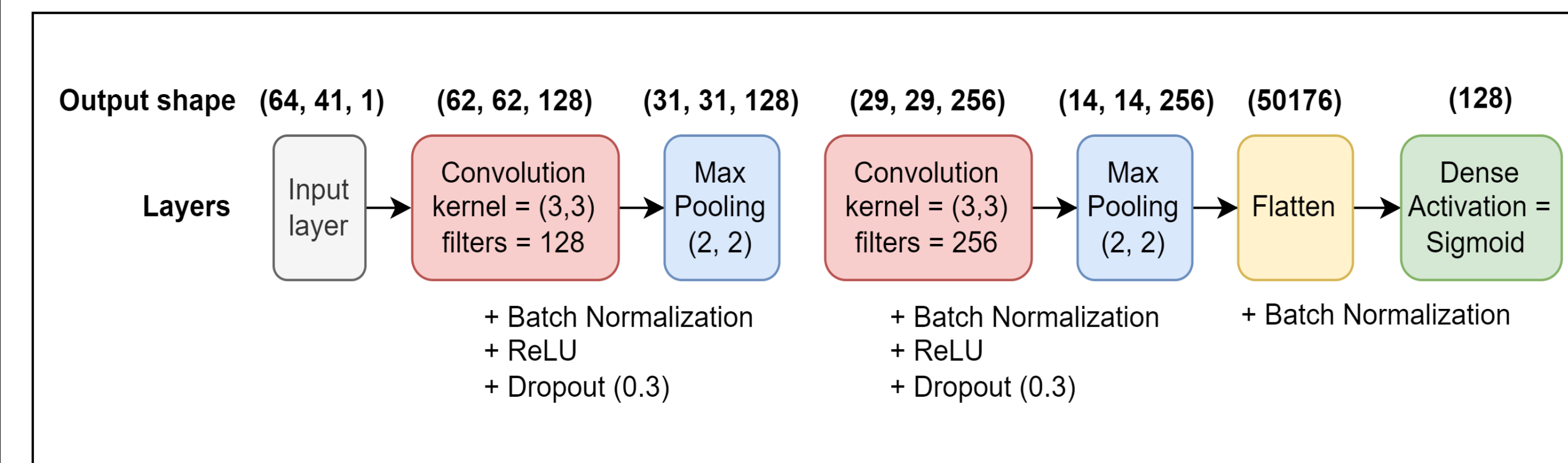


Figure 2. Summary of feature extraction layers

## Training

AT&T Lab. Cambridge Olivetti Faces Dataset; images of **40 individuals**, with **10 images** per person.  
Training: created **80,000 pairs** from **30 individuals** (evenly split **similar/different**).

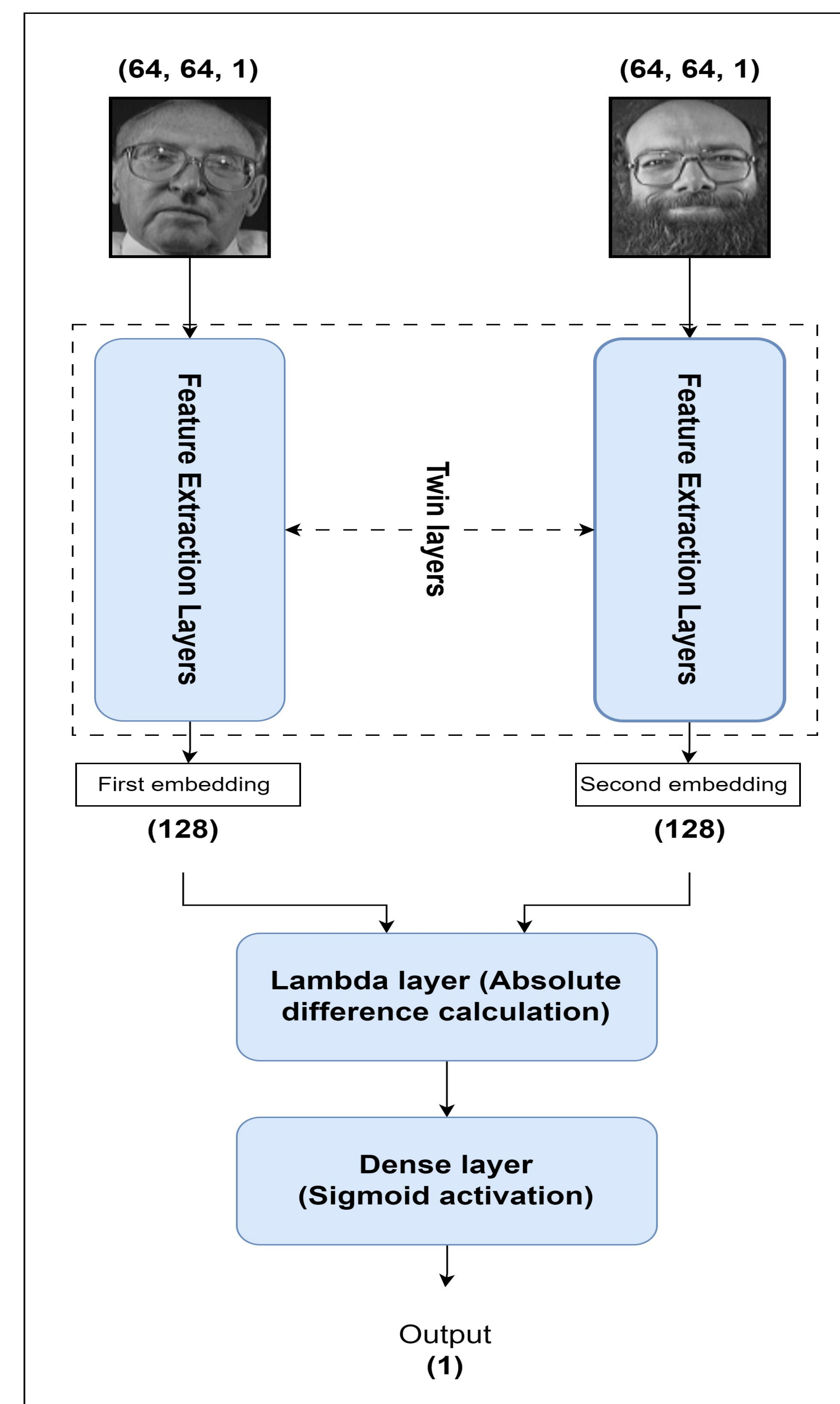


Figure 3. Summary of the Siamese Neural Network structure

## Evaluation: One-Shot Trials

**One-shot trials** are used to evaluate the model's ability to classify unseen classes using a single example per class. In each trial:

- A **support set** is formed from randomly selected images, each representing a unique class.
- A **query image** (not in the support set) is classified.
- Accuracy calculation: **proportion of correct predictions** across trials.

### First Evaluation: Unseen Identities from the Original Dataset

- 10,000** 10-way one-shot classification trials are performed using the Olivetti Faces dataset
- The model achieves an accuracy of **68.07%**.

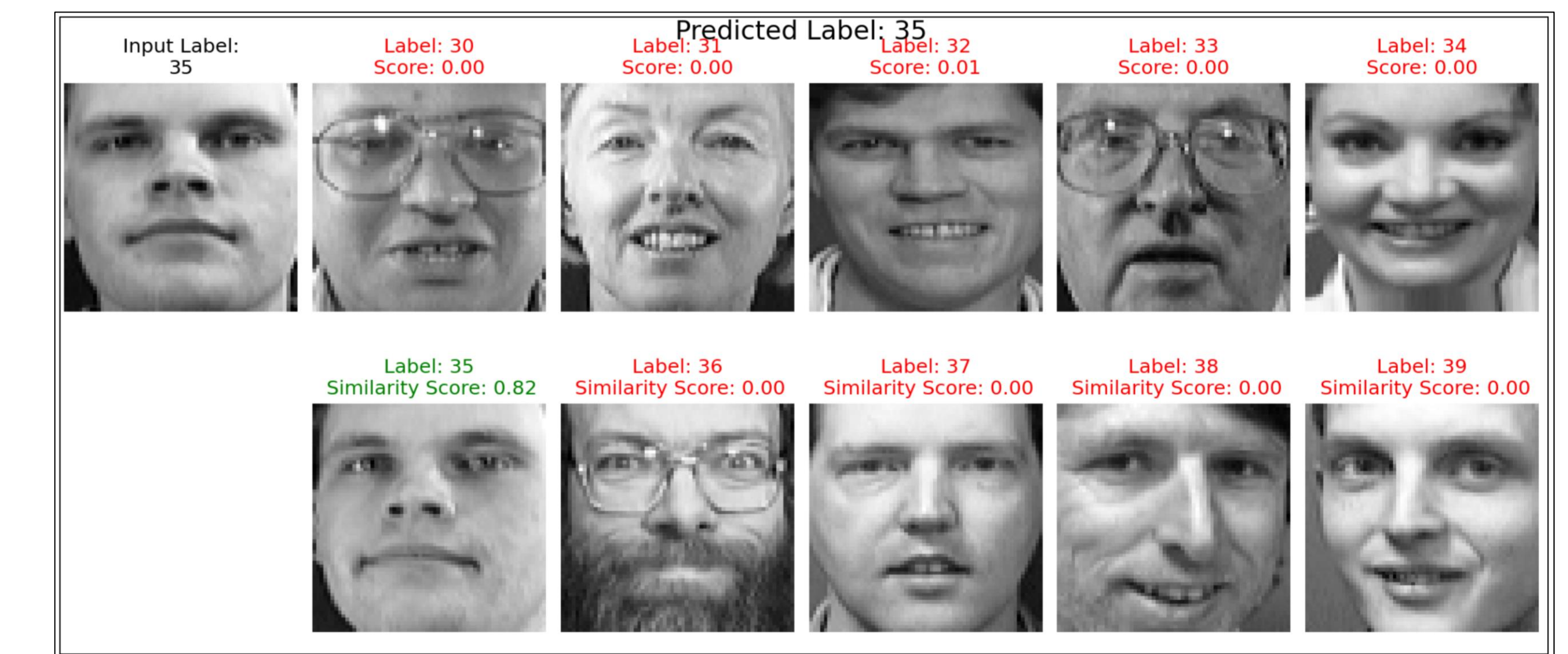


Figure 4: Example one-shot trial on the Olivetti dataset (original dataset)

### Second Evaluation: Unseen Identities from an Unseen Dataset

- The model was further tested on the **Yale Face Database** (unseen during training), with **no fine-tuning**.
- 10,000** 10-way one-shot trials using **10 randomly selected individuals**, yielding **44.27% accuracy**. The drop reflects the training data's limitations.

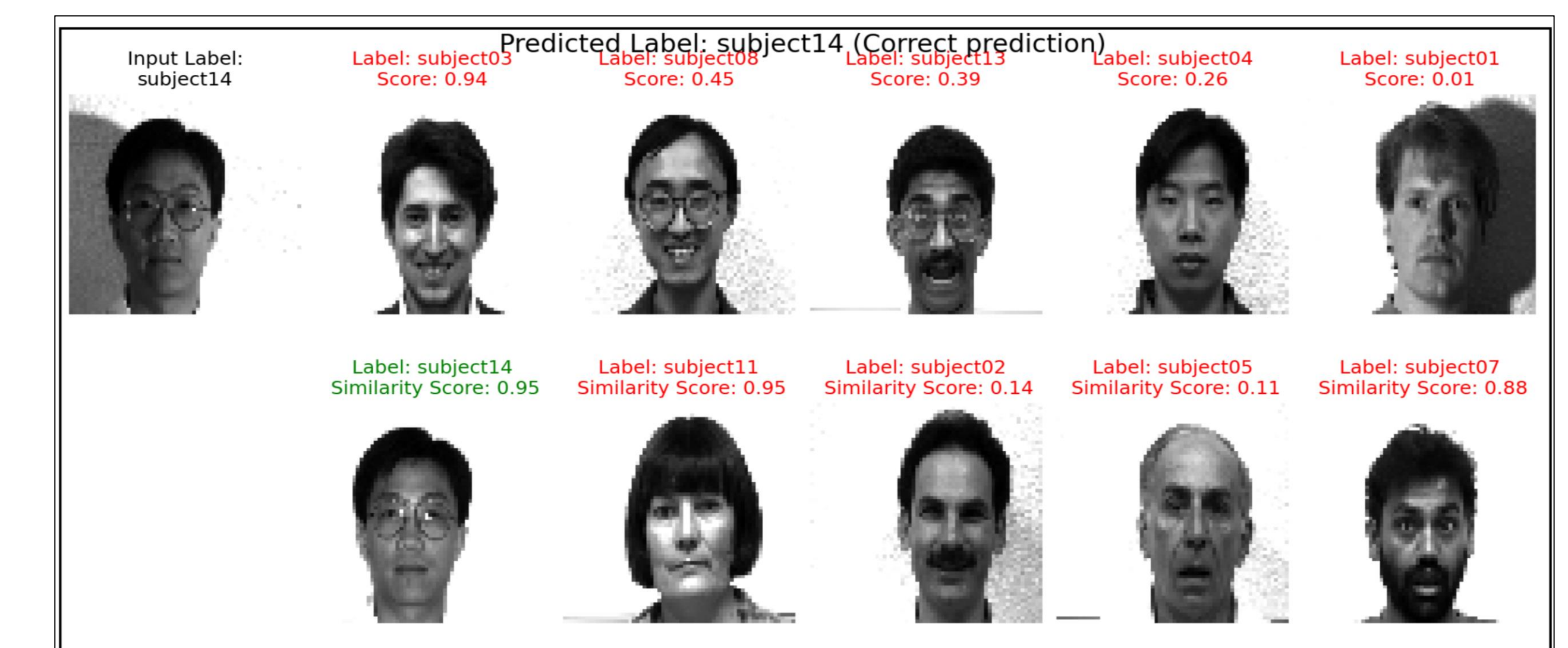


Figure 5: Example one-shot trial on the Yale Face Database (Unseen dataset)

## Limitations and Future Work

- The training dataset lacked variation in **lighting, pose, and angle**.
- While uniformity in the training data supported faster learning, it **limited generalization** to new conditions. The model struggled with variations that were not represented during training, like strong lighting in the Yale dataset.
- Future work: use **more diverse training dataset**; improve **data augmentation**, **loss function refinement**, and **optimization strategies**.