EPFL

Chocolate Recognition: Approach, Design, and Results, IAPR Final Project

Zaynab Hajroun
Mathieu Sanchez
Mischa Mez

Kaggle Team Name:
Group 2

May 23, 2025

Introduction & Problem Statement

- The Challenge: develop a model capable of accurately identifying different types of chocolates from images.

- Objective: achieve high classification accuracy on a diverse dataset of chocolate images, participating in the 'Deep Learning' Kaggle challenge.

- Constraints: Max 12M parameters, No pre-trained models No external data

- Dataset
  - 270 images (90 training images + 180 augmented images)
  - 180 testing
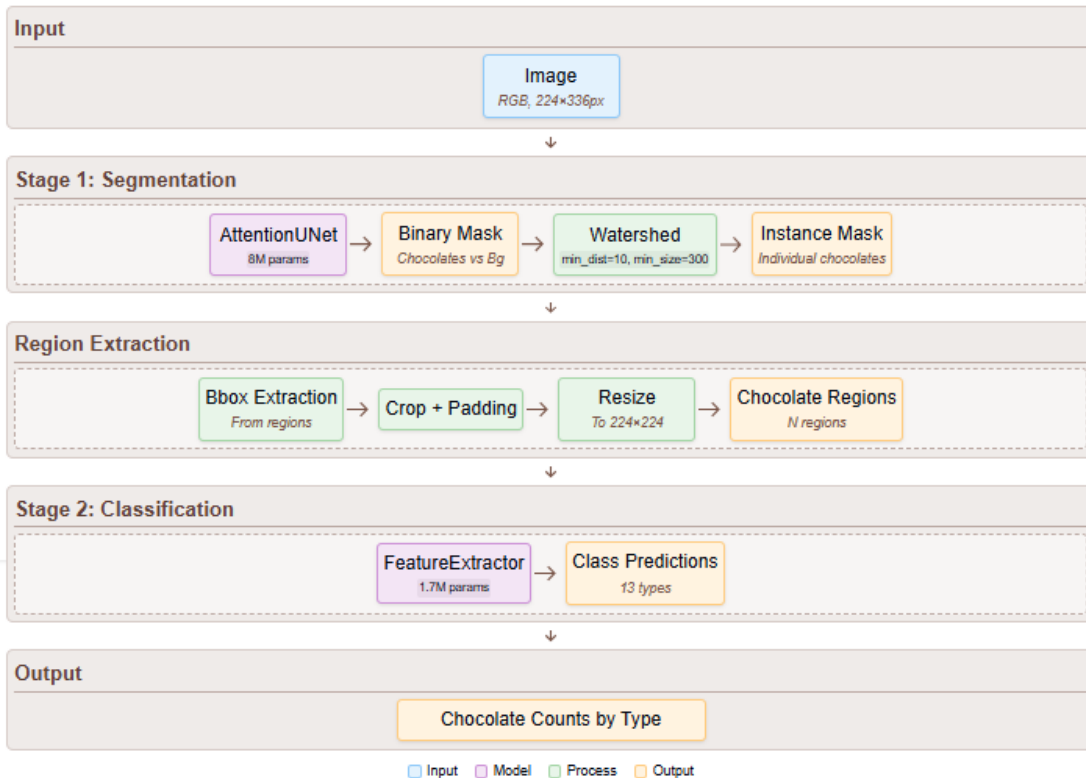 - 7 background types
  - 13 types of chocolates

Image Analysis and Pattern Recognition

Input Image

our Model
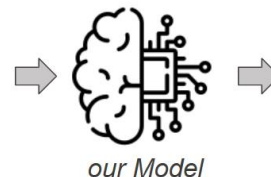
Category	Count
Jelly White	1
Jelly Milk	0
Jelly Black	0
Amandina	0
Crème brûlée	0
Triangolo	0
Tentation Noir	0
Comtesse	1
Noblesse	0
Noir authentique	0
Passion au lait	1
Arabia	1
Stracciatella	1

Prediction

Speaker

Model Architecture & Design Decisions

Image Analysis and Pattern Recognition

Chocolate Detection Pipeline

Input

Image
RGB, 224×336px

↓

Stage 1: Segmentation

AttentionUNet
8M params
→
Binary Mask
Chocolates vs Bg
→
Watershed
min_dist=10, min_size=300
→
Instance Mask
Individual chocolates

↓

Region Extraction

Bbox Extraction
From regions
→
Crop + Padding
→
Resize
To 224×224
→
Chocolate Regions
N regions

↓

Stage 2: Classification

FeatureExtractor
1.7M params
→
Class Predictions
13 types

↓

Output

Chocolate Counts by Type

☐ Input ☐ Model ☐ Process ☐ Output

Technical Architecture

Image Analysis and Pattern Recognition

AttentionUNet

- Attention Gates: Focus on chocolates, suppress background
- Skip Connections: Preserve spatial details
- BCEDiceLoss: Handle class imbalance

Watershed Algorithm

- Separate touching chocolates
- Tuned parameters (min_distance=10, min_size=300)

FeatureExtractor

- 5-block CNN backbone
- Skip connection (block 3→5)
- Global avg pooling + FC head
- LabelSmoothingLoss

Speaker

Pipeline in Action

Image Analysis and Pattern Recognition

- Stage 1: Segmentation (Binary mask + Watershed separation)

Original Image | Binary Mask | Watershed Mask (4 components) | Detected Chocolates: 4

- Stage 2: Classification (Individual chocolate recognition)

Region 1:
Passion au lait: 0.93
Creme brulee: 0.01
Stracciatella: 0.01

Region 2:
Comtesse: 0.82
Jelly White: 0.07
Amandina: 0.02

Region 3:
Tentation noir: 0.77
Triangolo: 0.07
Stracciatella: 0.02

Region 4:
Tentation noir: 0.91
Triangolo: 0.02
Noblesse: 0.01

- Final Counts (13 chocolates detected "almost" perfectly)

Main Results

- Segmentation Dice Score: 98%

- Classification Accuracy: 97%
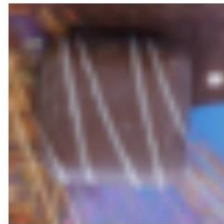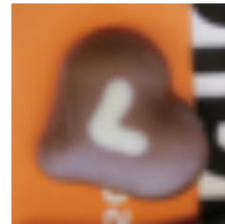
- Total Parameters: 9.7M
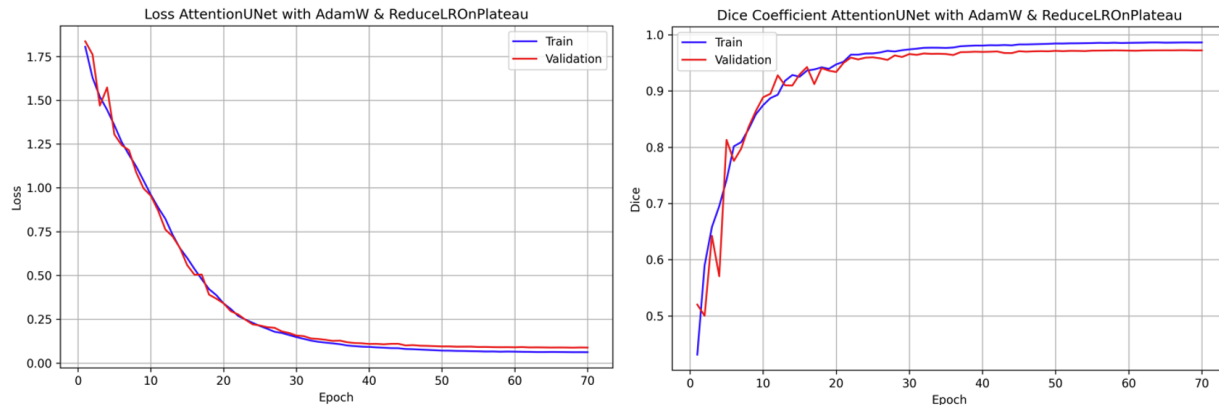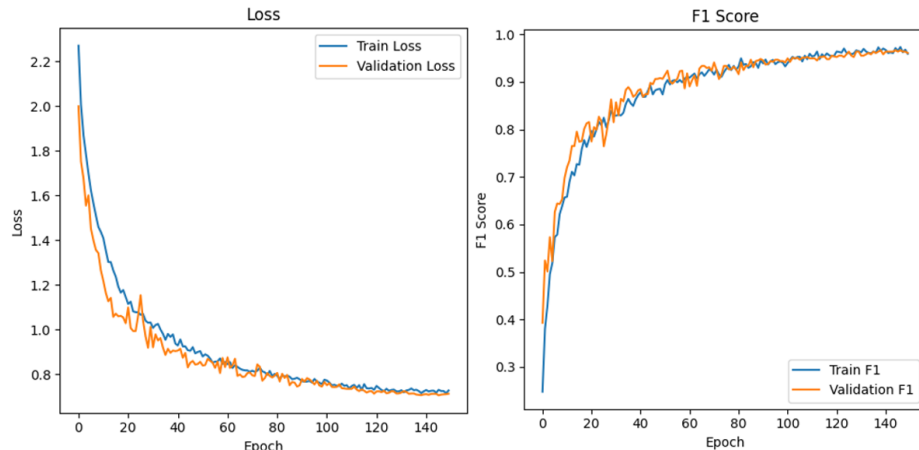
Strengths & Performance

- **Robust Segmentation**
  - Handles complex backgrounds
  - Separates touching chocolates
- **Accurate Counting**
  - Very good detection in validation set
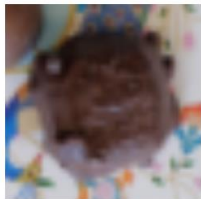
Original Image

Detected Chocolates (6 total)

Passion au lait (conf: 0.94)

Arabia (conf: 0.92)
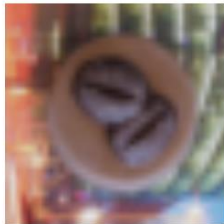
Stracciatella (conf: 0.89)
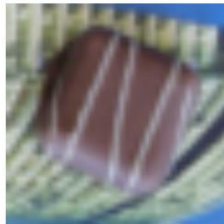
Stracciatella (conf: 0.88)

Passion au lait (conf: 0.97)

Arabia (conf: 0.91)

Strengths & Limitations

Image Analysis and Pattern Recognition
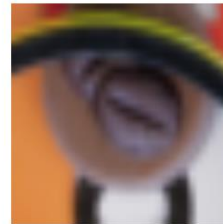
Tentation noir
(conf: 0.90)

Comtesse
(conf: 0.88)

Passion au lait
(conf: 0.96)

Comtesse
(conf: 0.89)
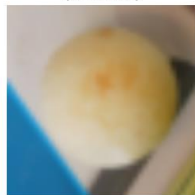
Comtesse
(conf: 0.58)

Tentation noir
(conf: 0.61)

Tentation noir
(conf: 0.66)

Jelly White
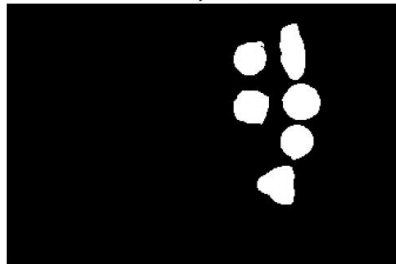(conf: 0.45)

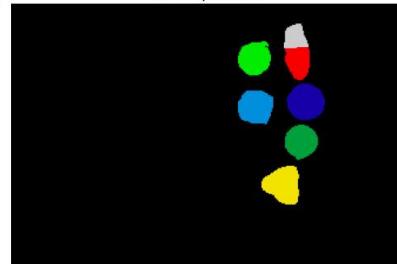Triangolo
(conf: 0.94)

Triangolo
(conf: 0.67)

Original Image

Binary Mask

Watershed Mask
(7 components)

Detected Chocolates: 7

Speaker

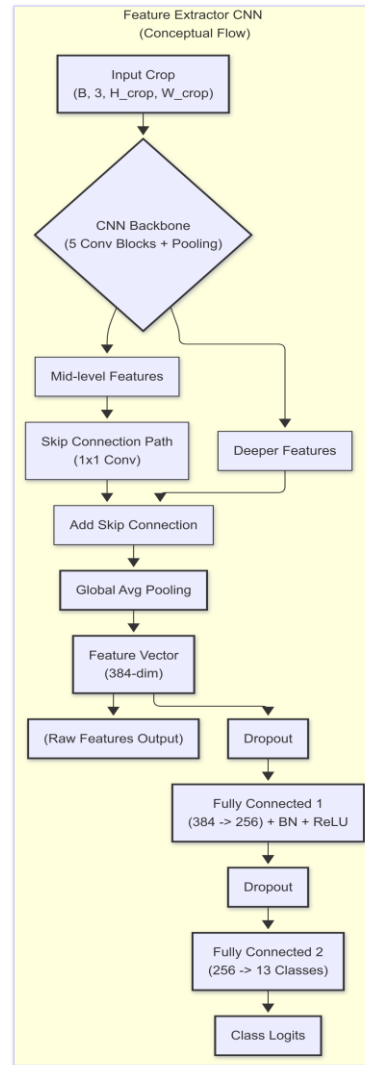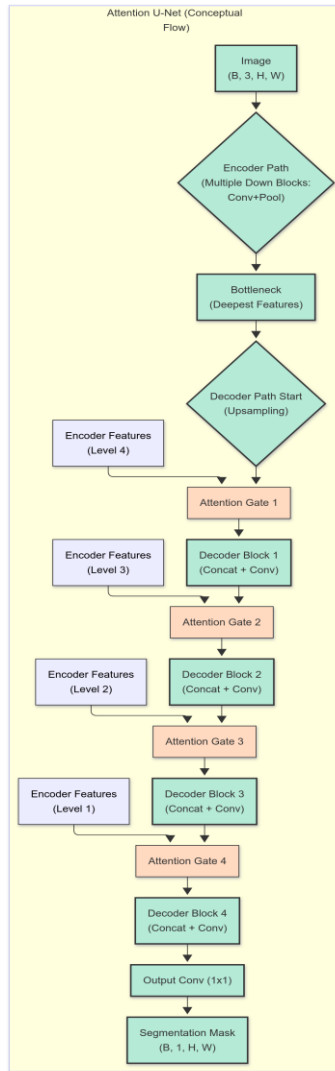# Conclusion & Future Work

Image Analysis and Pattern Recognition

- Implemented a two-stage pipeline for chocolate image recognition:
  - **Stage 1**: Segmentation with AttentionUNet + watershed for accurate detection.
  - **Stage 2**: Classification using a custom CNN with regularization and skip connections.
- Achieved high segmentation accuracy and robust classification under a 12M parameter limit.
- Design effectively handles small objects, occlusions, and varied backgrounds.

Improvements:

- Data augmentation (Copy-paste, on specific types)
- Watershed improvement for Amandina

Thank You For Your Attention Questions?

Speaker

Image Analysis and Pattern Recognition

Generated by ChatGPT

Detailed Model Architectures

Image Analysis and Pattern Recognition

Augmentation Details & Impact

Image Analysis and Pattern Recognition

- **Stage 1: Segmentation (Attention U-Net Training)**
  - Objective: Improve U-Net's ability to identify diverse chocolate regions & maintain image-mask alignment.
  - Method:
    - Geometric (Paired):
      - Random Horizontal/Vertical Flips
      - Random Rotation
    - Photometric (Image Only):
      - Color Jitter (Brightness, Contrast, Saturation)
    - Sizing & Final (Paired):
      - Aspect-Ratio Preserving Resize (smaller dim to 224px)
      - ToTensor, Normalize (Image only)
  - Impact: Enhanced U-Net invariance to view/lighting; consistent ground truth.

- **Stage 2: Classification (FeatureExtractor Training on Crops)**
  - Objective: Make classifier robust to crop variations from Stage 1 & diverse chocolate appearances.
  - Method: Initial Sizing: Aspect-Ratio Preserving Resize
    - Geometric:
      - Random Rotation
      - Random Resized Crop
      - Random Horizontal/Vertical Flips
    - Photometric:
      - Color Jitter
      - Random Grayscale
      - Random Gaussian Blur
      - Final: ToTensor, Normalize
  - Impact: Classifier generalizes better from varied/imperfect crops; learns core features

Future Consideration: Copy-Paste Augmentation (for Stage 1 U-Net)

Future Consideration: Copy-Paste Augmentation (for Stage 1 U-Net)

- **Concept (Ghiasi et al., 2021):** Randomly paste reference chocolates (with masks) onto training images.
- **Potential Benefits:**
  - Increase instance density & variety.
  - Improve occlusion handling.
  - Better balance class contexts.
- **Implementation:** Modify AugmentedChocolateBinarySegmentationDataset to composite reference stamps before other augmentations.