ETRI School

Multi-Task Learning for Recognizing Customer Characteristics

DongHwan Lee

2023. 11. 30. Seminar

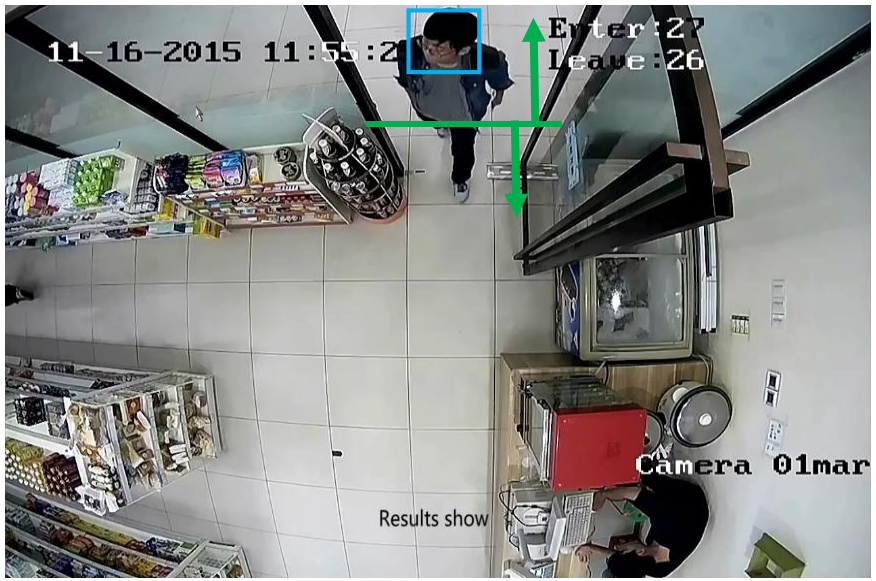
UST 과학기술연합대학원대학교
UNIVERSITY OF SCIENCE & TECHNOLOGY

Contents

Research Objective

Customer Characteristics Analysis

- My final goal is to build an integrated system for analyzing customer characteristics.



People Counting

Age/Gender/Expression ··· Recognition

System Utilization Examples

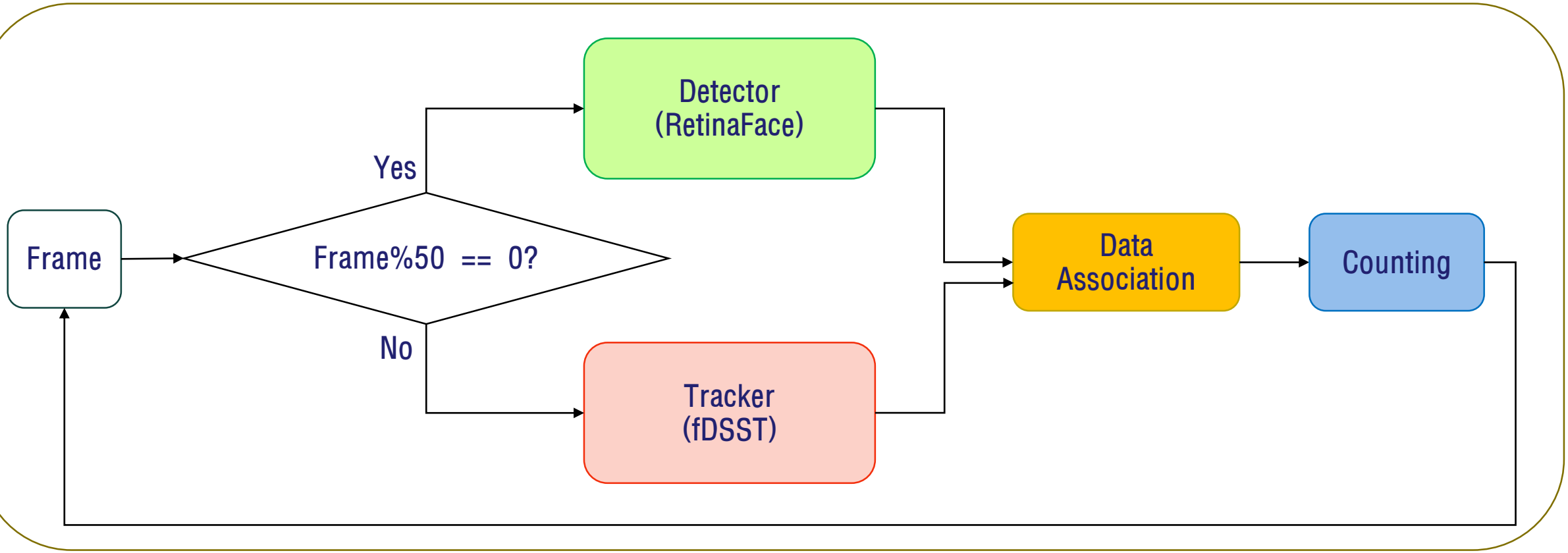
- Sales can be improved by providing a shopping experience that can increase satisfaction.
- Store can be managed efficiently.

Section B (visited 3) & C (visited 15)
→ Move popular products from section B to section C

20, Female, Section A,
→ Discount coupons for the most purchased products by women in their 20s

People Counting System Overview

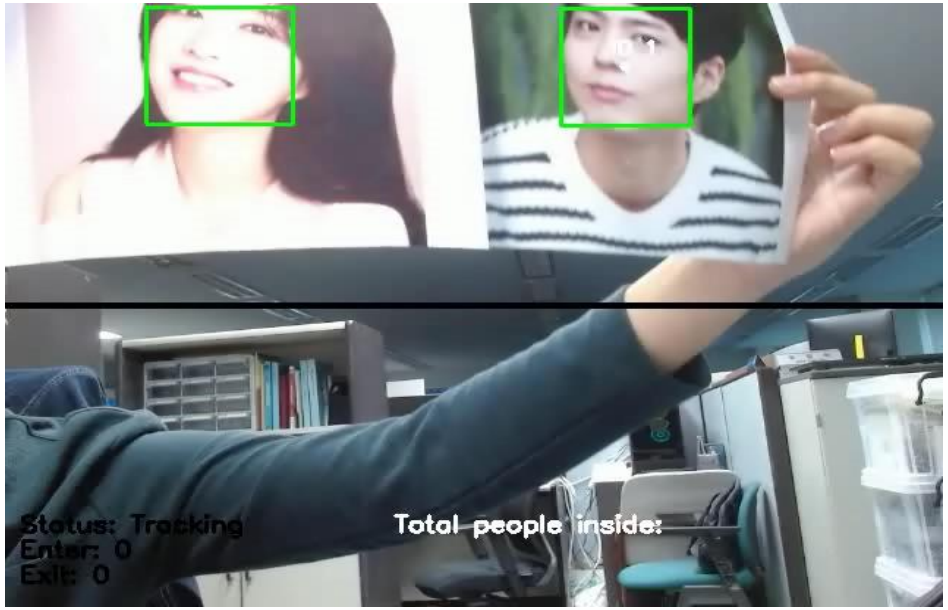
- Detector is used to find specific objects in an image.
- Tracker is used for object tracking in a series of images.



Demonstration

- It needs to be modified. (Not yet published)

Enter ↓



Exit ↑

Method – Facial Expression Recognition

D. -H. Lee and J. -H. Yoo, "CNN Learning Strategy for Recognizing Facial Expressions," *in IEEE Access*, vol. 11, pp. 70865-70872, Jul. 2023, doi: 10.1109/ACCESS.2023.3294099.

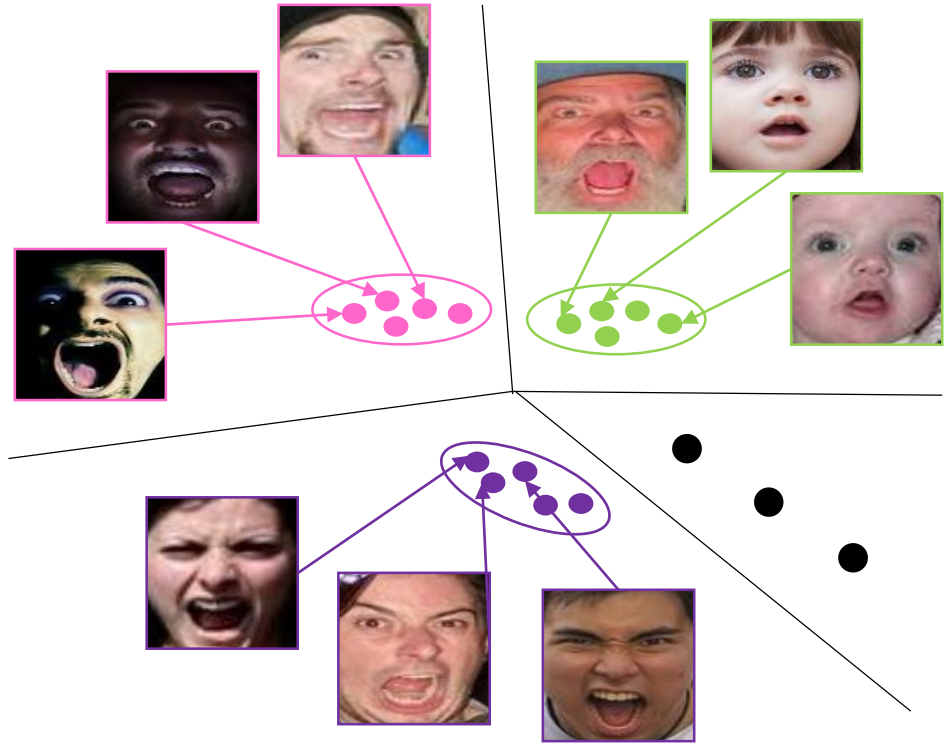
Motivation

- Solving the performance degradation problem caused by an increase in the number of similar facial expressions to classify at a time.

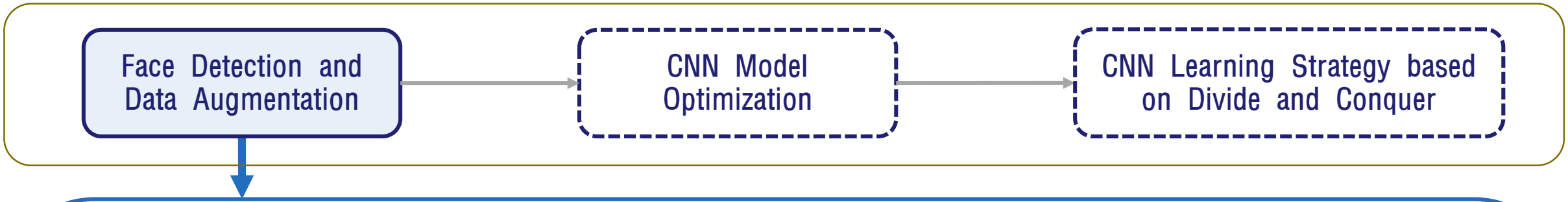
[Causes of reduced performance]

⦿ Intra-class difference

⦿ Inter-class similarity



Preprocessing

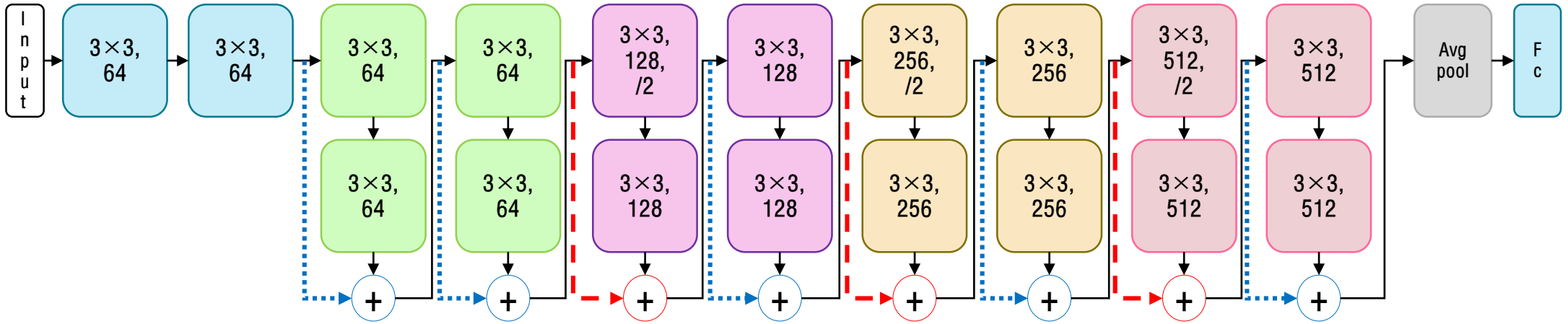


- ⦿ Face Detection (Normalization)
  - ✓ Use RetinaFace to remove unnecessary areas
  - ✓ Resolution: 130×160
- ⦿ Data Augmentation
  - ✓ Horizontal Flip, Randomly Rotation（ -10° ~ +10° ）

CNN Model Optimization

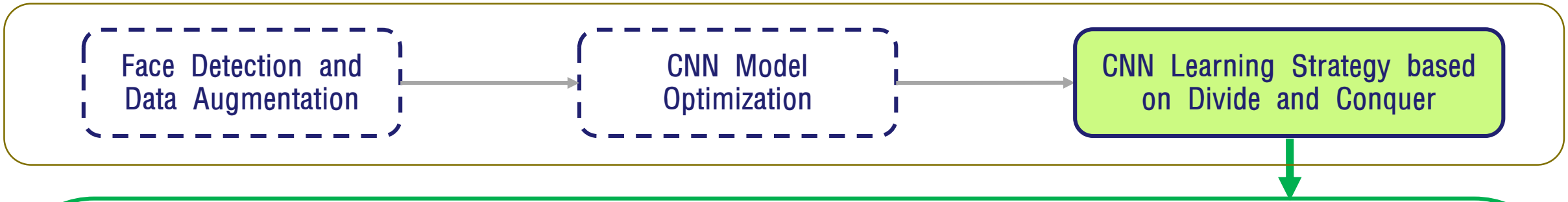
Face Detection and Data Augmentation → CNN Model Optimization → CNN Learning Strategy based on Divide and Conquer

⦿ Based on the pre-trained ResNet-18

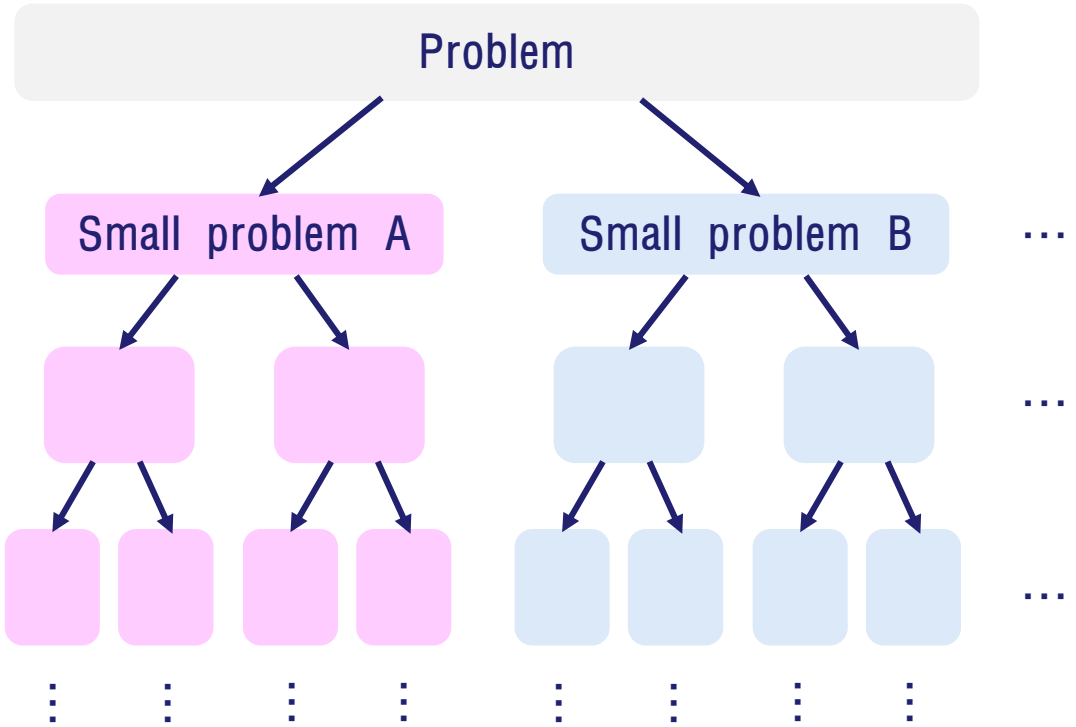


Proposed Model

Learning Strategy (1/3)



⊙ Divide and conquer is a computer algorithm for solving one problem by dividing it into several smaller problems.



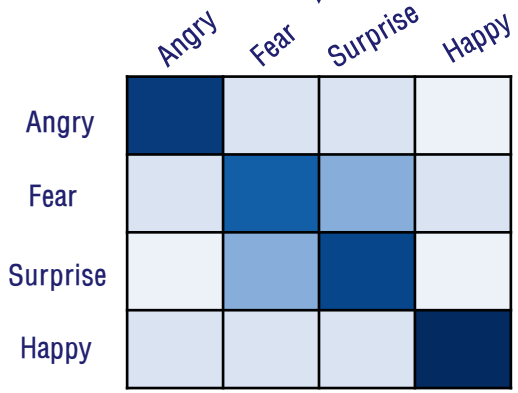
Learning Strategy (2/3)

Face Detection and Data Augmentation → CNN Model Optimization → CNN Learning Strategy based on Divide and Conquer

⊙ Utilize the confusion matrix obtained from all facial expression learning results.

Q. How to find the most similar expressions?

Problem (Example: 4class FER)



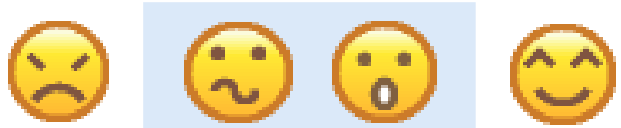
Learning Strategy (3/3)

Face Detection and Data Augmentation → CNN Model Optimization → CNN Learning Strategy based on Divide and Conquer

① Convert confusion matrix to weighted directed graph

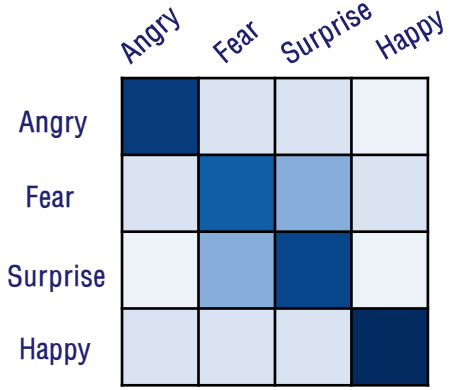
② Applying the shortest path search algorithm and grouping expressions

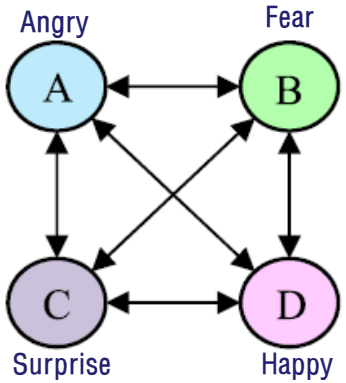
③ CNN model learning by group.



⊙ Group1: 3Class

⊙ Group2: 2Class





Algorithm 1 Divide and Conquer based CNN Learning Strategy

Input: a facial expression dataset E

$E \leftarrow [C_1, C_2, \ldots, C_n]$, n = the label of facial expression

Phase 1 - Grouping similar facial expressions

Train dataset E using the optimized CNN for creating the confusion matrix

Analyze the confusion matrix for grouping similar facial expressions

$S \leftarrow [S_1, S_2, \ldots, S_m]$, m = the label of similar facial expression group

Phase 2 - Classifying similar facial expression groups

Train generated set S using the optimized CNN

$S_m \leftarrow [R_{1,F_1}, R_{2,F_2}, \ldots, R_{m,F_m}]$

F_m = the number of final facial expressions in S_m

for all attribute $R_{m,F_m} \in S_m$ **do**

  while $F_m > 1$ **do**

    if $F_m > 2$ **then**

      if F_m is even **then**

        Generate new facial expression group S'_m

        $S'_m \leftarrow [R_{1',F_m/2}, R_{2',F_m/2}]$

        $R_{m,F_m} \leftarrow S'_m$

        $F_m \leftarrow F_m/2$

      else

        Train final facial expression in R_{m,F_m} using the optimized CNN

        $F_m \leftarrow F_m/2$

    else

      Train final facial expression in R_{m,F_m} using the optimized CNN

      $F_m \leftarrow F_m/2$

  end

end

Experiment & Result

Environment

- GPU: NVIDIA RTX A6000
- Hyperparameter : Epoch – 10,000, learning rate – 0.001, optimizer – Adam, loss function – L1

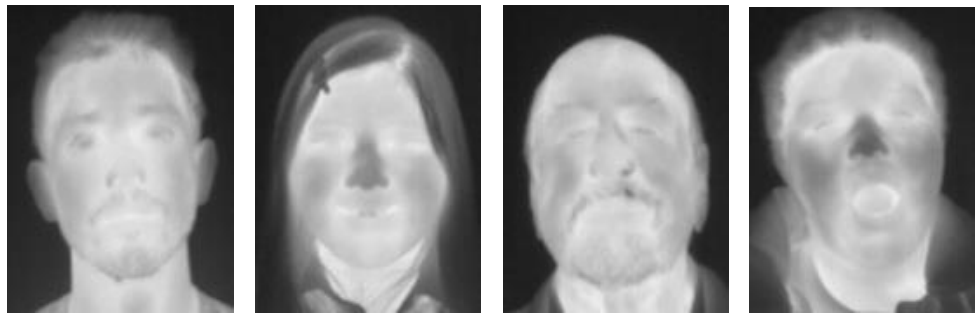
⦿ Grouping Expressions $k = \left\lceil \frac{N}{2} \right\rceil$

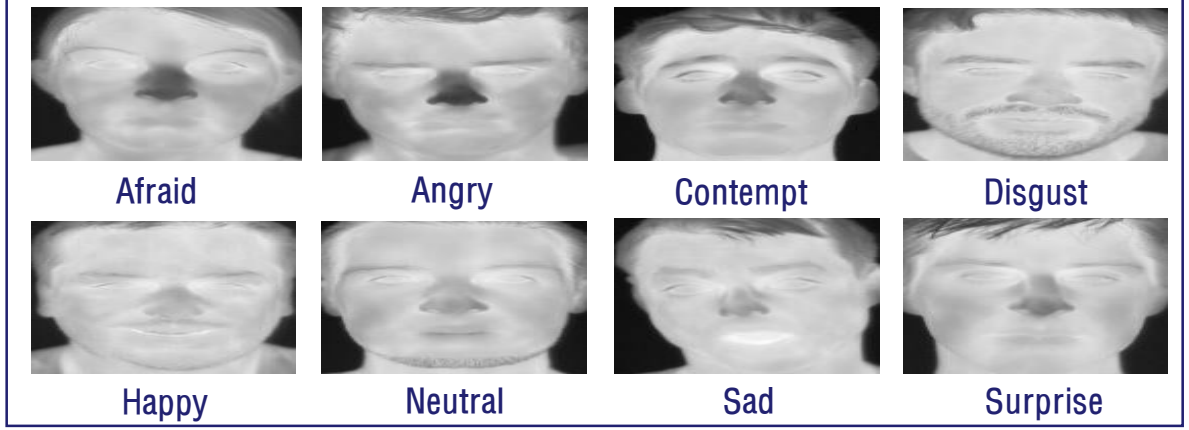
- N: Total number of labels in the dataset
  - Ex) N = 4, k = 2
  - Ex) N = 7, k = 4, k = 2
  - Ex) N = 8, K = 4, k = 2

Thermal DB

- Small scale, collected from the laboratory.

Dataset	Samples	Subject	Resolution	Expressions
Tufts	565	113	336×256	4
RWTH	1,782	90	1024×768	8

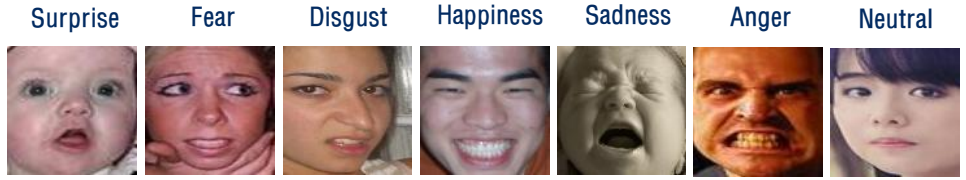




RGB DB

- Large scale, collected from the web.

Dataset	Samples	Subject	Resolution	Expressions
RAF	15,339	-	100×100	7
FER2013	35,887	-	48×48	7



Result

Datasets	Expressions	Methods	Group	Accuracy(%)
Tufts	4	Kamath *et al.*	-	96.20
		Optimized model	-	96.07
		Proposed method	ⓐⓑ, ⓒ, ⓓ	**97.75**
RWTH	8	Kopaczka *et al.*	-	46.70
		Optimized model	-	54.72
		Proposed method	ⓐⓔⓕⓖ, ⓗ, ⓘ, ⓙ, ⓚ	**62.50**
RAF	7	Zhang *et al.*	-	90.35
		Optimized model	-	85.10
		Proposed method	ⓐⓖⓗⓙ, ⓔ, ⓘ, ⓚ	**90.81**
FER2013	7	Pham *et al.*	-	76.82
		Optimized model	-	69.04
		Proposed method	ⓐⓔⓖⓚ, ⓗ, ⓘ, ⓙ	**77.83**

ⓐ: Neutral, ⓑ: Smile, ⓒ: Sleepy, ⓓ: Shock, ⓔ: Fear, ⓕ: Contempt, ⓖ: Sadness,
ⓗ: Disgust, ⓘ: Surprise, ⓙ: Happiness, ⓚ: Anger

Conclusion

Summary & Future Works

[Summary]
- My goal is to create an all-in-one system for analyzing customer characteristics.
- This semester, I plan to implement each module.
  1. People Counting Module
  2. Age/Gender/Expression … Recognition Module

[Future Works]
- Extend from facial expression to age, gender.
- Combine two modules.

Q & A

Thank you