ETRI School

Multi-Task Learning for Recognizing Customer Characteristics

DongHwan Lee

2023. 09. 21. Seminar

UST 과학기술연합대학원대학교
UNIVERSITY OF SCIENCE & TECHNOLOGY

Contents

01 Introduction

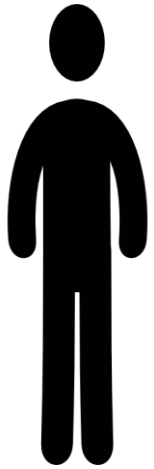
02 Related works

03 Conclusion

Introduction

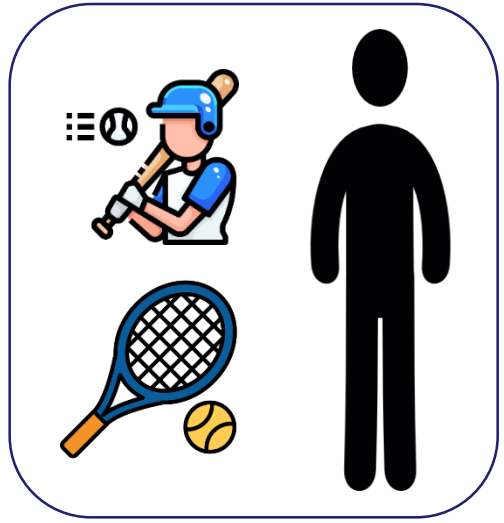
What is Multi-Task Learning?

Goal: Perform tasks well (baseball, tennis)



How?

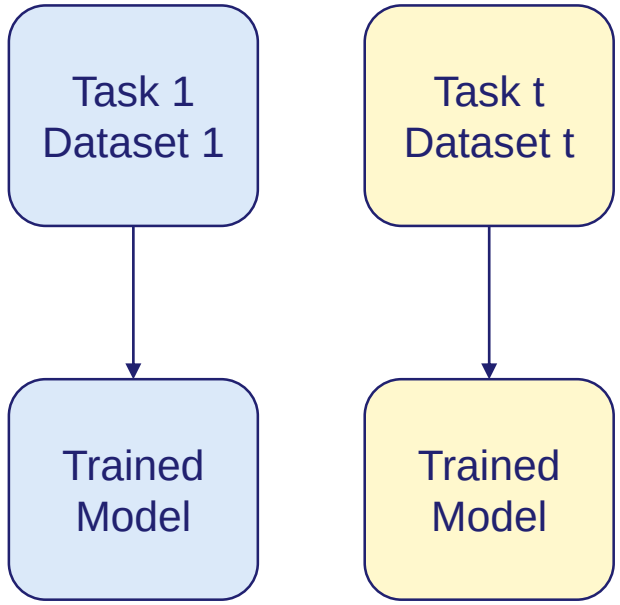
Learn tasks at the same time (baseball, tennis)



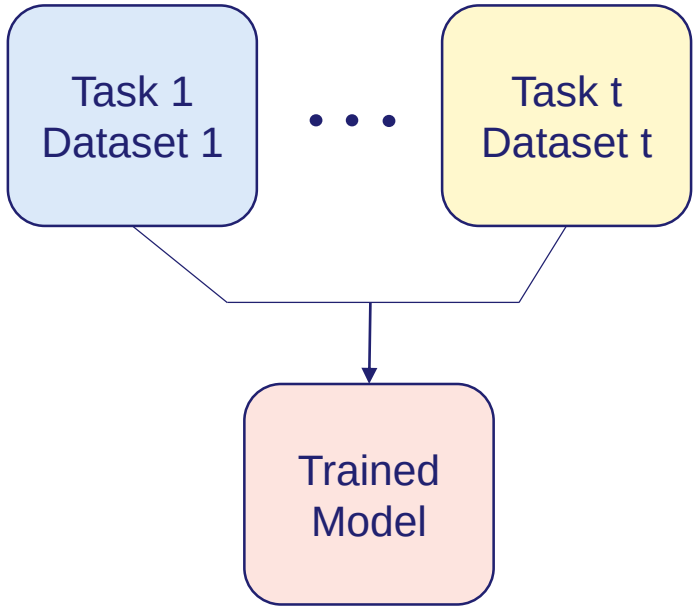
- Multi-task learning (MTL) is a subfield of machine learning in which multiple learning tasks are solved at the same time, while exploiting commonalities and differences across tasks. (Wikipedia)

Why use Multi-Task Learning?

Single task learning



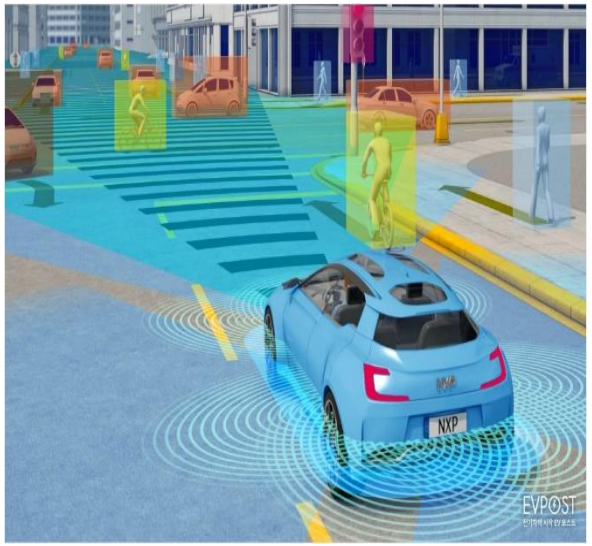
Multi-task learning



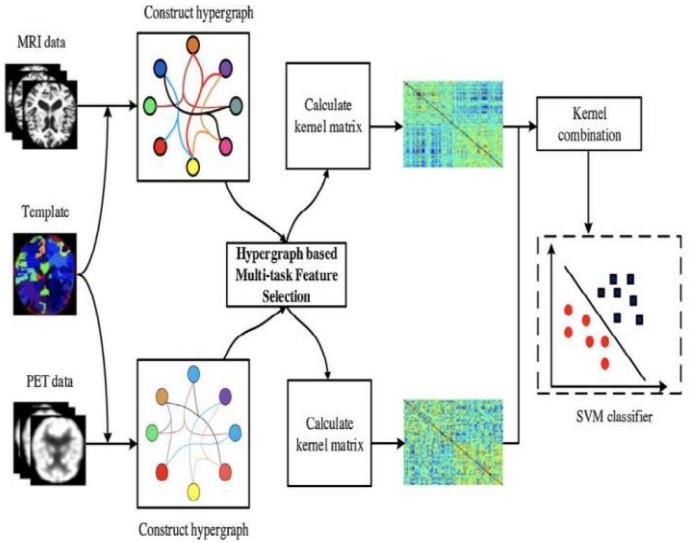
- MTL offers advantages like improved faster model convergence, and reduced model overfitting due to shared representations.

Where can Multi-Task Learning be applied?

Autonomous driving system



Disease diagnostic system



Smart retail system

• • •

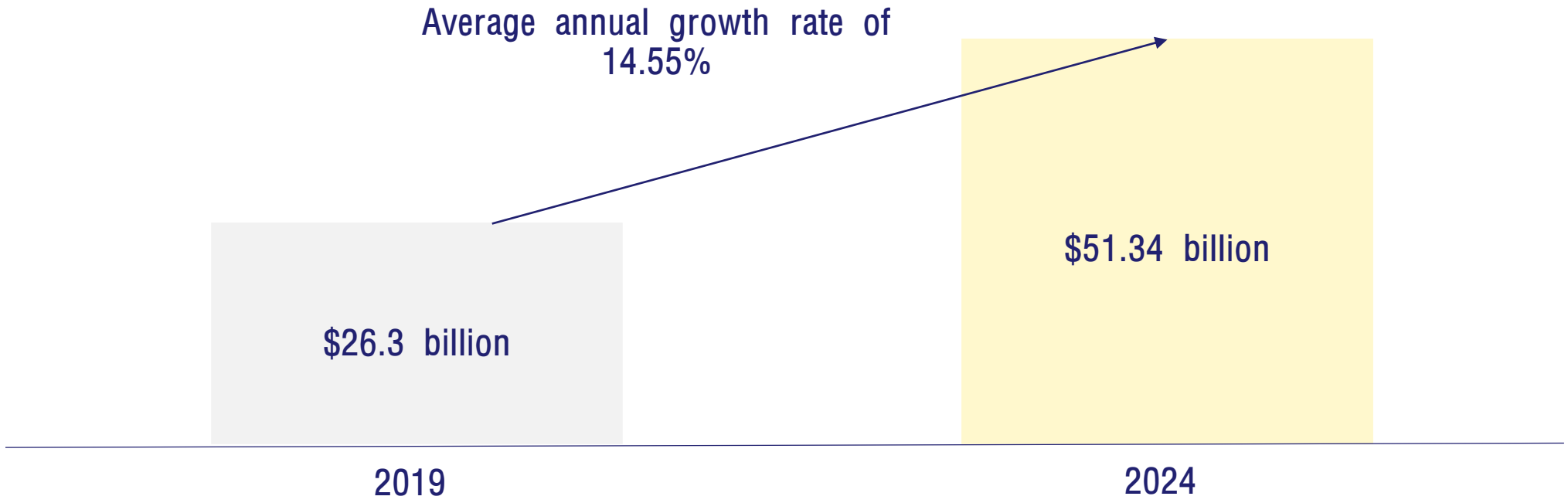
- MTL is applicable to all areas using deep learning.

Smart Retail System (1/2)

- My research topic is building a smart retail system.
- The goal of smart retail system is to increase sales by providing a shopping experience that can increase consumers´ satisfaction.

Image: https://www.einfochips.com/blog/facial-recognition-in-retail-enhance-in-store-customer-experience-and-improve-retailer-operations/

Smart Retail System (2/2)

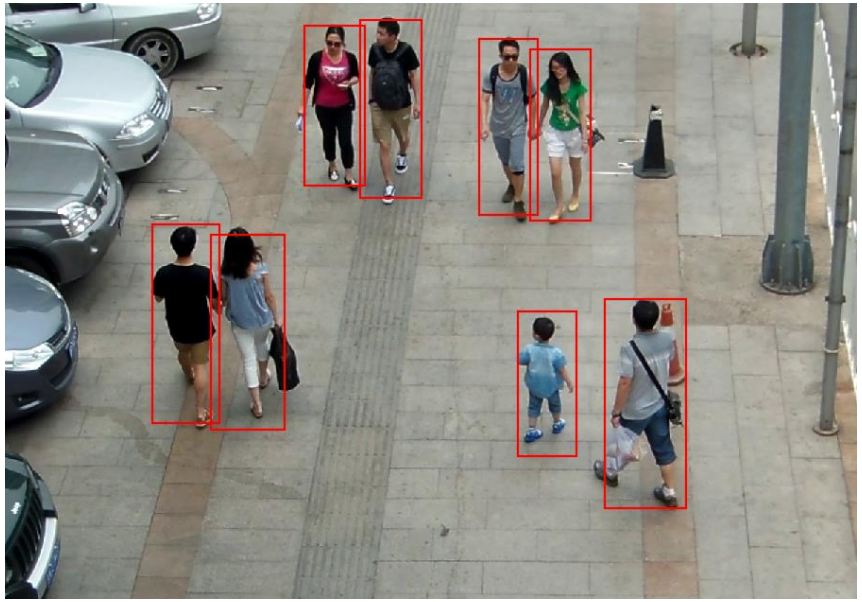


- The global smart retail market is growing steadily every year.
- Therefore, the demand for technology is expected to be in the future.

TechNavio, Global Smart Retail Market, 2020

Customer Characteristics Analysis (1/5)

Full body analysis



Face analysis

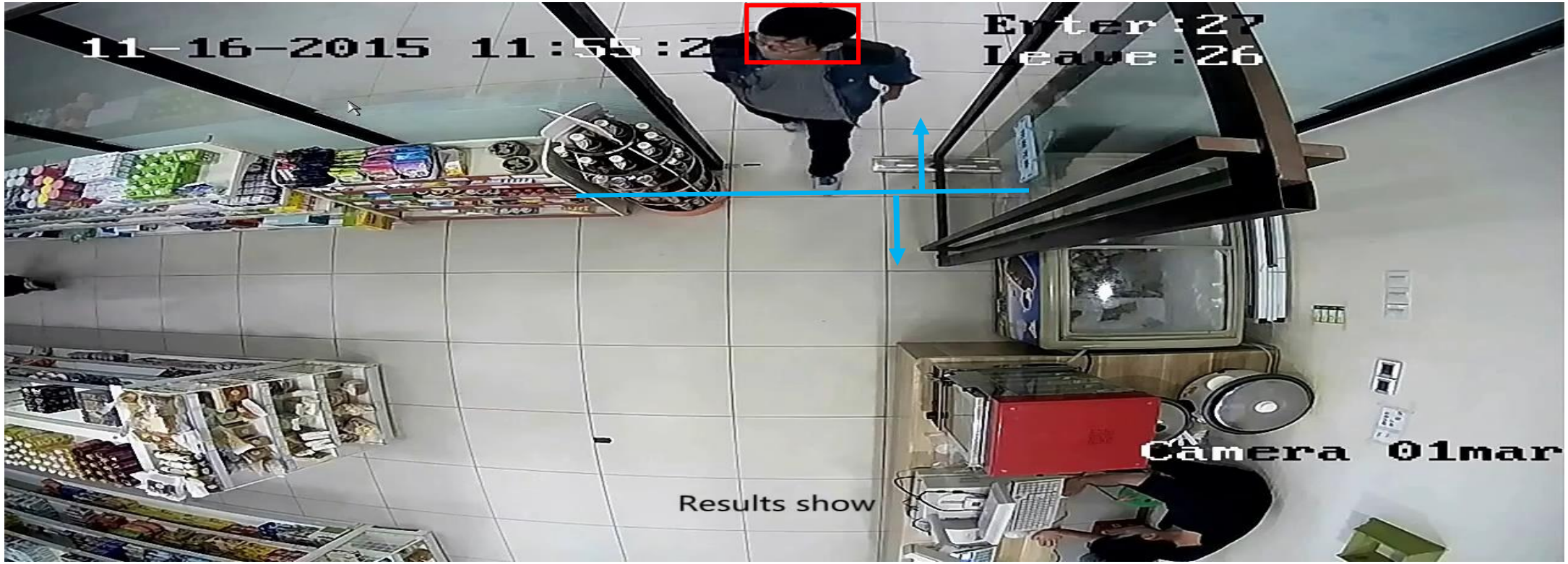
- The most important thing to build a system is to understand customer characteristics.
- I will focus on facial area analysis to reduce computing costs.

Customer Characteristics Analysis (2/5)

Country	Company(Solution)	Descriptions
China	Hikvision	People counting, Predicting age, gender, …
Taiwan	CyberLink(FaceMe)	Predicting age, gender, satisfaction(positive, negative), …
USA	Affectiva(AFFDEX)	Prediction facial expressions, action units, …
Japan	OMRON(OKAO Vision)	Predicting age, gender, facial expressions, …
USA	Generated photos	Create a new face by changing gender, head pose, age, …

- I found a company that offers facial-related services.

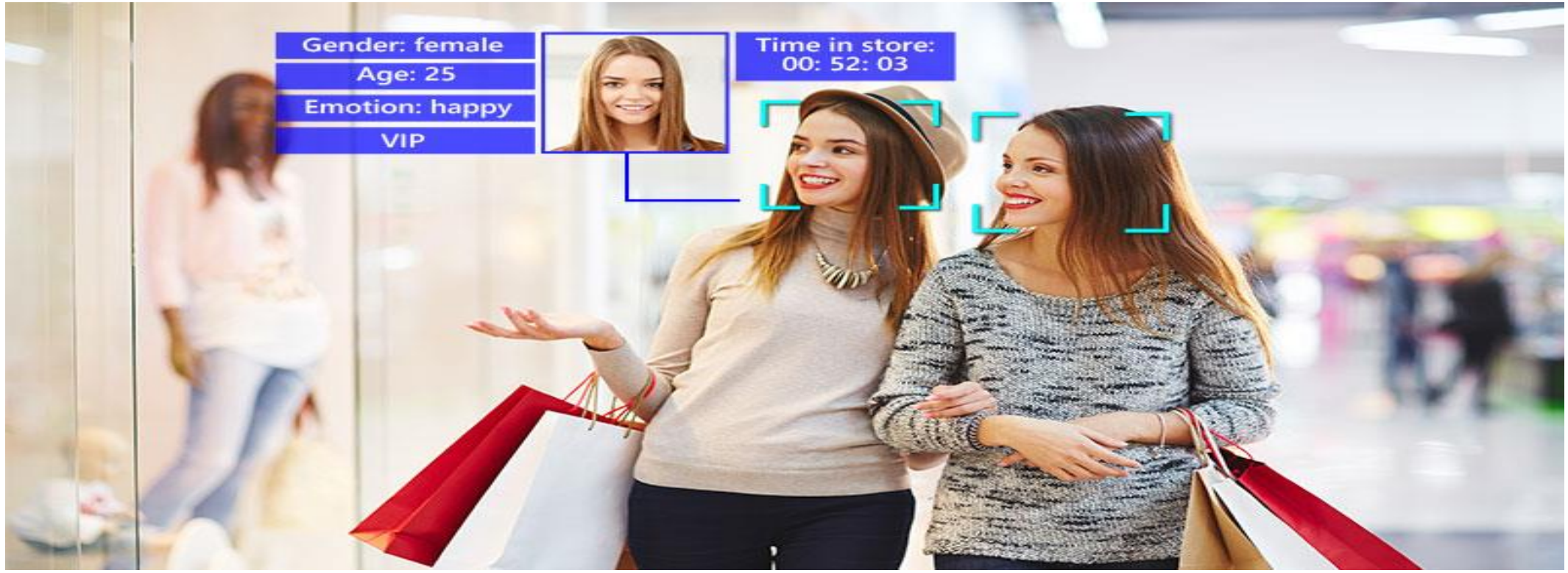
Customer Characteristics Analysis (3/5)



- Hikvision, People counting example.

Image: https://www.youtube.com/watch?v=2kx5Q6Eh49s

Customer Characteristics Analysis (4/5)



- FaceMe, Smart retail example.

Image: https://kr.cyberlink.com/faceme/insights/cases/retail

Customer Characteristics Analysis (5/5)

Customer Characteristics Analysis System

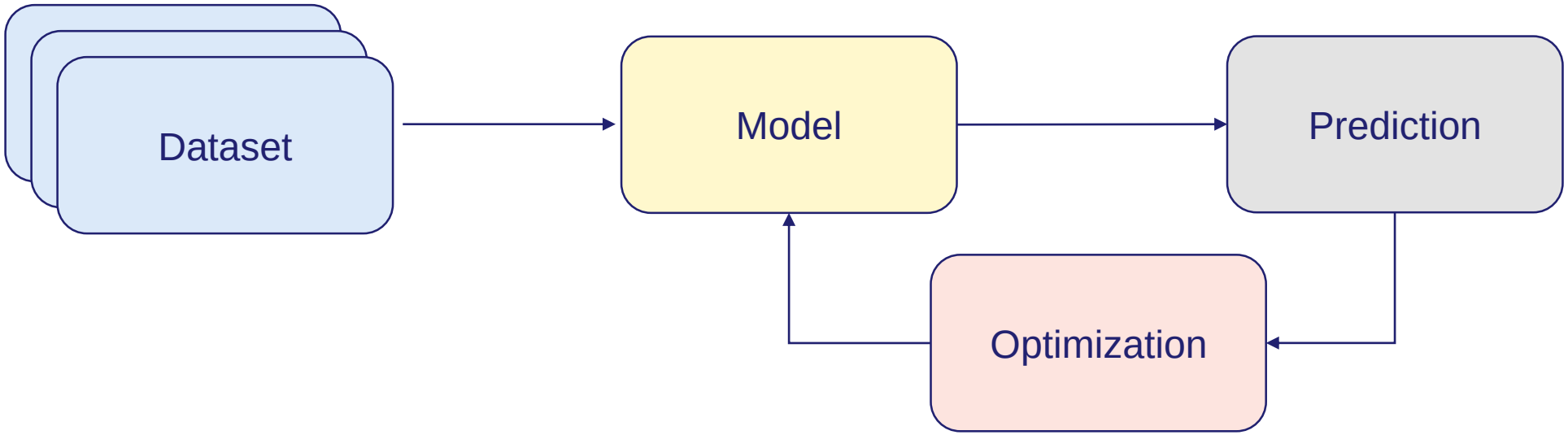
People Counting

Age/Gender/Expression … Recognition

- My final goal is to build an all-in-one system.
- This semester, I plan to implement each module.

Related works

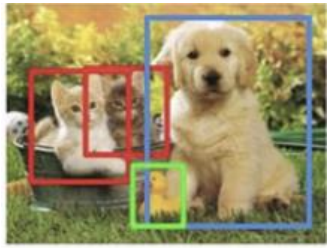
Multi-Task Learning Considerations (1/7)



- MTL is the same as a typical deep learning procedure.
- Dataset preprocessing, model design, and optimizer selection must be considered.

Multi-Task Learning Considerations (2/7)

Dataset1 (Detection)



CAT: [20, 20, 60, 60],
[30, 30, 70, 70]
DOG: [50, 50, 80, 80]
DUCK: [70, 70, 80, 80]

Dataset2 (Classification)

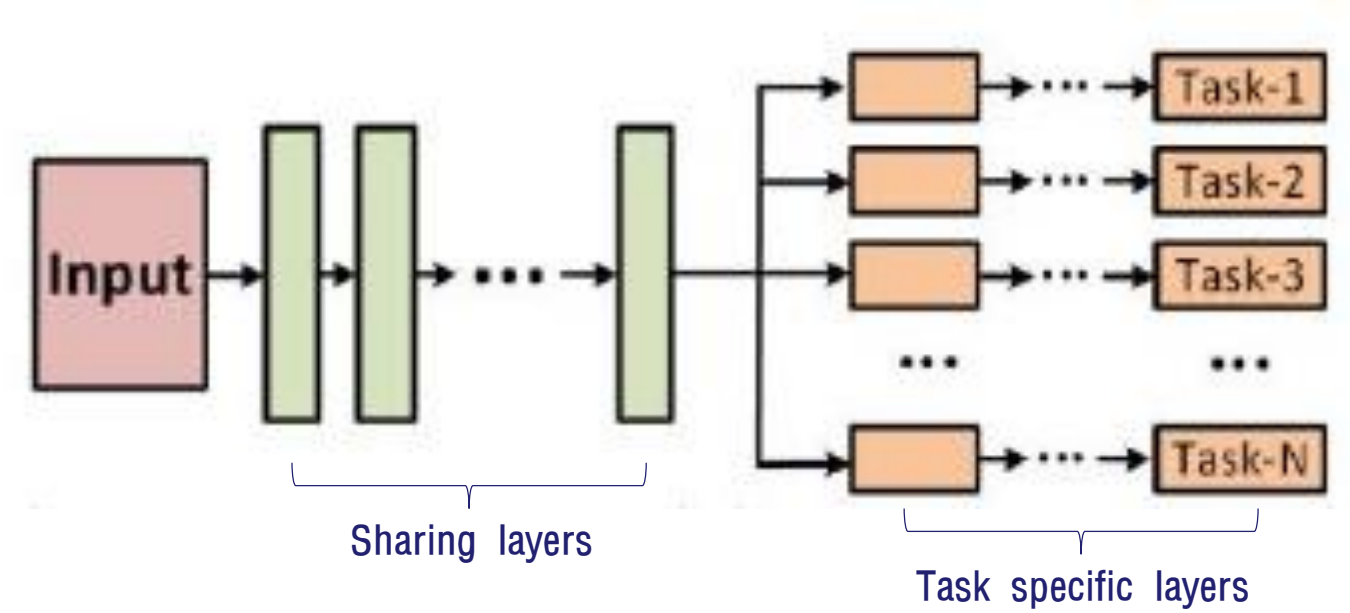
CAT: 0

- It should be considered that each dataset may have different input data or labels.
- For example, the input data for Dataset1 can be a bounding box, and the label for Dataset2 can be in discrete data format.
- In order to train various data from one model, a method of distinguishing data types is required.

Image: https://objectways.medium.com/best-practices-to-manage-semantic-segmentation-data-labeling-projects-229bd0c67a01

Multi-Task Learning Considerations (3/7)

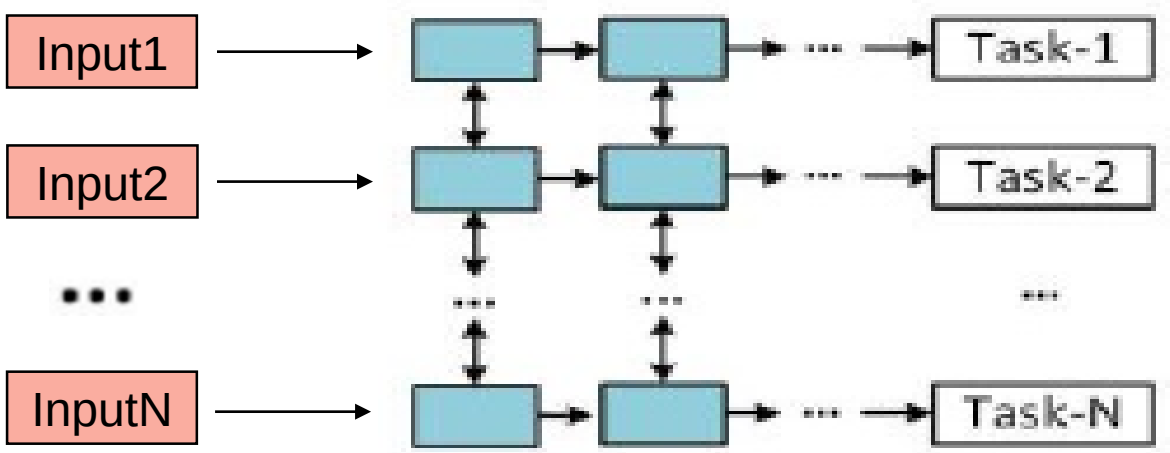
Hard parameter sharing



- Models for MTL can be built in various forms.
- This structure is divided into shared parameters and task-specific parameters.
- It has the advantage of reducing the number of parameters.

Multi-Task Learning Considerations (4/7)

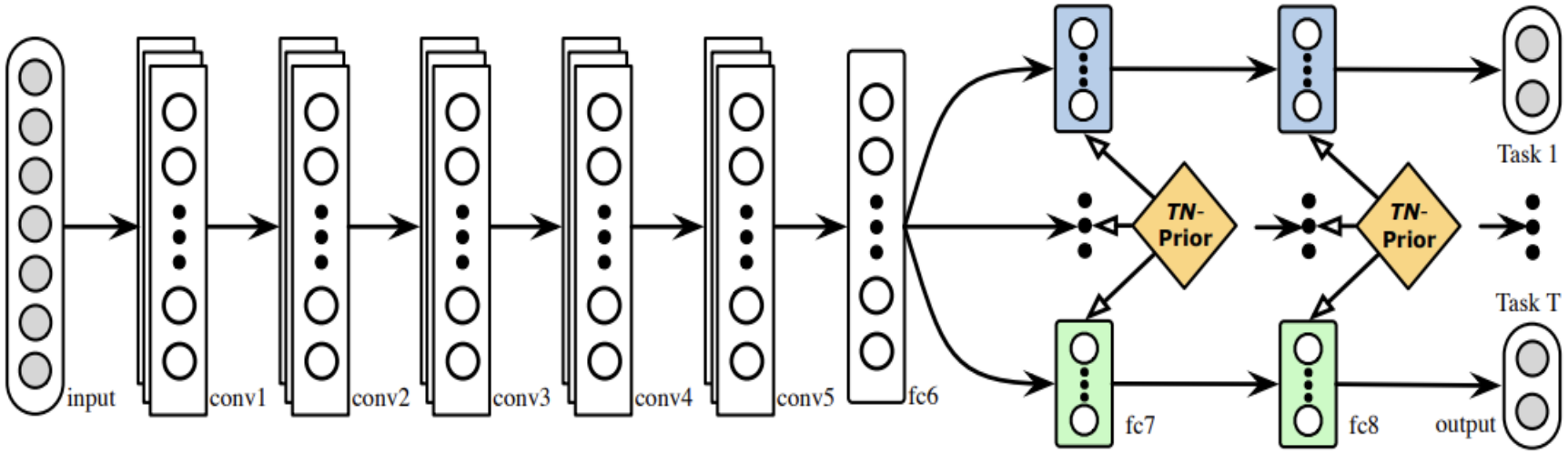
Soft parameter sharing



- This structure has a separate layer for each task.
- It gives more flexibility for the tasks by only loosely coupling the shared space representations.

Multi-Task Learning Considerations (5/7)

Deep Relation Networks
(Based on pre-trained AlexNet)

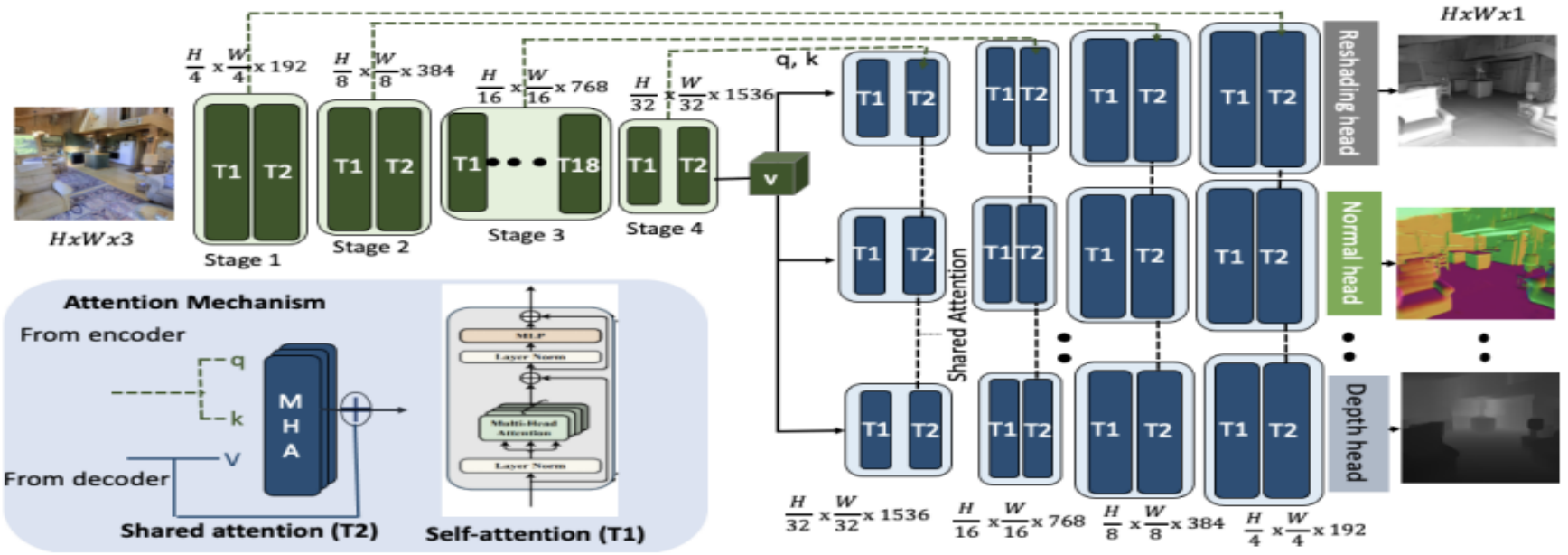


- Of course, the previous two methods can be combined, and various other models can be designed.
  (Hard, Soft parameter sharing)

M. Long, Z. Cao, J. Wang and S. Y. Philip, "Learning multiple tasks with multilinear relationship networks", *in Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, pp. 1594-1603, Long Beach, CA, USA, 2017.

Multi-Task Learning Considerations (6/7)

Multitask Learning Transformer framework, MulT
(Based on Swin transformer)

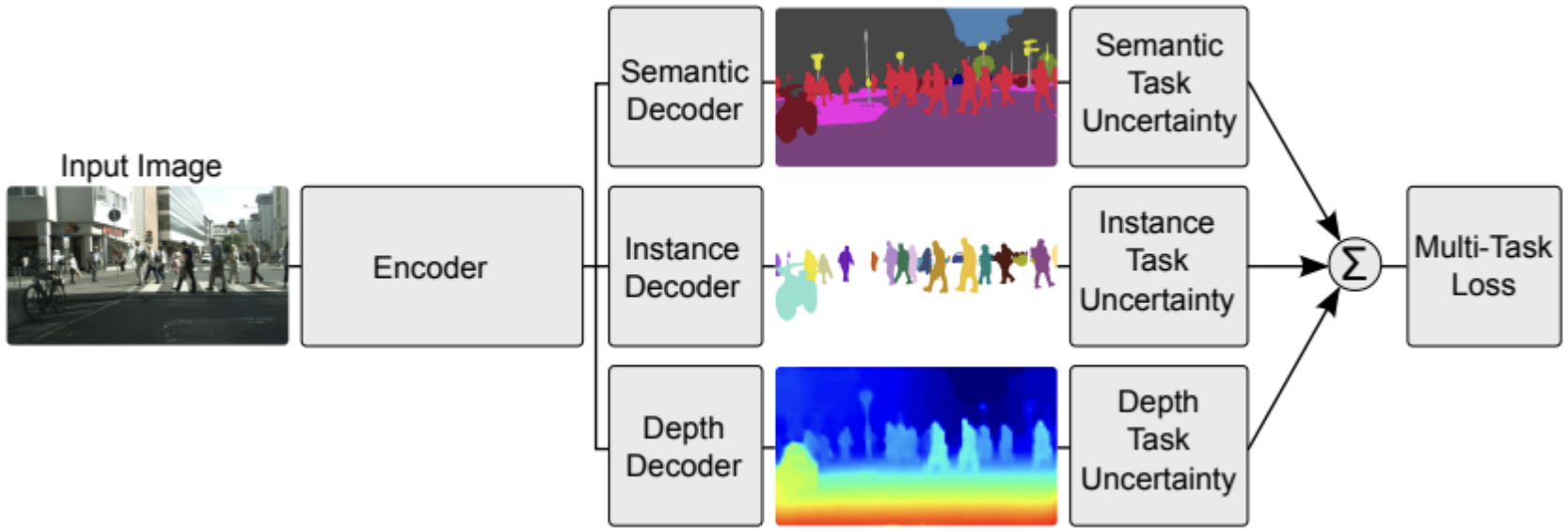


- Recently, a transformer framework has been developed that enables end-to-end MTL.
  - Encoder: Hard parameter sharing, Decoder: Soft parameter sharing

D. Bhattacharjee *et al*., "Mult: An end-to-end multitask learning transformer," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* New Orleans, Louisiana, USA, pp. 12031-12041, 2022.

Multi-Task Learning Considerations (7/7)

Loss function for MTL



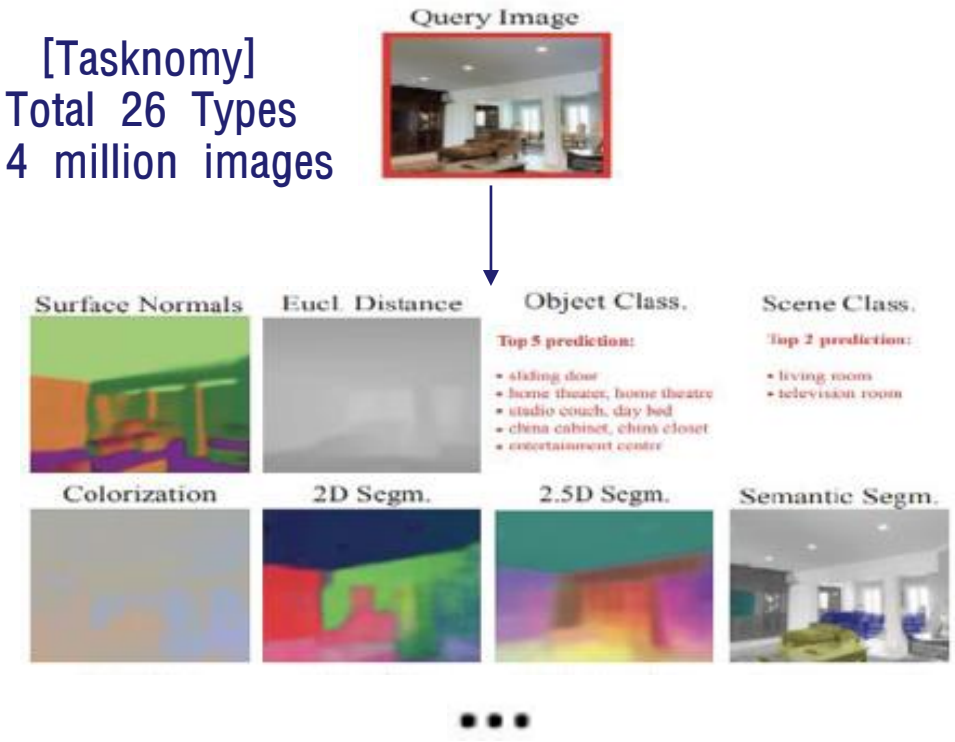
- A method of combining multiple regression and classification loss functions for MTL has been proposed.

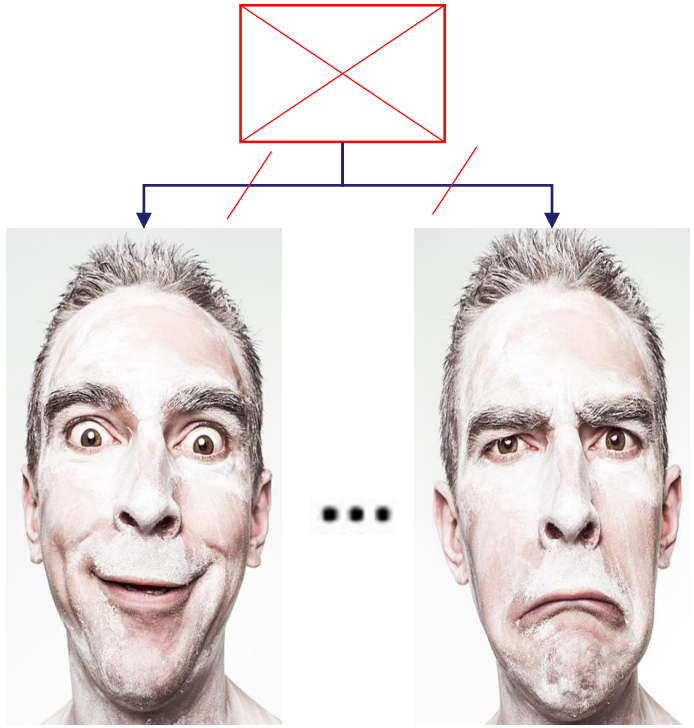
A. Kendall *et al*., "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics", *in Proc. CVPR*, Salt Lake City, Utah, USA, pp. 7482-7491, 2018.

Code Implementation (1/3)

[Tasknomy]

- Total 26 Types
- 4 million images

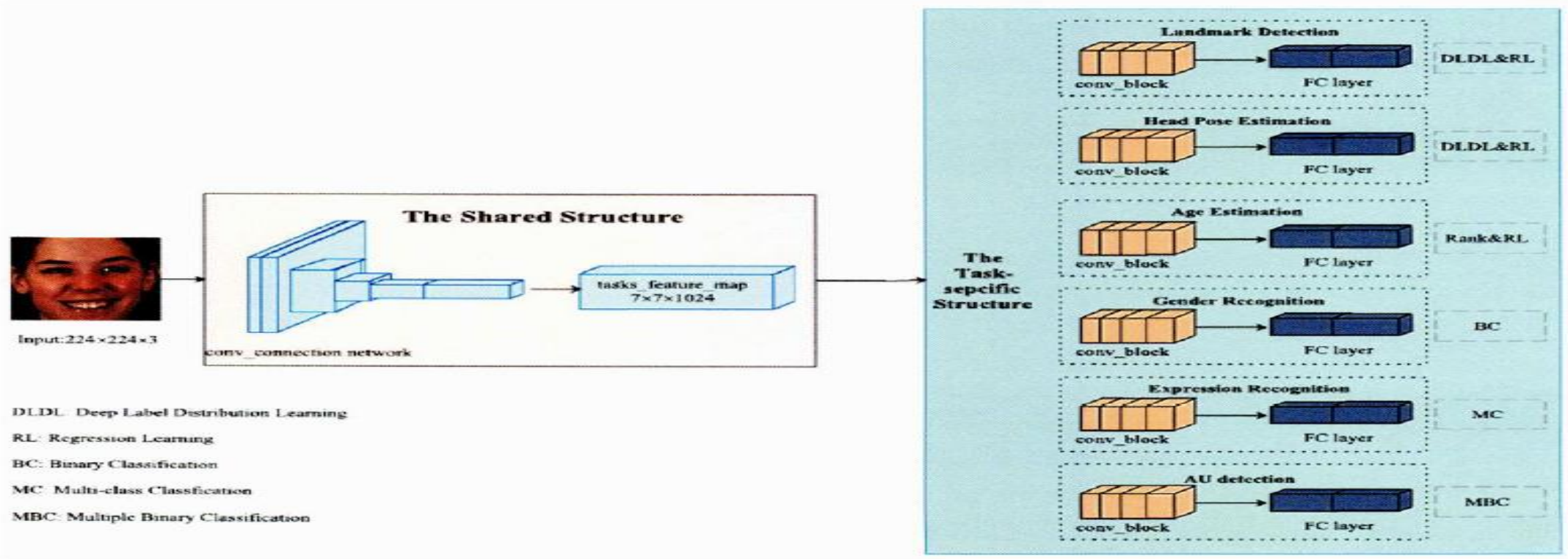




- In general, Tasknomy dataset is used for MTL experiments.
- However, there is no dataset that labels a person´s gender, age, and facial expressions at once.

A. R. Zamir *et al*., "Taskonomy: Disentangling Task Transfer Learning," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, Utah, USA, pp. 3712-3722, 2018.

Code Implementation (2/3)



- Fortunately, a paper using facial images for MTL was published this year.
  - Different task-specific datasets were used for experiments.
- I will think about how to improve performance by building CNN based model first.

Code Implementation (3/3)

Task	Dataset	# Images	Descriptions
Age/Gender	UTKFace [1]	~ 24,000	Age range from 0 to 116 years old.
	LAGENDA [2]	~ 67,000	Age range from 0 to 95 years old.
Expression	RAF-DB [3]	~ 30,000	Angry, Disgust, Fear, Happy, Sad, Surprise + Neutral
	AffectNet [4]	~ 450,000	Angry, Contempt, Disgust, Fear, Happy, Sad, Surprise + Neutral
Head pose	BIWI [5]	~ 15,000	The head pose range covers about +-75 degrees yaw and +-60 degrees pitch.
	300W-LP [6]	~ 120,000	Yaw angles in [-90, +90], Pitch and roll angles in [-99,+99]

- Data balance should be considered for experiments.

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. (MTL Model Learning First)
  1. People Counting Module
  2. Age/Gender/Expression … Recognition Module
     - Need training time

[Future Works]
- Prepare Datasets
- Model Design
- Experiments

Reference

[3page]

- Autonomous driving system image: https://www.evpost.co.kr/wp/%EC%9E%90%EC%9C%A8%EC%A3%BC%ED%96%89%EC%B0%A8%EA%B0%80-%EC%84%B8%EC%83%81%EC%9D%84-%EB%B3%B4%EB%8A%94-%EC%97%AC%EB%9F%AC%EA%B0%80%EC%A7%80-%EB%B0%A9%EB%B2%95/
- Medical science image: W. Shao, Y. Peng, C. Zu, M. Wang, and D. Zhang, “Hypergraph based multi-task feature selection for multimodal classification of Alzheimer’s disease,” *Computerized Med. Imag. Graph.*, vol. 80, Mar. 2020.
- K. Nguyen, M. Le, B. Martin et al., When AI meets store layout design: a review. *Artif. Intell. Rev*. Vol. 55, 5707–5729 (2022).

[9page]

- Full body image: https://docs.openvino.ai/2023.0/omz_models_model_person_detection_0202.html
- Face image: https://docs.openvino.ai/2023.0/omz_demos_face_detection_mtcnn_demo_python.html

[20page]

[1] Z. Zhang, Y. Song and H. Qi, "Age progression/regression by conditional adversarial autoencoder", *CVPR*, pp. 5810-5818, 2017.

[2] M. Kuprashevich, and I. Tolstykh, “MiVOLO: Multi-input Transformer for Age and Gender Estimation", *arXiv:2307.04616*, 2023, [online] Available: https://arxiv.org/abs/2307.04616

[3] S. Li, W. Deng, and J. Du, “Reliable crowdsourcing and deep localitypreserving learning for expression recognition in the wild,” *CVPR*, 2017, pp. 2584–2593.

[4] A. Mollahosseini, B. Hasani and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," in *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18-31, 1 Jan.-March 2019.

[5] G. Fanelli, T. Weise, J. Gall and L. V. Gool, "Real time head pose estimation from consumer depth cameras", *Proc. Joint Pattern Recognit. Symp.*, pp. 101-110, 2011.

[6] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, “Face alignment across large poses: A 3D solution,” *CVPR*, pp. 146–155, 2016.

Q & A

Thank you