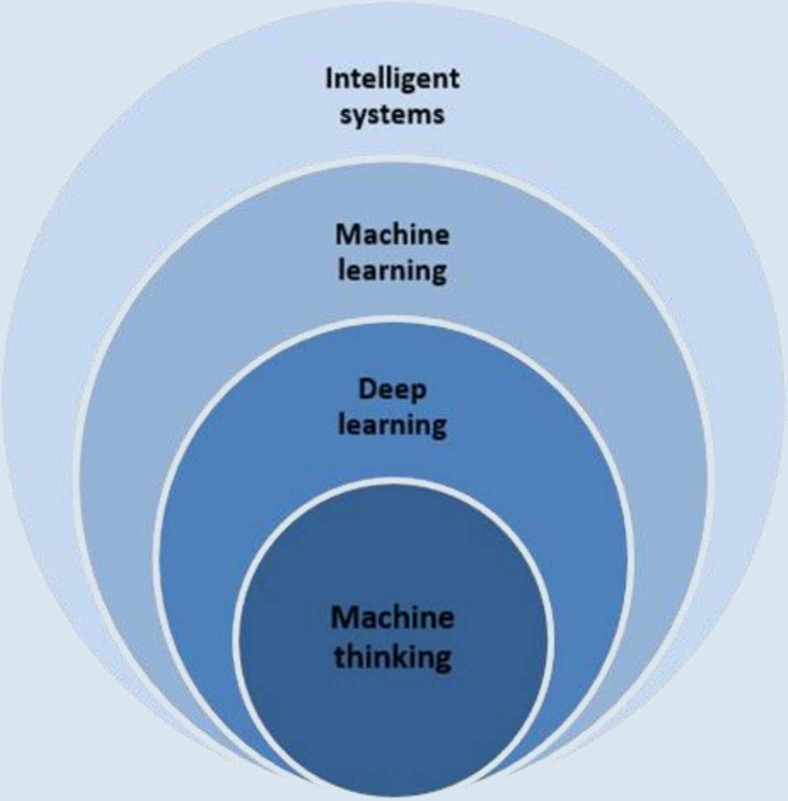
Multi-Task Learning Applications in Intelligent Systems

Benedetta Pacilli

A.Y. 2023/2024

Introduction



What is Multi-Task Learning?

- Machine Learning approach
- Multiple Diverse learning tasks
- Commonalities and differences

Why Multi-Task Learning?

- Improved learning
- Improved Efficiency

Where can Multi-Task Learning be found?

- Several fields
- Healthcare, Smart Transportation

Methodology

Research Objectives and Questions

Search Strategy

Study Selection

Quality Assessment

Data Extraction

Research Objectives and Questions

1. What are the prevalent design choices and algorithms utilized in MTL frameworks for intelligent systems?
2. What challenges and limitations are commonly encountered in the application of MTL?
3. What benefits have been observed from implementing MTL in different domains of intelligent systems?
4. How do different MTL approaches compare in terms of performance and scalability?

Search Strategy

Searched Databases:

Google Scholar
IEEE Xplore
Springer Link
ACM Digital Library

First Search Keywords:

Multi-Task Learning
Network
Applications
Design
Implementations
Deep Learning
CNN

Second Search Keywords:

Healthcare
Medical
Diagnosis
Semantic Segmentation
Smart Transportation
Autonomous Driving
Prediction
Tracking
Monitoring

Study Selection

Multiple Stages:

1. Initial Screening
2. Full-text review
3. Final Selection

Quality Assessment:

- Relevance
- Rigor
- Contribution
- Clarity

Inclusion/Exclusion Criteria:

- **Source**:
  - Journal articles; conference papers; academic publications
- **Language** and **time**:
  - English
  - 2018 - 2024
- Pertinent **content**:
  - To the MLT application topic
  - To the research questions
- **Accessibility**

Results

Q1. What are the prevalent design choices when implementing MTL?

- Parameter Sharing
- Attention Mechanism
- Task Balancing and Loss Weighting
- Adversary and Auxiliary Tasks

Parameter Sharing

Soft vs Hard Parameter Sharing

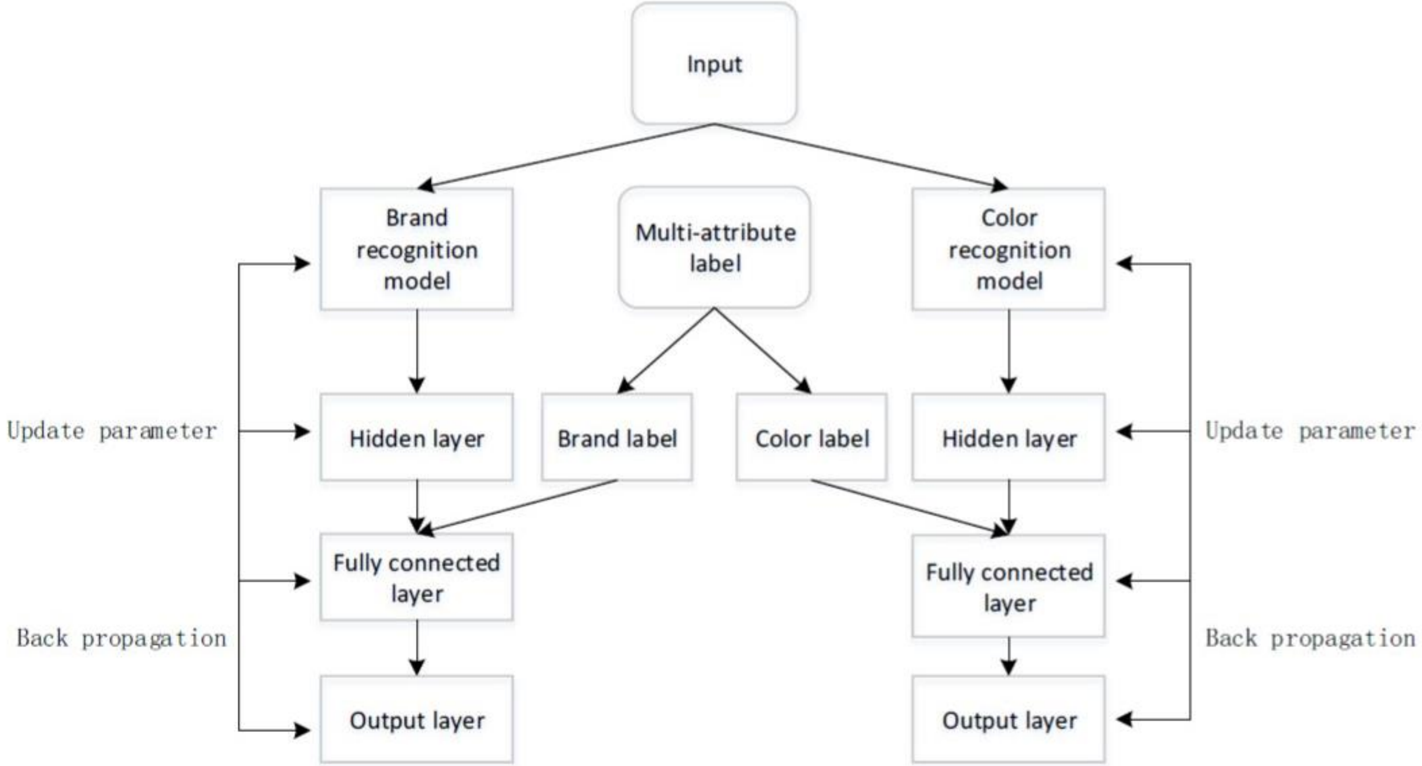
Soft parameter sharing

- Each task has its own model and parameters
- Regularization is applied to encourage the parameters of the different task-specific networks to be close to each other
- **Benefits:**
  - More flexibility
  - Better handling of tasks with different modalities or data distributions
  - Reduced Task Inference
- **Drawbacks:**
  - Increased computational cost and memory usage
  - Regularization Term

Hard parameter sharing

- The shared hidden layers learn a common representation for all tasks
- Task-specific output layers to make predictions for each task
- **Benefits:**
  - Reduced risk of overfitting
  - Decreased number of parameters
- **Drawbacks:**
  - Risk of suboptimal representations of individual tasks if they are not closely related

Soft Parameter Sharing - i

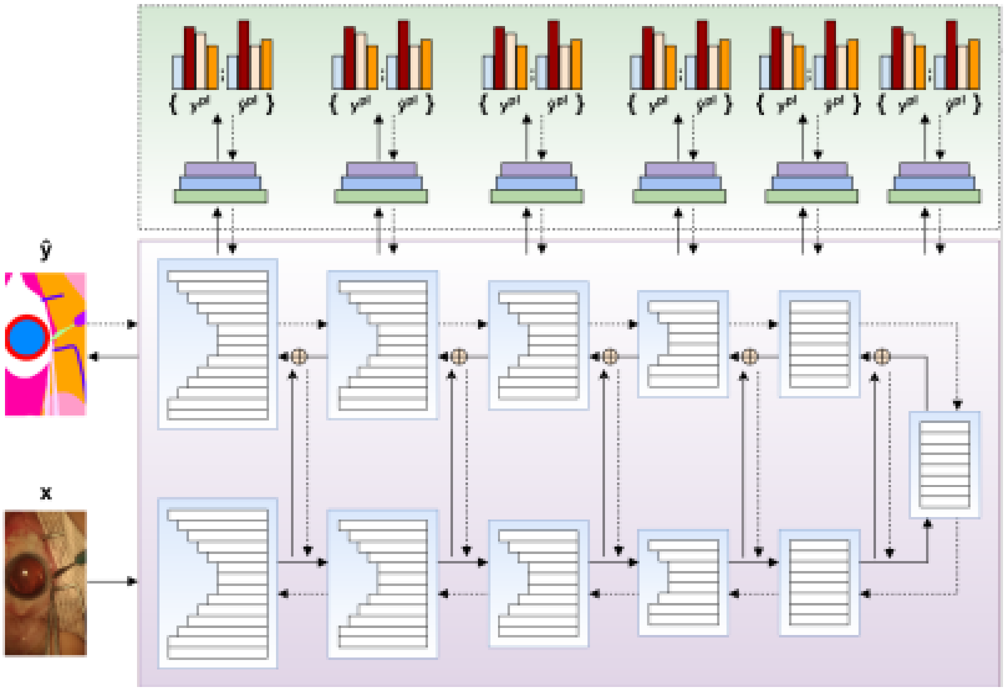


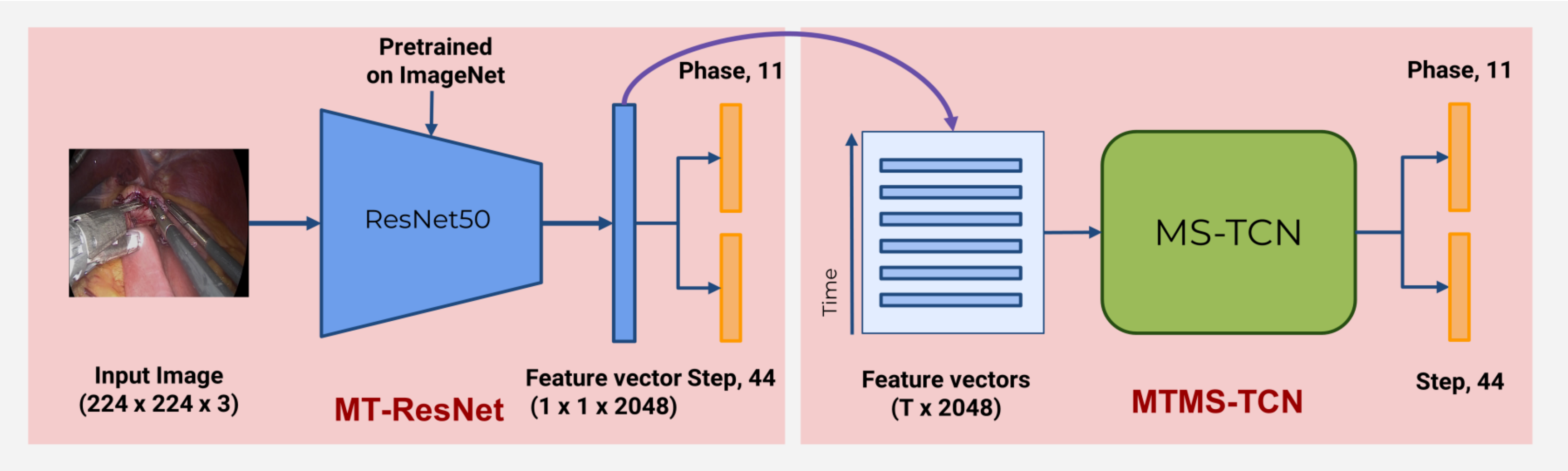
- *[Sun et al., 2019]* Improved vehicle attribute recognition
- Two tasks
  - Brand and Color recognition
- Tailored CNN model for each task
- Separate models combined into a unified MTL framework
  - Learn common features useful for both tasks
- Regularization

Soft Parameter Sharing - ii

[Bhattarai et al., 2023] Multi-task deep network for 2D surgical image semantic segmentation

- Two tasks:
  - Semantic segmentation of surgical images
  - Regression of Histogram of Oriented Gradients (HOG) features
- Shared early and intermediate layers
- Task-specific layers added on top
- Joint Optimization

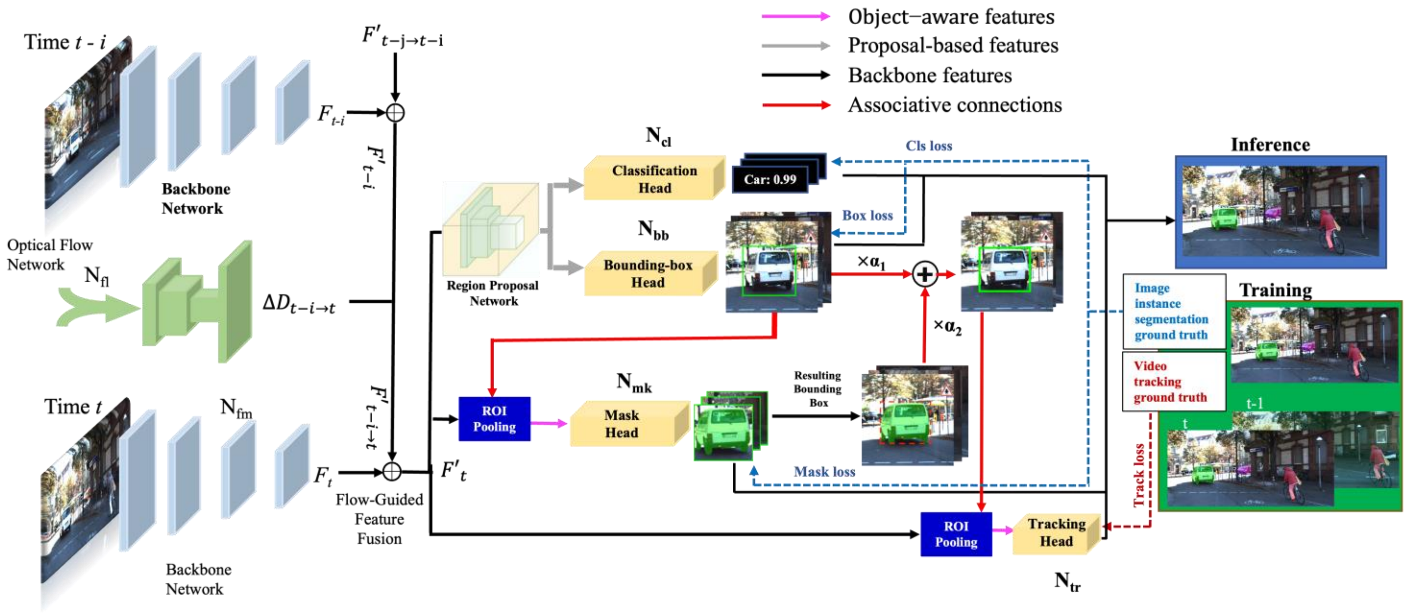




Hard Parameter Sharing

- Two tasks:
  - Joint surgical steps and phases recognition
- Shared feature extractor
- Shared feature representation
- Shared Temporal Convolutional Network (TCN) for prediction refinement
- Task-specific output heads

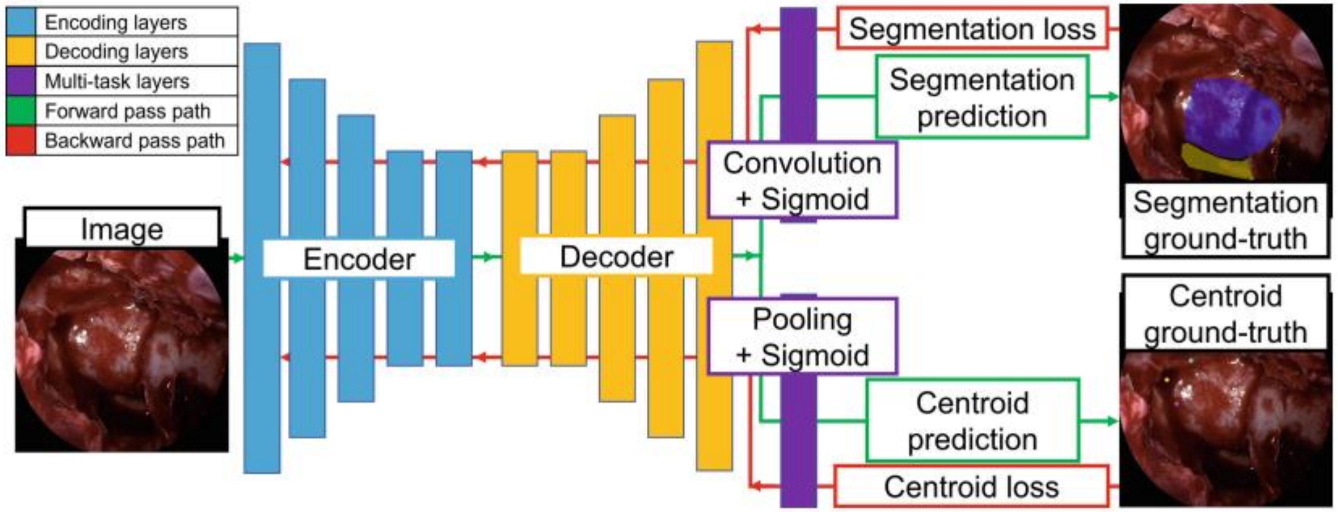
Hybrid Parameter Sharing - i



[Cui et al., 2024] Collaborative Multi-task Learning for Multi-Object Tracking and Segmentation

- Hard Parameter Influence:
  - Shared feature extraction layer acts as a common basis for all tasks
- Soft Parameter Influence:
  - Associative connections link the outputs of the detection head to the inputs of the segmentation and tracking heads

Hybrid Parameter Sharing - ii



[Das et al., 2023] A multi-task network for anatomy identification in endoscopic pituitary surgery

- Hard Parameter Influence:
  - Shared encoder
  - Common features
- Soft Parameter Influence:
  - Task-specific decoders

Attention Mechanism

Attention Mechanism

- Enhances MTL framework performance
  - Selection of the most relevant aspects
  - Focus on the important features

- Various ways to integrate attention mechanisms into MTL frameworks
- Two case studies:
  - *[Yuan et al., 2023]* Multi-task Joint Prediction of Infant Cortical Morphological and Cognitive Development
  - *[Liu et al., 2018]* End-to-end multi-task learning with attention

[Yuan et al., 2023] Attention-based Feature Disentanglement (AFD) module

- Two components:
  - **Channel attention module** highlights most relevant feature channel for age and identity-related tasks
  - **Spatial attention module** emphasizes spatial regions within the feature maps to accurately capture the patterns of cortical morphology.
- **Benefits**: separation of the mixed cortical surface features into identity-related (time-invariant) and age-related (variable) features

$$z_t^i = z_t^i \odot \phi(z_t^i) + z_t^i \odot (1 - \phi(z_t^i))$$

[Liu et al., 2018] Multi-Task Attention Network (MTAN)

- Two networks:
  - **Single Shared Network**: used to learn a shared global pool of features considered useful across all tasks
  - **Task-Specific Attention Network**: weights the shared feature through soft attention masks
- Each task-specific network has several attention modules, one for each shared network layer
- Soft attention masks determine the importance of shared features based on the task

$$\hat{a}_i^{(j)} = a_i^{(j)} \circ p^{(j)}$$

Task Balancing and Loss Weighting

Task Balancing and Loss Weighting

- *Task Balancing*: Adjusting the importance of tasks during training to ensure effective learning
- *Loss Weighting*: Weighting task-specific loss functions to manage their impact on overall loss

→ all tasks learned **effectively**

→ **prevents dominance** of one task over others

- Three case studies
  - *[Liu et al., 2018]* End-to-end multi-task learning with attention
  - *[Yuan et al., 2023]* Multi-task Joint Prediction of Infant Cortical Morphological and Cognitive Development
  - *[Ramesh et al., 2021]* Multi-task Temporal Convolutional Networks for Joint Recognition of Surgical Phases and Steps in Gastric Bypass Procedures

[Liu et al., 2018] End-to-end multi-task learning with attention

- MTAN module balances the tasks through the attention masks

$$\hat{a}_i^{(j)} = a_i^{(j)} \circ p^{(j)}$$

- A weighting scheme Dynamic Weight Average (DWA) adjusts and balances the loss contribution of each task **dynamically**

$$\lambda_k(t) = \frac{K \exp\left(\frac{w_k(t-1)}{T}\right)}{\sum_i \exp\left(\frac{w_i(t-1)}{T}\right)} \qquad w_k(t-1) = \frac{L_k(t-1)}{L_k(t-2)}$$

[Yuan et al., 2023] Attention-Based Disentanglement Module and Identity Conditional Block

- AFD disentangles the features of distinct tasks
- → ICB combines identity-related features (invariant) with age-related features (evolving)
- Different components have different loss functions

$$L(E, ID, AgeP) = \lambda_{IT} L_{IT} + \lambda_{AE} L_{AE}$$

$$L(ICB, Cog, G, D) = \lambda_{Cog} L_{Cog} + \lambda_{D} L_{D} + \lambda_{G} L_{G} + \lambda_{rec} L_{rec}$$

- Weights **empirically** set based on experimental tuning

[Ramesh et al., 2021] Multi-task Temporal Convolutional Networks for Joint Recognition of Surgical Phases and Steps

- Loss weighting is used to address the imbalance in the class distribution of phases and steps
- Class weighted cross-entropy loss
- **Benefits:** infrequent classes are given appropriate importance during training
- Weights are calculated using **median frequency balancing**

$$L_{\text{total}} = L_{\text{phase}} + L_{\text{step}}$$

Adversarial and Auxiliary Tasks

Adversarial and Auxiliary Tasks

Adversarial Tasks are designed to ensure that the shared representation learned does not capture certain unwanted features or biases

[Yuan et al., 2023]

- **Auxiliary**: Identity recognition and age estimation
  - Supports the AFD module
  - Minimizes the distance between features of the same individual at different ages
  - Maximizes the distance between features of different individuals
- **Adversarial**: a generator and a discriminator
  - Generator produces maps → Discriminator evaluates
  - Real/generated data
  - The generator improves itself → more realistic and vivid cortical morphological maps, subtle details and variances
  - Two adversarial loss functions

Auxiliary Tasks support the primary task by providing additional supervision and improving feature representation

[Bhattarai et al., 2022]

- Additional pseudo-labels created through HOGs
- Demarcate boundaries between different anatomical structures and surgical instruments
- Aids to the primary task of segmentation
- Additional supervision and robust features without manual annotations

Q2. What challenges and limitations are commonly encountered in MTL application?

Complexity

Optimization and Training

Architectural Design

Scalability

Data and Label Availability

Task Balancing

Negative Transfer

Task Relatedness and Transferability

Challenges and Solutions - i

Increased Complexity → *[Cui et al., 2023]* Associative Connection

Architectural Design → *[Yuan et al., 2023]* DITSAA
[Das et al., 2023] PAINet
[Zhang et al., 2022] pmlLSTM

Scalability → *[Liu et al., 2018]* MTAN parameter sharing
[Ramesh et al., 2021] Higher temporal resolution and larger receptive fields

Challenges and Solutions - ii

Negative Transfer → [Das et al., 2023] Network separated into sub-tasks to minimize task inference

Task Balancing → [Liu et al., 2018] MTAN and DWA to adapt task weighting depending on the loss rate of change

Challenges and Solutions - iii

Data and Label Availability → *[Bhattarai et al., 2022]* Novel approach using HOGs pseudo labels

Evaluation and Benchmarking → *[Cui et al., 2023]* Multiple valuation metrics used (MOTSA, MOTSP, sMOTSA)

Q3. What benefits have been observed from implementing MTL?

Parameter Efficiency

Robustness to Weighting Schemes

Improved Accuracy
(Mutual Benefits between Tasks)

Training Efficiency

Enhanced Generalization

Q4. How do MTL approaches compare in performance and scalability?

- *[Das et al., 2023]* PAINet outperformed single-task models by combining semantic segmentation and centroid detection tasks
- *[Zhang et al., 2022]* Co-predicting pickup and drop-off demands improved overall prediction accuracy due to inherent correlations
- *[Yuan et al., 2023]* DITSAA framework showed improved performance due to the simultaneous prediction of cognitive scores and cortical morphological maps
- *[Ramesh et al., 2021]* MTMS-TCN network improved accuracy and precision over single-task models by exploiting hierarchical relationship between phases and steps
- *[Bhattarai et al., 2022]* Integrated HOG auxiliary task resulted in performance improvements across varying training datasets, demonstrating higher generalization
- *[Liu et al., 2018]* MTAN demonstrated that the model could reach higher performance while having fewer parameters compared to traditional Single-Task models
- *[Ramesh et al., 2021]* TCN provided higher temporal resolution and larger receptive fields

Conclusion

Conclusion

- Various implementations of MTL
  - Smart Transportation
  - Healthcare
- Improvements in accuracy, efficiency, robustness, generalization, and scalability compared to Single-Task models

  - Parameter sharing
  - Attention Mechanisms
  - Shared representations
- Overcoming Challenges
  - Data scarcity, task balancing, scalability, ...

- Continuous advancement hold promises for the future

Thank you for your attention

References - i

- [1] Binod Bhattarai, Ronast Subedi, Rebati Raman Gaire, Eduard Vazquez, and Danail Stoyanov. Histogram of oriented gradients meet deep learning: A novel multi-task deep network for medical image semantic segmentation, 2022
- [2] Yiming Cui, Cheng Han, and Dongfang Liu. Cml-mots: Collaborative multi-task learning for multi-object tracking and segmentation, 2023
- [3] Adrito Das, Danyal Z. Khan, Simon C. Williams, John G. Hanrahan, Anouk Borg, Neil L. Dorward, Sophia Bano, Hani J. Marcus, and Danail Stoyanov. A multi-task network fornbsp;anatomy identification innbsp;endoscopic pituitary surgery. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2023: 26th International Conference, Vancouver, BC, Canada, October 8–12, 2023, Proceedings, Part IX, page 472–482, Berlin, Heidelberg, 2023. Springer-Verlag
- [4] Shikun Liu, Edward Johns, and Andrew J. Davison. End-to-end multi-task learning with attention. CoRR, abs/1803.10704, 2018

References - ii

- [5] Sanat Ramesh, Diego Dall'Alba, Cristians Gonzalez, Tong Yu, Pietro Mascagni, Didier Mutter, Jacques Marescaux, Paolo Fiorini, and Nicolas Padoy. Multi-task temporal convolutional networks for joint recognition of surgical phases and steps in gastric bypass procedures. CoRR, abs/2102.12218, 2021
- [7] Jingying Sun, Chengzhe Jia, and Zhiguo Shi. Vehicle attribute recognition algorithm based on multi-task learning. In 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), pages 135–141, 2019
- [8] Xinrui Yuan, Jiale Cheng, Fenqiang Zhao, Zhengwang Wu, Li Wang, Weili Lin, Yu Zhang, and Gang Li. Multi-task joint prediction of infant cortical morphological and cognitive development. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2023: 26th International Conference, Vancouver, BC, Canada, October 8–12, 2023, Proceedings, Part IX, page 545–554, Berlin, Heidelberg, 2023. Springer-Verlag
- [9] Chizhan Zhang, Fenghua Zhu, Xiao Wang, Leilei Sun, Haina Tang, and Yisheng Lv. Taxi demand prediction using parallel multi-task learning model. IEEE Transactions on Intelligent Transportation Systems, 23(2):794–803, 2022