Cross-Task Attention Network: Improving Multi-Task Learning for Medical Imaging Applications

Sangwook Kim[1], **Thomas G. Purdie**[1,2,6], and **Chris McIntosh**[1,2,3,4,5]

[1]Department of Medical Biophysics, University of Toronto, Toronto, Canada
[2]Princess Margaret Cancer Centre, Princess Margaret Research Institute, University Health Network (UHN), Toronto, Canada
[3]Toronto General Research Institute, Peter Munk Cardiac Centre, UHN, Toronto, Canada
[4]Department of Medical Imaging, University of Toronto, Toronto, Canada
[5]Vector Institute, Toronto, Canada
[6]Department of Radiation Oncology, University of Toronto, Toronto, Canada

Highlights

1. Proposed a novel multi-task learning architecture using cross task-attention modules: Cross-Task Attention Network (CTAN)
2. CTAN improved overall performance by 4.67% for eight tasks using four different medical imaging datasets
3. CTAN outperformed single-task learning and two widely used multi-task learning networks: hard-parameter sharing and multi-task attention network.

Introduction

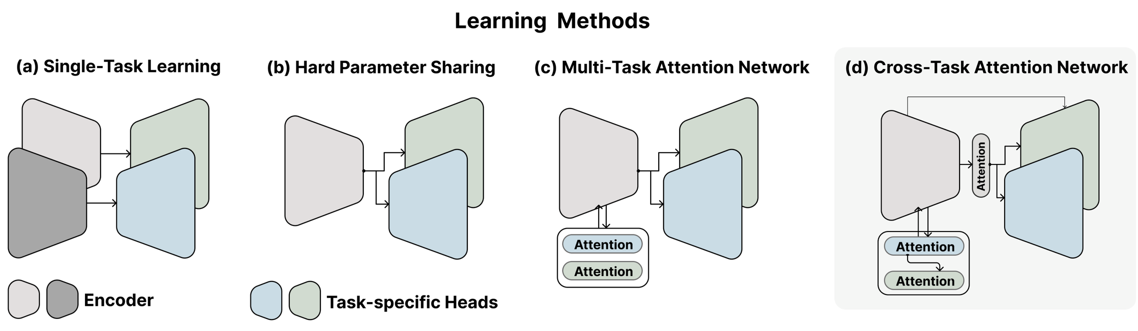

Figure 1. Comparison with single-task learning (STL), hard-parameter sharing (HPS), multi-task attention network (MTAN), and our proposed cross-task attention network (CTAN).

1. Multi-task learning (MTL) is utilized to **solve two or more tasks simultaneously using shared parameters to encourage beneficial cooperation.**
2. Most MTL architectures are based on hard-parameter sharing (HPS) which maximizes encoder regularization between tasks but **limits all tasks to an identical feature set as opposed to some common features.**
3. Multi-task attention network (MTAN)[2] uses one-to-many mapping and task-specific attention mechanisms to modify embeddings for each task, **but it cannot share information across tasks.**
4. Thus, we focus on solving MTL in **hybrid scenarios including both pixel and image-level tasks by utilizing cross-task attention in MTL using medical imaging datasets.**

Datasets

1. **Two Pixel-level tasks**: Dose prediction and segmentation of organs at risk (OAR) and clinical target volume (CTV) for prostate (Prostate) and head and neck cancer treatment (OpenKBP)[1].
2. **One image-level and one pixel-level task**: dermatoscopic images of pigmented skin lesion datasets (HAM10000)[4] is used to segment and diagnose skin lesions.
3. **Two image-level tasks**: Classification of COVID-19 and disease severity using chest CT scans (STOIC)[3].

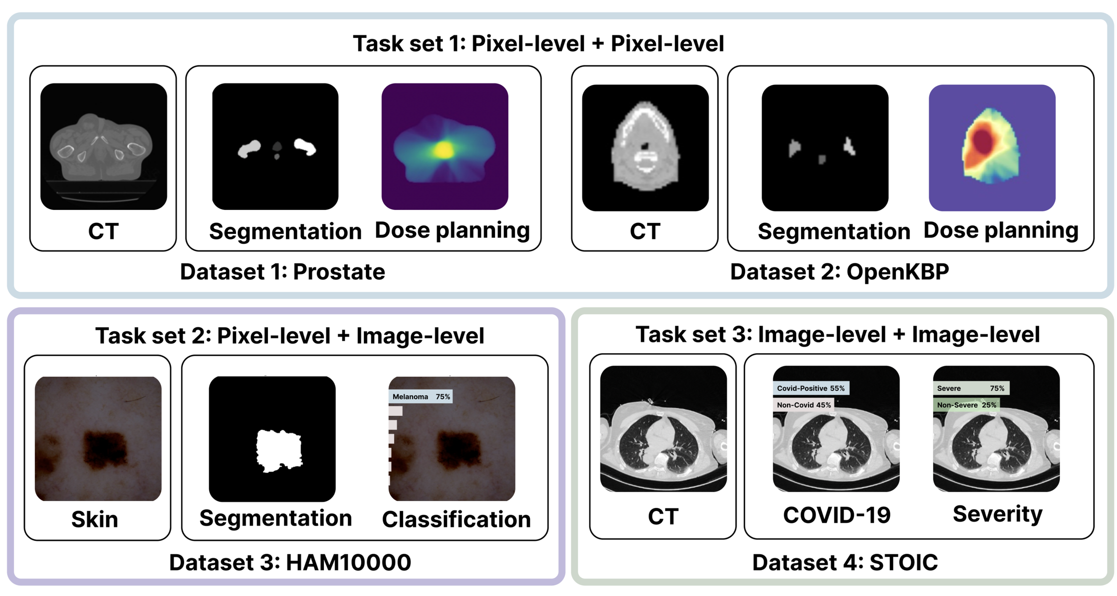

Figure 2. Graphical summary of datasets categorized into three different task groups.

Cross-Task Attention Network

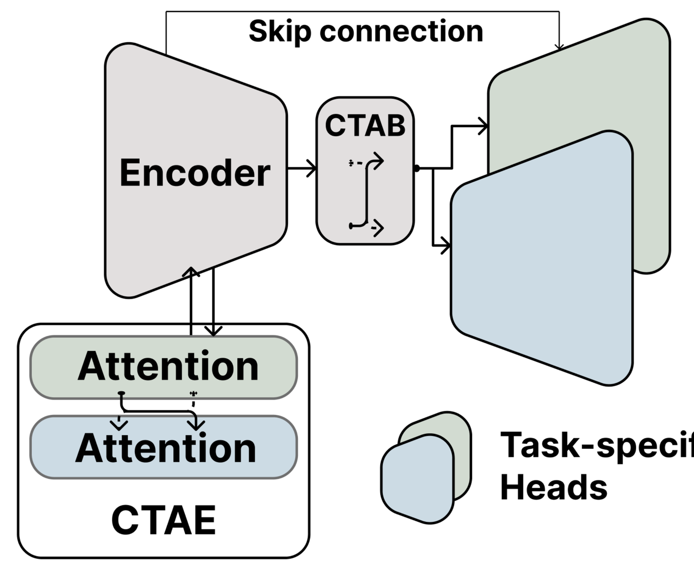

Figure 3. Graphical abstract of Cross-task attention network.

1. **Cross-Task Attention Encoder (CTAE)**: Task interactions within encoder using squeeze-and-excitation style attention block to maximize both understanding of input images and transfer of necessary information across tasks.
2. **Cross-Task Attention Bottleneck (CTAB)**: Task interactions in the bottleneck using scaled dot product attention between two latent representation.

Results

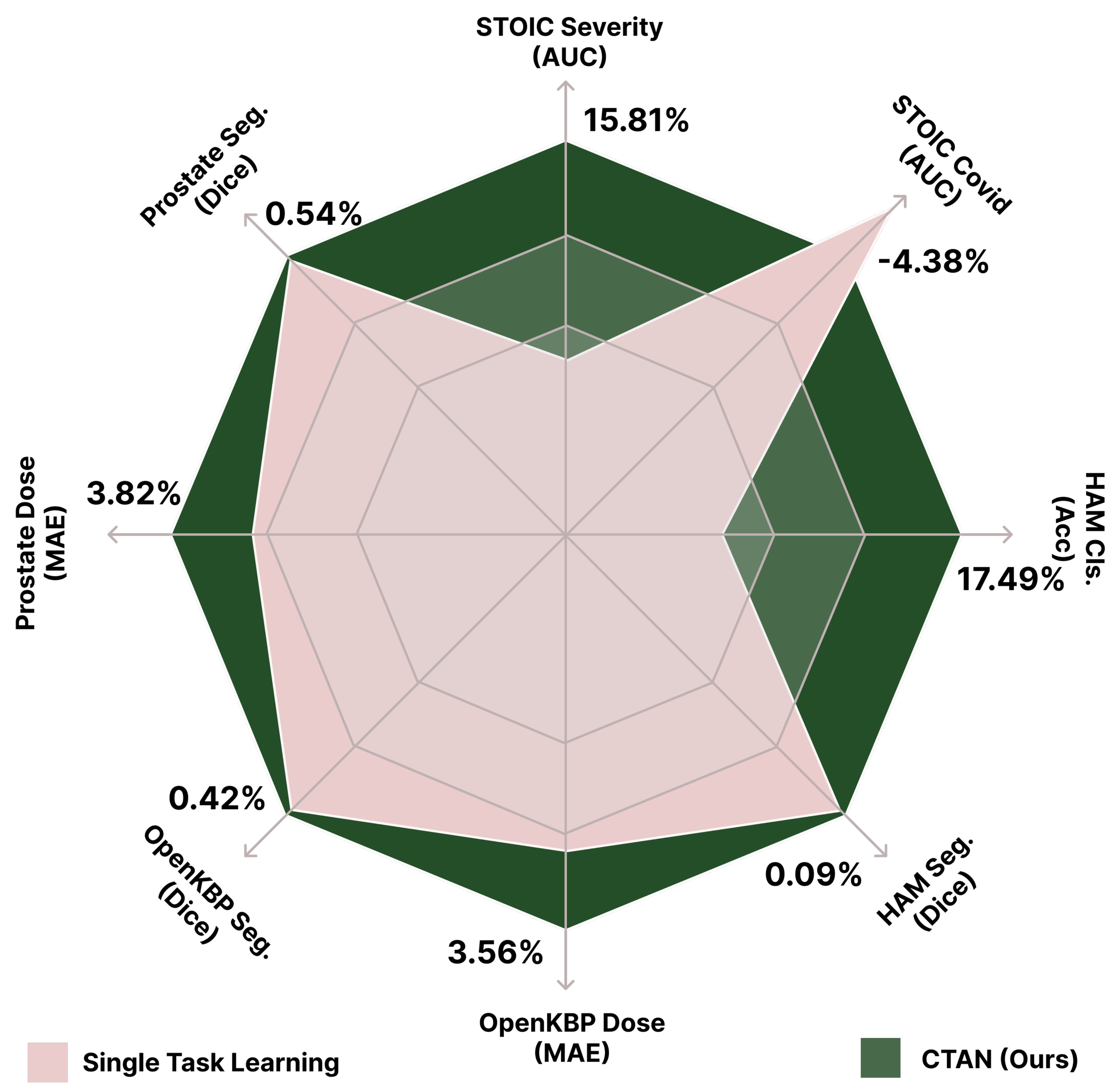

Figure 4. Relative performance increase (%) of Cross-task attention network (CTAN) compared to single-task learning for each task. Seg. refers to segmentation task, Cls. refers to classification task, and HAM refers to HAM10000 dataset. Evaluation metrics for each task are in parentheses below each task.

1. **CTAN improved performance of seven out of eight tasks** regardless of the types of tasks from pixel-level to image-level.
2. For radiotherapy datasets, training segmentation task with dose prediction task can improve dose prediction task by **3.82% and 3.56% increase** for Prostate and OpenKBP dataset, respectively.
3. **Classification performance of melanoma and skin diseases was greatly improved by 17.49%**, when simultaneously trained with segmentation task for contours of pigmented skin lesion.
4. However, **COVID classification performance was decreased by 4.38% when trained with classify severity of patients**, which shows task collision between two tasks (**negative transfer**).

Evaluation

1. Relative performance difference
   We measure the relative performance difference between baseline single-task learning and multi-task learning methods: which are hard-parameter sharing, multi-task attention network, and CTAN.

$$\Delta_{task}(\%) = 100 * \frac{(-1)^{l_i}(M_{b,i} - M_{m,i})}{M_{b,i}},\ l \in \{0, 1\}, \quad (1)$$

   i: task index b: baseline (single-task learning)
   m: multi-task learning, l: metric-specific flag (1 if the metric is higher the better)
2. Task-specific metrics

Table 1. We use task-specific evaluation metrics for each task.

Task	Task-specific metrics
Segmentation	Dice Similarity Coefficient (%)
Dose prediction	Mean absolute error (Gy)
Classification (HAM10000)	Accuracy (%)
Classification (STOIC)	Area Under the Receiver Operating Characteristic (ROC) Curve

Implementation

Table 2. Summary of loss functions for each task. We use Adam with the learning rate of 10^{-4} and the weight decay of 10^{-5}. We use task-specific losses. We utilized Dynamic Weight Averaging [2] to stabilize the combined training losses of all tasks.

Task	Loss function	Dataset
Segmentation	Combo Loss	Prostate, OpenKBP, HAM10000
Dose prediction	Mean absolute error Loss	Prostate, OpenKBP
Classification	Cross-entropy Loss	HAM10000, STOIC

Acknowledgement

The *Cross-Task Attention Network: Improving Multi-Task Learning for Medical Imaging Applications* is supported by the ***Data Sciences Institute, University of Toronto***

References

[1] Aaron Babier, Binghao Zhang, Rafid Mahmood, Kevin L. Moore, Thomas G. Purdie, Andrea L. McNiven, and Timothy C. Y. Chan.
OpenKBP: The open-access knowledge-based planning grand challenge and dataset.
Medical Physics, 48(9):5549–5561, 2021.

[2] Shikun Liu, Edward Johns, and Andrew J Davison.
End-to-end multi-task learning with attention.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1871–1880, 2019.

[3] Marie-Pierre Revel, Samia Boussouar, Constance de Margerie-Mellon, Inès Saab, Thibaut Lapotre, Dominique Mompoint, Guillaume Chassagnon, Audrey Milon, Mathieu Lederlin, Souhail Bennani, et al.
Study of thoracic ct in covid-19: the stoic project.
Radiology, 301(1):E361–E370, 2021.

[4] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler.
The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions.
Scientific data, 5(1):1–9, 2018.

LinkedIn: linkedin.com/in/swkim95
sangwook.kim@mail.utoronto.ca