NEURAL INFORMATION PROCESSING SYSTEMS

BiSPL KAIST

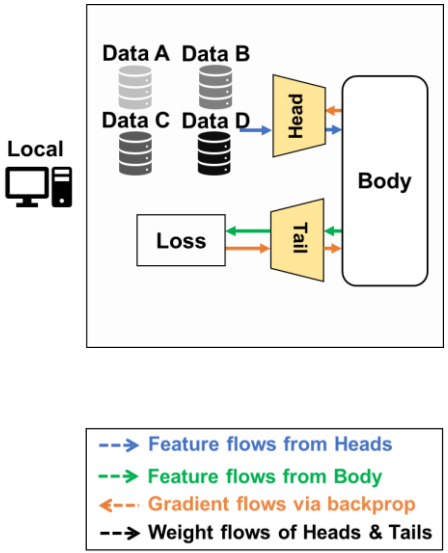
Federated Split Task-Agnostic Vison Transformer for COVID-19 CXR Diagnosis

Sangjoon Park*, Gwanghyun Kim*, Jeongsol Kim, Boah Kim, Jong Chul Ye

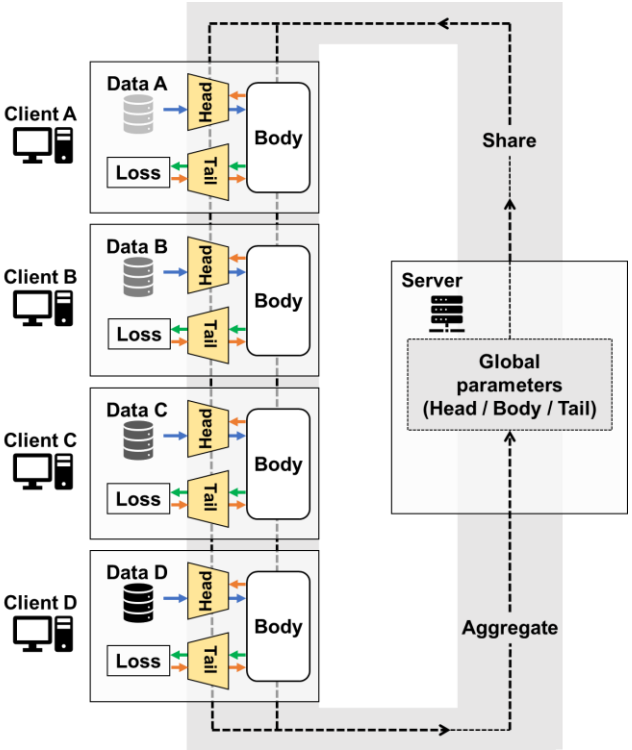
BISPL - BioImaging, Signal Processing, and Learning Lab.
Dept. of Bio and Brain Engineering
KAIST, Korea

Federated learning & Split learning

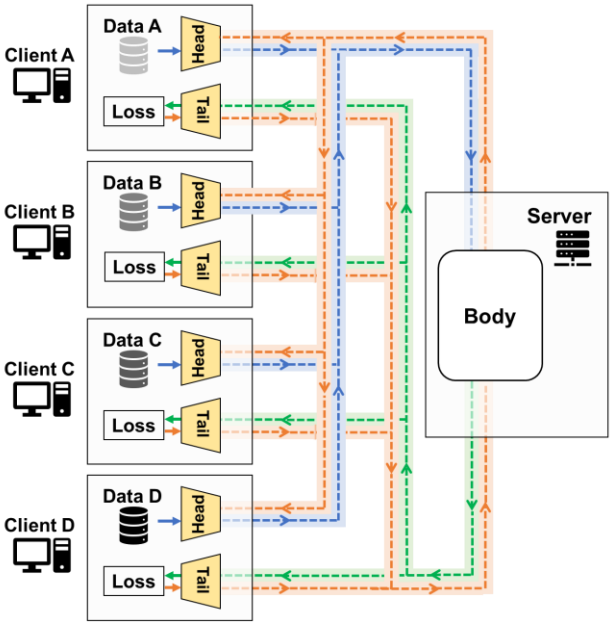
Data-centralized



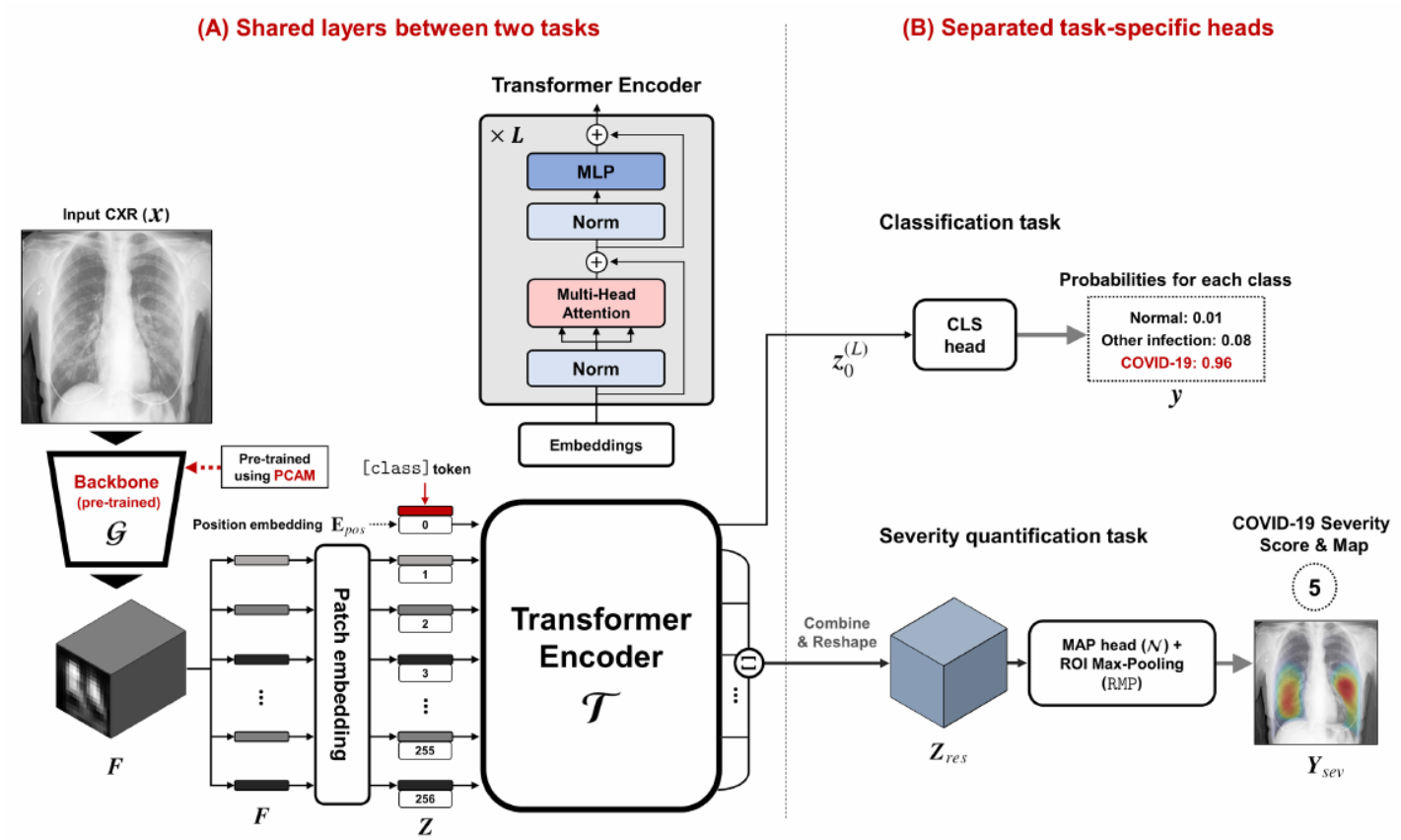
Federated Learning (FL)



Split Learning (SL)



Vision Transformer for COVID-19 Diagnosis



Park and Kim et al. under review, 2021

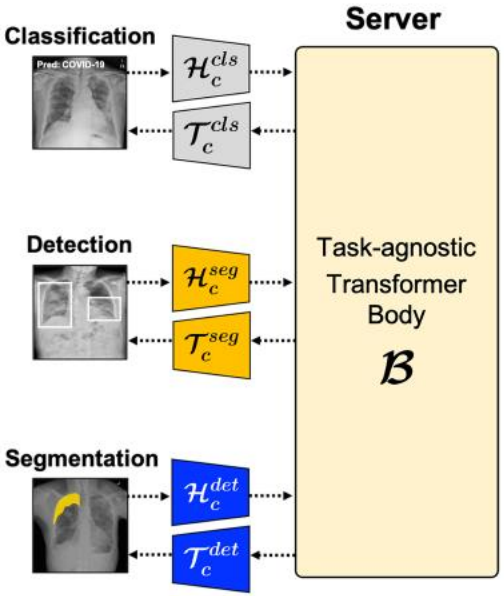
Federated Split Task-Agnostic Vison Transformer

Here we propose a novel Federated Split Task-Agnostic (FeSTA) framework equipped with Vision Transformer (ViT) to simultaneously process multiple CXR tasks including diagnosis of COVID-19.

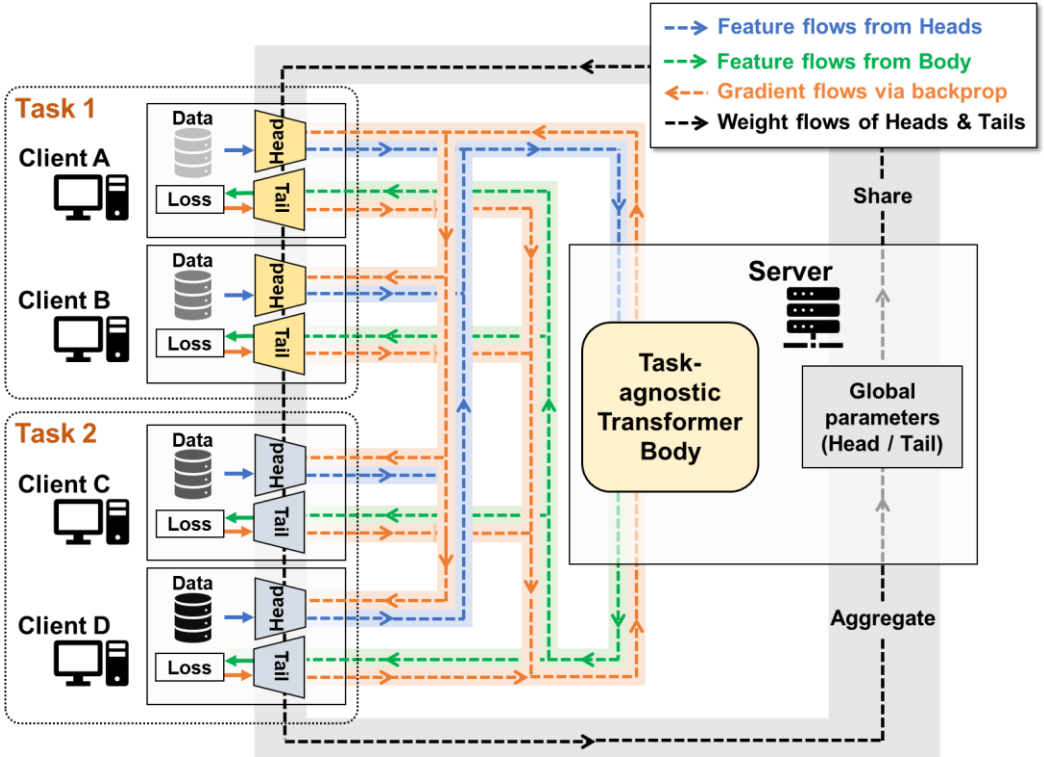
- Feasibility of FeSTA (amalgamate pros of FL and SL) method
- Benefit of Multi-task learning (MTL) with ViT and FeSTA

Federated Split Task-Agnostic Vison Transformer

Split Task-Agnostic Transformer



FeSTA framework



Detailed Procedure of FeSTA

Function `ServerMain`:

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow \texttt{ClientHead}(c)$

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow \texttt{ClientTail}(c, b_c^{(i)})$ & Backprop.

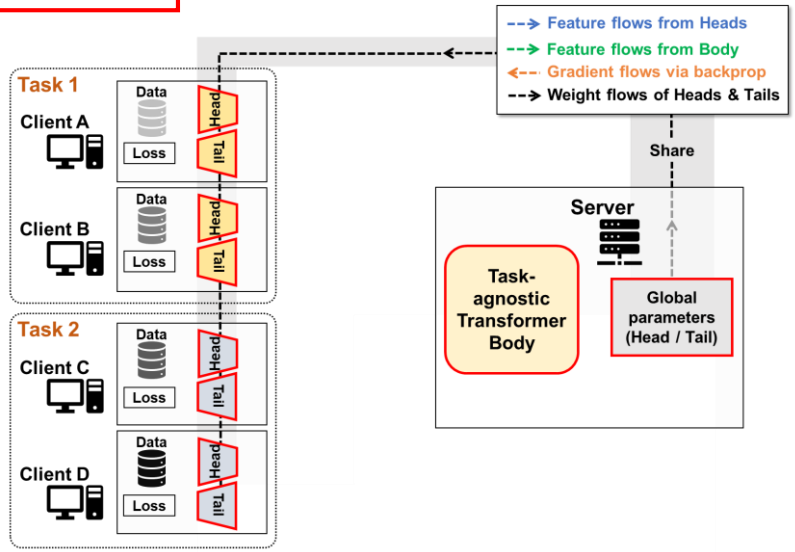
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow \texttt{ClientUpdate}(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Detailed Procedure of FeSTA

Function `ServerMain`**:**

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow$ `ClientHead`(c)

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow$ `ClientTail`$(c, b_c^{(i)})$ & Backprop.

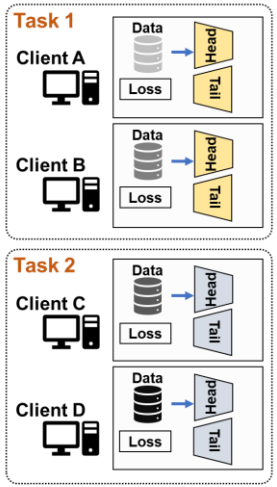
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow$ `ClientUpdate`$(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

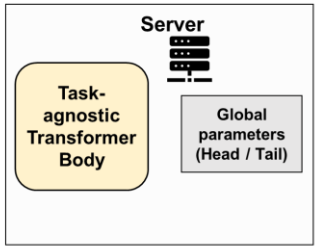
  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$





Function `ClientHead`(c)**:**

  $x_c \leftarrow$ Current batch of input from client c

  return $\mathcal{H}_c(x_c)$

Detailed Procedure of FeSTA

Function `ServerMain`**:**

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow \texttt{ClientHead}(c)$

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow \texttt{ClientTail}(c, b_c^{(i)})$ & Backprop.

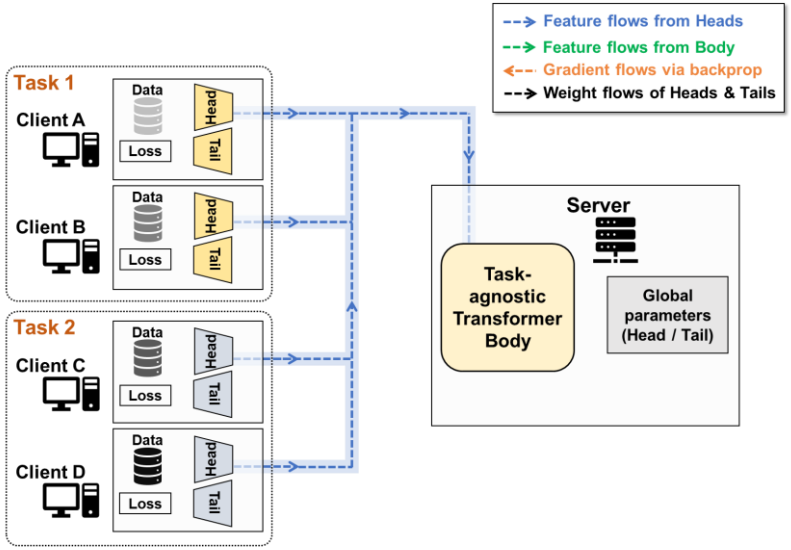
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow \texttt{ClientUpdate}(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Function `ClientHead`(c)**:**

  $x_c \leftarrow$ Current batch of input from client c

  return $\mathcal{H}_c(x_c)$

Detailed Procedure of FeSTA

Function `ServerMain`:

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow \texttt{ClientHead}(c)$

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow \texttt{ClientTail}(c, b_c^{(i)})$ & Backprop.

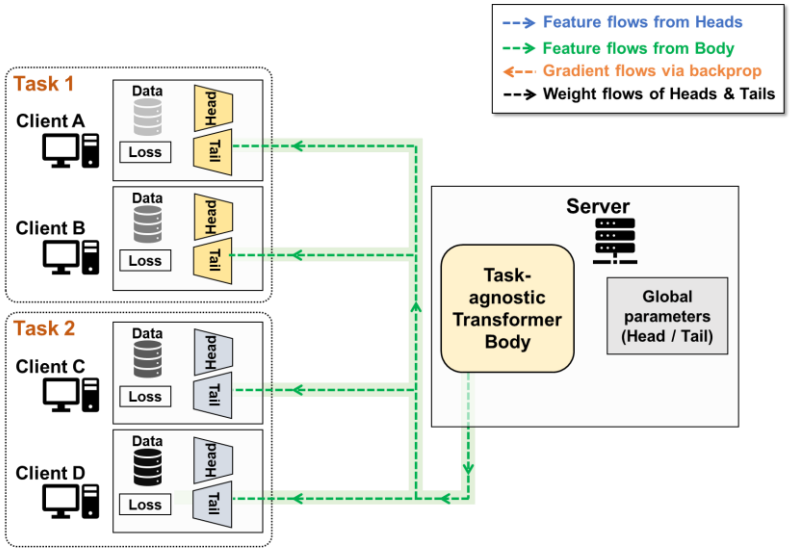
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow \texttt{ClientUpdate}(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Detailed Procedure of FeSTA

Function `ServerMain`**:**

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow$ `ClientHead`(c)

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow$ `ClientTail`$(c, b_c^{(i)})$ & Backprop.

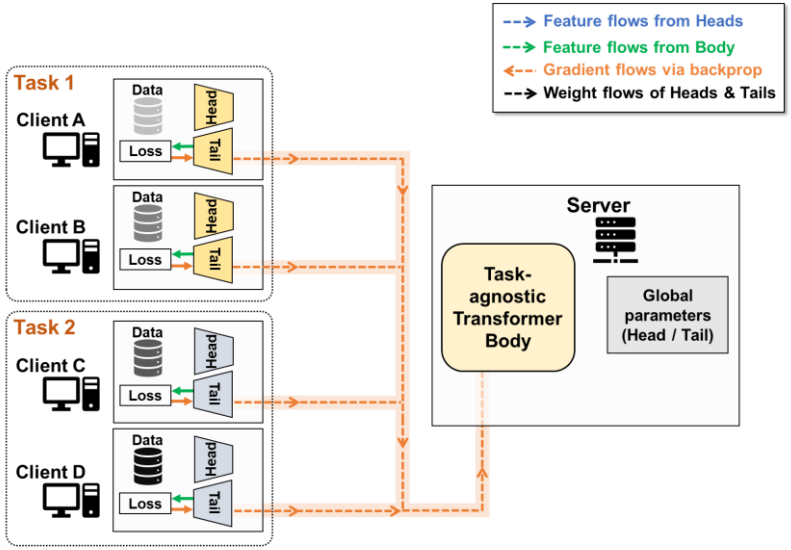
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow$ `ClientUpdate`$(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Function `ClientTail`(c, b_c)**:**

  $y_c \leftarrow$ Current batch of label from client c

  $L_c \leftarrow \ell_c(y_c, \mathcal{T}_c(b_c))$ & Backprop.

  return $\frac{\partial L_c}{\partial b_c}$

Detailed Procedure of FeSTA

Function `ServerMain`**:**

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow$ `ClientHead`(c)

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow$ `ClientTail`$(c, b_c^{(i)})$ & Backprop.

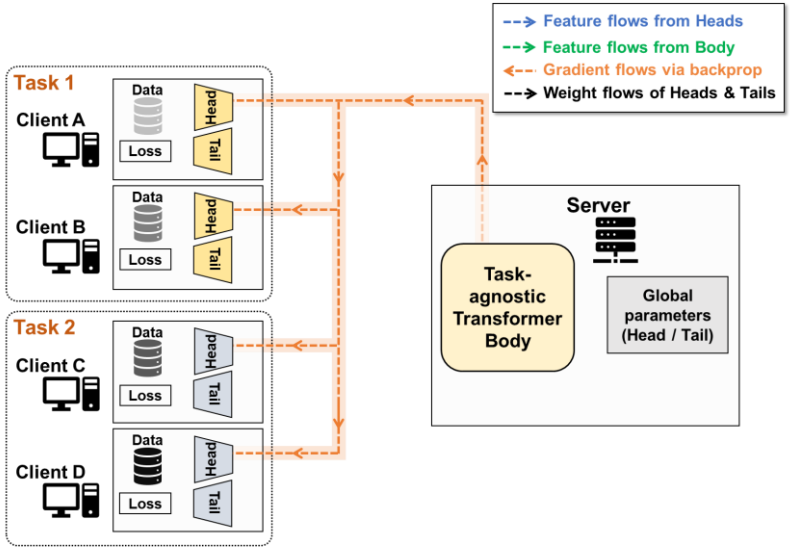
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow$ `ClientUpdate`$(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Function `ClientUpdate`$(c, \frac{\partial L_c}{\partial h_c})$**:**

  Backprop. & $(w_{\mathcal{H}_c}, w_{\mathcal{T}_c}) \leftarrow (w_{\mathcal{H}_c} - \eta \frac{\partial L_c}{\partial w_{\mathcal{H}_c}}, w_{\mathcal{T}_c} - \eta \frac{\partial L_c}{\partial w_{\mathcal{T}_c}})$

  return $(w_{\mathcal{H}_c}, w_{\mathcal{T}_c})$

Detailed Procedure of FeSTA

Function `ServerMain`**:**

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow$ `ClientHead`(c)

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow$ `ClientTail`$(c, b_c^{(i)})$ & Backprop.

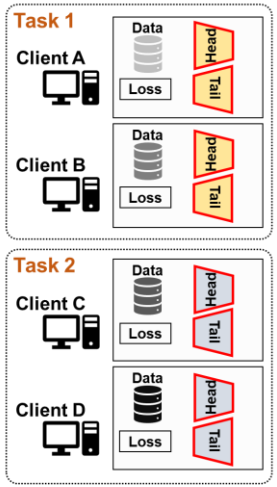
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow$ `ClientUpdate`$(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

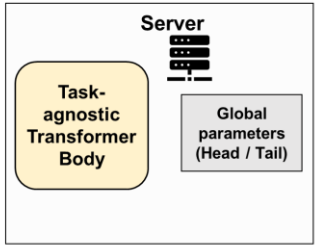
  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$





Function `ClientUpdate`$(c, \frac{\partial L_c}{\partial h_c})$**:**

  Backprop. & $(w_{\mathcal{H}_c}, w_{\mathcal{T}_c}) \leftarrow (w_{\mathcal{H}_c} - \eta \frac{\partial L_c}{\partial w_{\mathcal{H}_c}}, w_{\mathcal{T}_c} - \eta \frac{\partial L_c}{\partial w_{\mathcal{T}_c}})$

  return $(w_{\mathcal{H}_c}, w_{\mathcal{T}_c})$

Detailed Procedure of FeSTA

Function `ServerMain`:

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow \texttt{ClientHead}(c)$

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow \texttt{ClientTail}(c, b_c^{(i)})$ & Backprop.

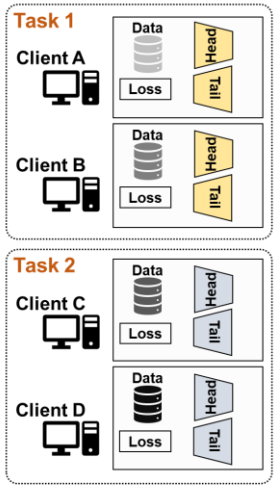
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow \texttt{ClientUpdate}(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

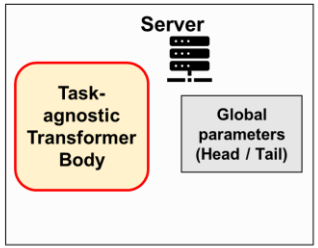
  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$





Detailed Procedure of FeSTA

Function `ServerMain`:

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow \texttt{ClientHead}(c)$

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow \texttt{ClientTail}(c, b_c^{(i)})$ & Backprop.

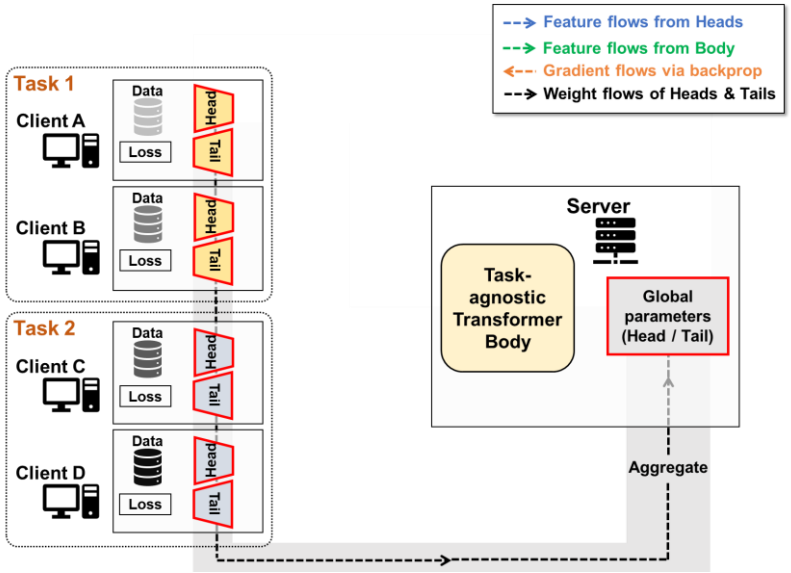
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow \texttt{ClientUpdate}(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Detailed Procedure of FeSTA

Function `ServerMain`**:**

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow \texttt{ClientHead}(c)$

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow \texttt{ClientTail}(c, b_c^{(i)})$ & Backprop.

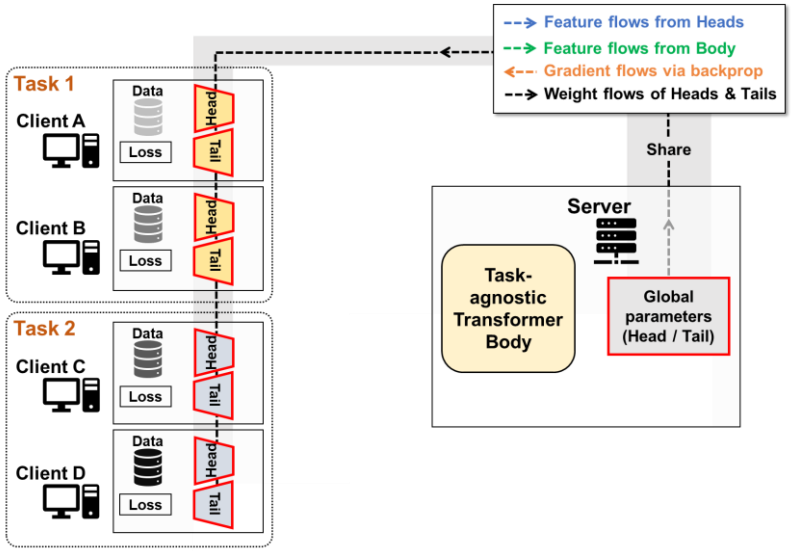
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow \texttt{ClientUpdate}(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Detailed Procedure of FeSTA

Function `ServerMain`**:**

Initialize the body weight $w_{\mathcal{B}}^{(1)}$ and client head/tail weights $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$ for each task $k \in \{1, ..., K\}$ in server

for rounds $i = 1, 2, \ldots R$ **do**

  for tasks $k \in \{1, 2, \ldots K\}$ **do in parallel**

    for clients $c \in C_k$ **do in parallel**

      if $i = 1$ **or** $(i-1) \in$ UnifyingRounds **then**

        Set client $(w_{\mathcal{H}_c}^{(i)}, w_{\mathcal{T}_c}^{(i)}) \leftarrow (\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k})$

      $h_c^{(i)} \leftarrow \texttt{ClientHead}(c)$

      $b_c^{(i)} \leftarrow \mathcal{B}(h_c^{(i)})$

      $\frac{\partial L_c^{(i)}}{\partial b_c^{(i)}} \leftarrow \texttt{ClientTail}(c, b_c^{(i)})$ & Backprop.

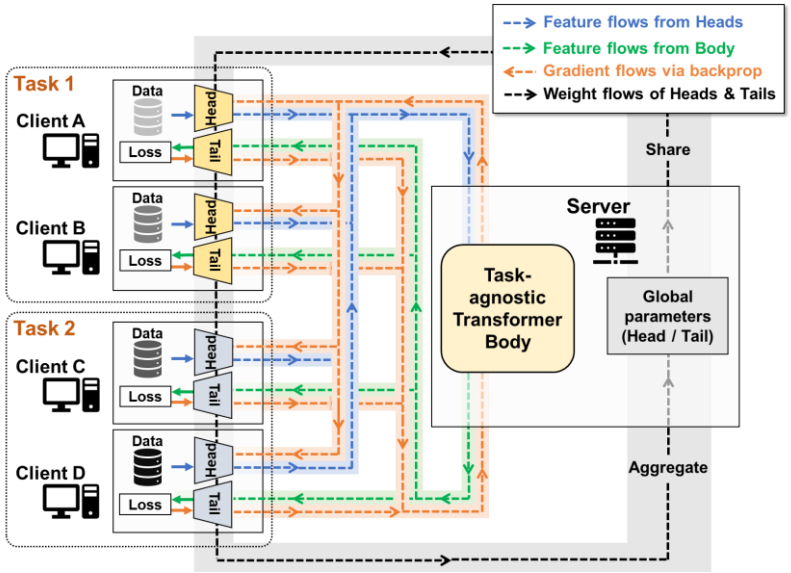
      $(w_{\mathcal{H}_c}^{(i+1)}, w_{\mathcal{T}_c}^{(i+1)}) \leftarrow \texttt{ClientUpdate}(c, \frac{\partial L_c^{(i)}}{\partial h_c^{(i)}})$

  Update body $w_{\mathcal{B}}^{(i+1)} \leftarrow w_{\mathcal{B}}^{(i)} - \frac{\eta}{K} \sum_{k=1}^{K} \sum_{c \in C_k} \frac{\partial L_c^{(i)}}{N_k \partial w_{\mathcal{B}}^{(i)}}$

  if $i \in$ UnifyingRounds **then**

    for tasks $k \in \{1, 2, \ldots K\}$ **do**

      Update $(\bar{w}_{\mathcal{H},k}, \bar{w}_{\mathcal{T},k}) \leftarrow (\frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{H}_c}^{(i+1)}, \frac{1}{N_k} \sum_{c \in C_k} w_{\mathcal{T}_c}^{(i+1)})$



Optimization in FeSTA

Task-agnostic body update

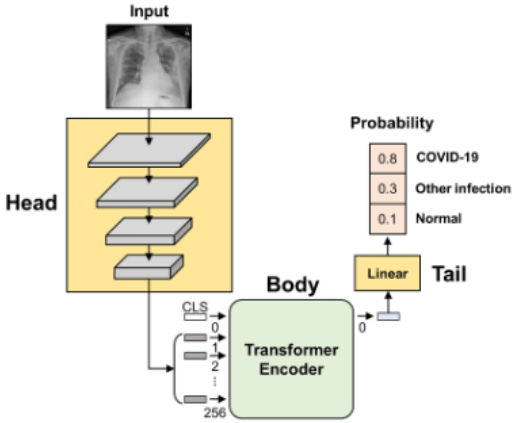
$$\min_{\mathcal{B}} \sum_{c \in \mathcal{C}} \sum_{i=1}^{N_c} \ell_c(y_c^{(i)}, \mathcal{T}_c(\mathcal{B}(\mathcal{H}_c(x_c^{(i)}))))$$

Task-specific head & tail update

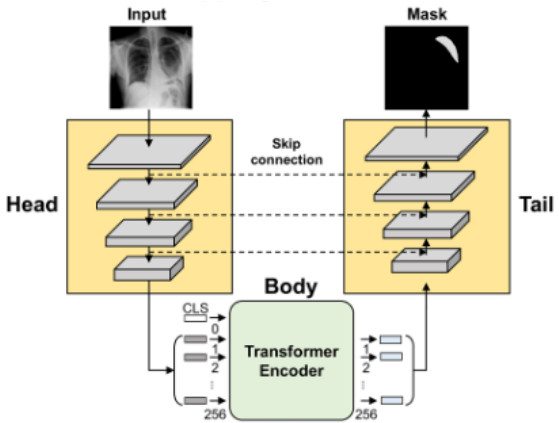
$$\min_{\mathcal{H}_c, \mathcal{T}_c} \sum_{i=1}^{N_c} \ell_c(y_c^{(i)}, \mathcal{T}_c(\mathcal{B}(\mathcal{H}_c(x_c^{(i)}))))$$

Network Configurations for CXR Tasks

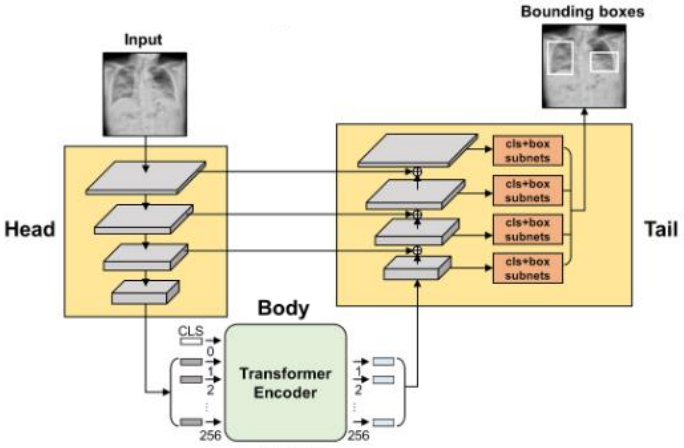
COVID-19 Classification



Pneumothorax Segmentation



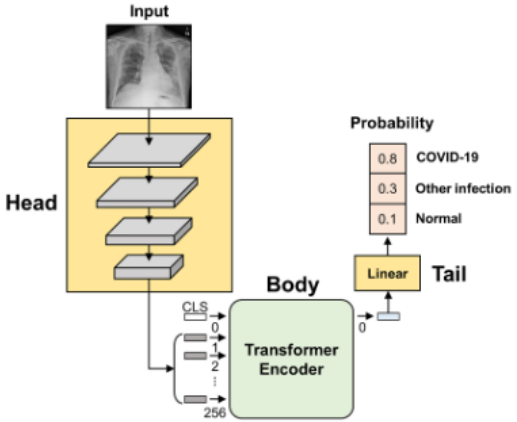
Pneumonia Detection



Task-agnostic Body: Shared

Network Configurations for CXR Tasks

COVID-19 Classification



Pneumothorax Segmentation

Pneumonia Detection

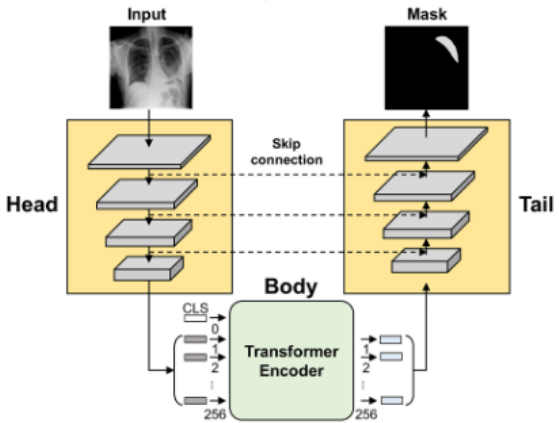
Task-agnostic Body: Shared

Network Configurations for CXR Tasks

COVID-19 Classification

Pneumothorax Segmentation

Pneumonia Detection



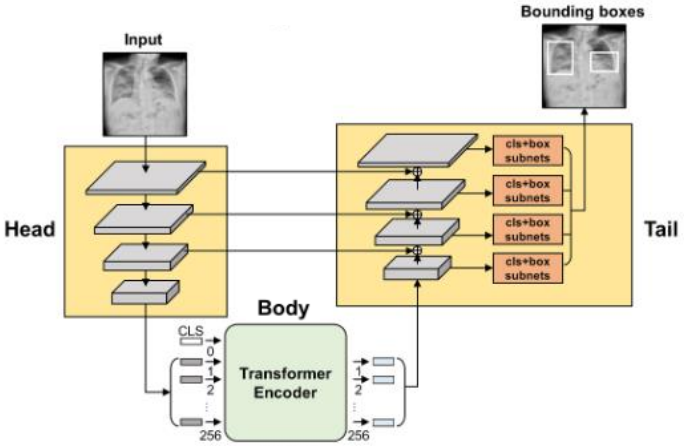
Task-agnostic Body: Shared

Network Configurations for CXR Tasks

COVID-19 Classification

Pneumothorax Segmentation

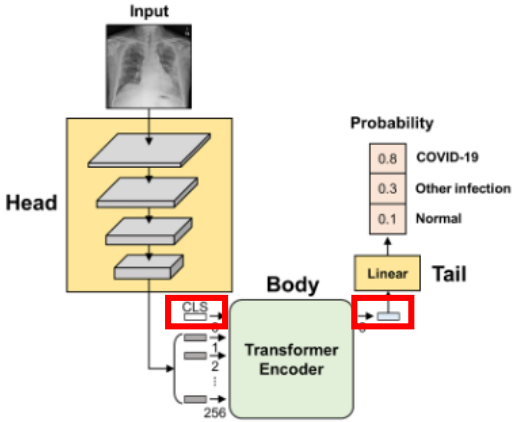
Pneumonia Detection



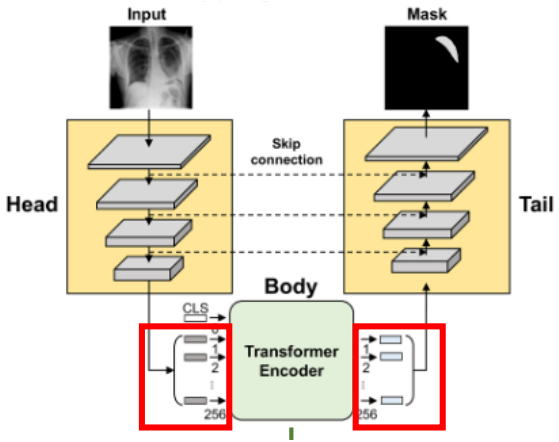
Task-agnostic Body: Shared

Network Configurations for CXR Tasks

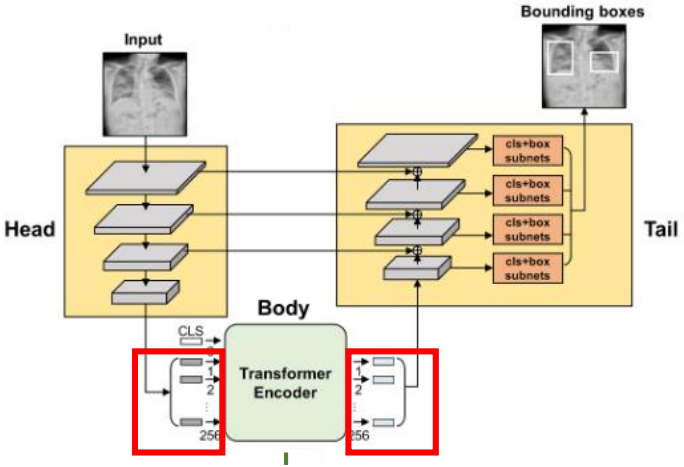
COVID-19 Classification



Pneumothorax Segmentation



Pneumonia Detection



Task-agnostic Body: Shared

Datasets for CXR Tasks

COVID-19 classification

Total CXR images		External	Training and validation dataset					
			Client 1	Client 2	Client 3	Client 4	Client 5	Client 6
		Hospital 1	Hospital 2	Hospital 3	Hospital 4	NIH	Brixia	BIMCV
Normal	13,649	320	300	400	8,861	3,768	-	-
Other infection	1,468	39	144	308	977	-	-	-
COVID-19	2,431	6	8	80	-	-	1,929	408
Total CXR	**17,548**	**365**	**452**	**788**	**9,838**	**3,768**	**1,929**	**408**

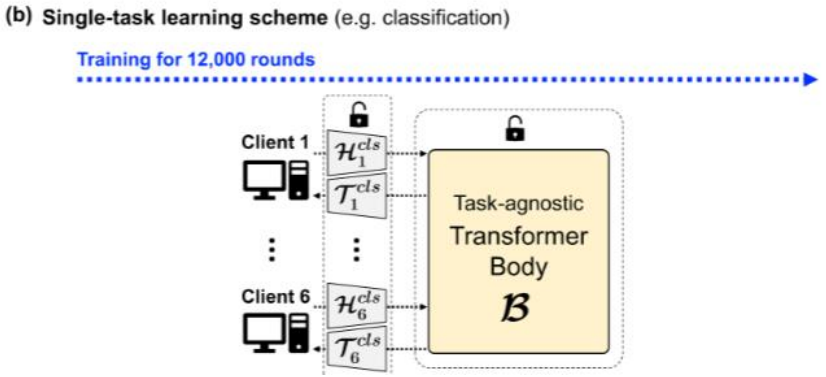
Pneumothorax segmentation (SIIM-ACR challenge)

- **12,047 subjects for training (→ randomly assigned 4:1 ratio for training and validation datasets) and 3,205 subjects for testing**

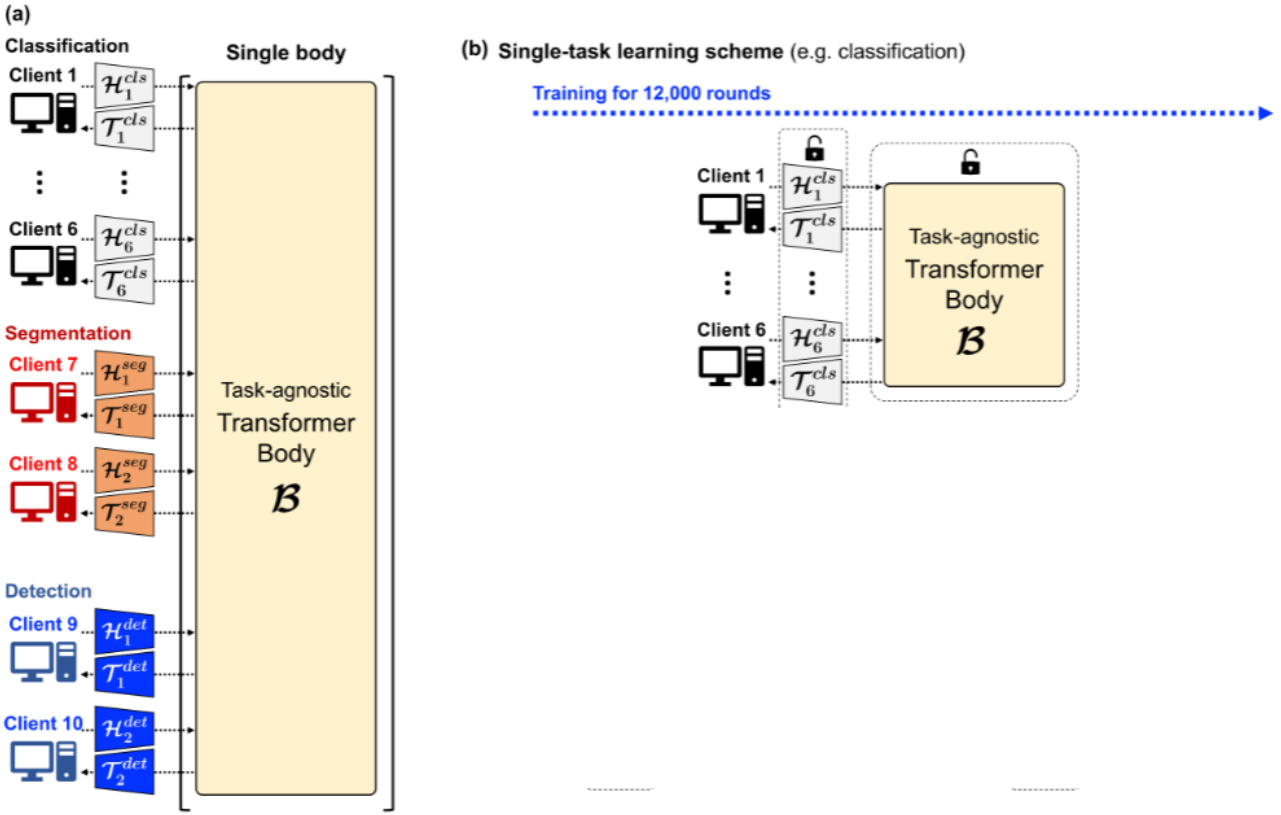
Pneumonia detection (RSNA challenge)

- **26,684 subjects → Randomly assigned 3:1 ratio to training and testing datasets**

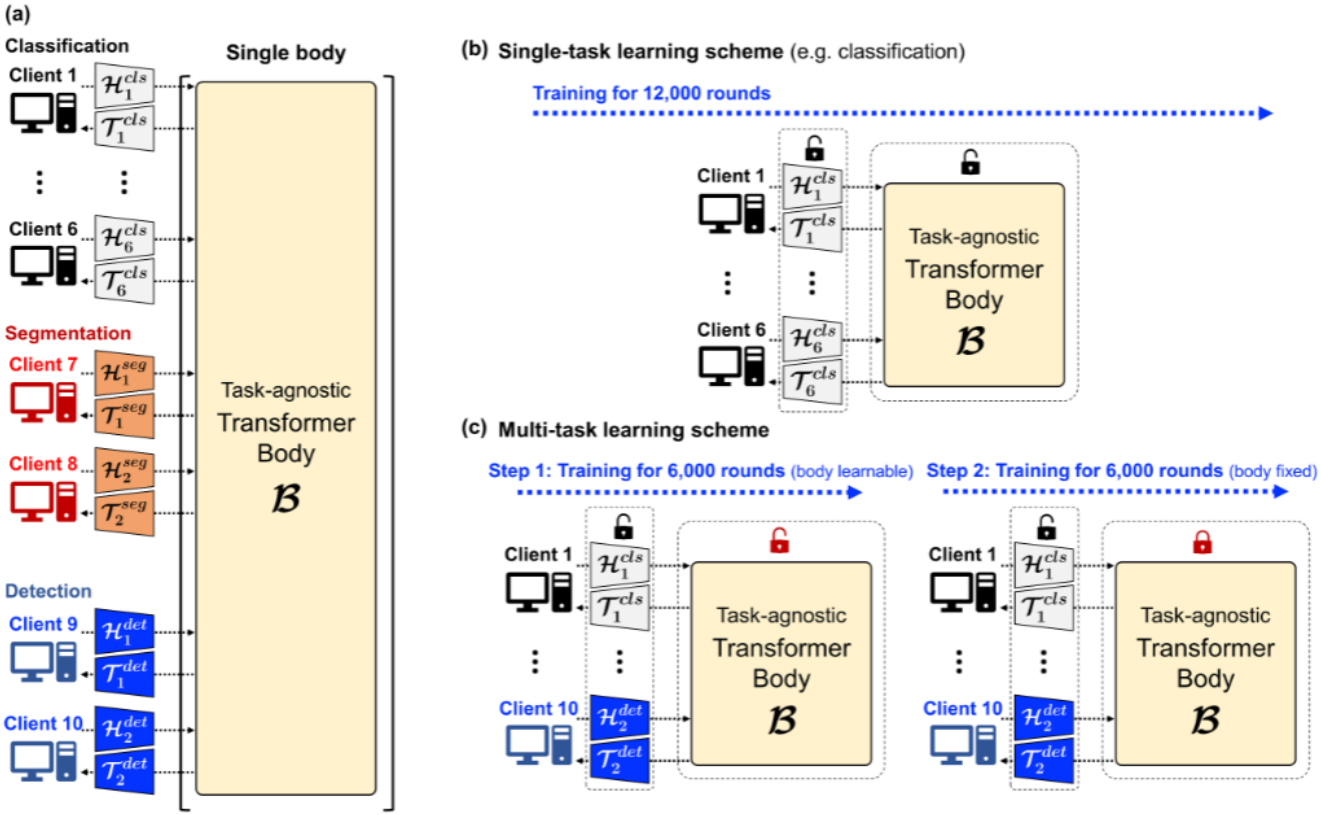
Detailed Experimental Setting



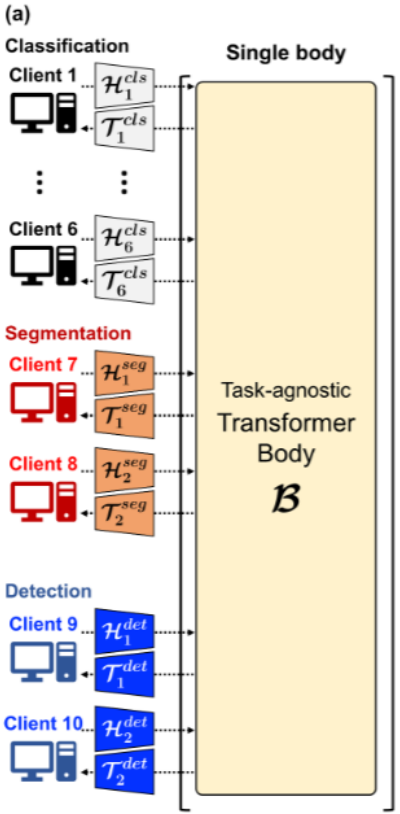
Detailed Experimental Setting



Detailed Experimental Setting

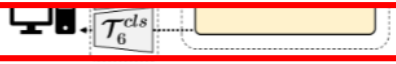


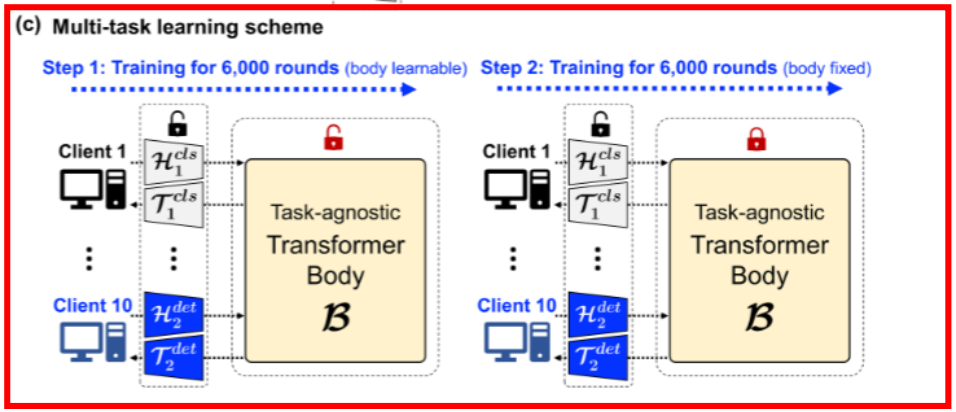
Detailed Experimental Setting



(b) **Single-task learning scheme** (e.g. classification)

Tasks	Classification	Segmentation	Detection
	AUC	Dice	mAP
Effect of training strategy			
One-step approach	0.930 ± 0.022	0.801 ± 0.024	0.188 ± 0.020
Alternating approach	0.915 ± 0.011	0.799 ± 0.021	0.179 ± 0.003
Two-step approach	**0.931 ± 0.004**	**0.821 ± 0.003**	**0.204 ± 0.002**





Results

Comparison of the FeSTA with other strategies

Strategy	AUC			
	Average	COVID-19	Others	Normal
Data-centralized	0.911 ± 0.016	0.883 ± 0.036	0.927 ± 0.013	0.923 ± 0.004
Federated learning	0.891 ± 0.019	0.840 ± 0.035	0.926 ± 0.018	0.906 ± 0.028
Split learning	0.863 ± 0.005	0.807 ± 0.012	0.892 ± 0.007	0.889 ± 0.019
FeSTA (STL)	$\mathbf{0.909 \pm 0.021}$	$\mathbf{0.880 \pm 0.008}$	$\mathbf{0.916 \pm 0.038}$	$\mathbf{0.931 \pm 0.021}$
FeSTA (MTL)	$\mathbf{0.931 \pm 0.004}$	$\mathbf{0.926 \pm 0.023}$	$\mathbf{0.929 \pm 0.016}$	$\mathbf{0.938 \pm 0.013}$

Comparison between single-task & multi-task learning

Tasks	Metrics	Single-task learning	Multi-task learning
Classification	AUC	0.909 ± 0.021	$\mathbf{0.931 \pm 0.004}$
Segmentation	Dice	0.798 ± 0.016	$\mathbf{0.821 \pm 0.003}$
Detection	mAP	0.202 ± 0.008	$\mathbf{0.204 \pm 0.002}$

Results

Comparison with task-specific expert network & CNN-based MTL model

Tasks	Metrics	Task-specific experts	CNN-based MTL model	Transformer-based MTL model (ours)
Classification	AUC	0.898 ± 0.004	0.907 ± 0.011	$\mathbf{0.931 \pm 0.004}$
Segmentation	Dice	0.736 ± 0.014	0.797 ± 0.018	$\mathbf{0.821 \pm 0.003}$
Detection	mAP	0.190 ± 0.006	0.159 ± 0.035	$\mathbf{0.204 \pm 0.002}$

Statistical comparison of performance between model with & without the transformer

Method	COVID-19		Others		Normal	
	AUC (95% CI)	p-value	AUC (95% CI)	p-value	AUC (95% CI)	p-value
w/o Transformer body	0.867 (0.696 - 1.000)	-	0.883 (0.817 - 0.948)	-	0.889 (0.837 - 0.941)	-
w Transformer body (STL)	**0.868 (0.749 - 0.987)**	**0.988**	**0.905 (0.852 - 0.958)**	**0.498**	**0.927 (0.889 - 0.965)**	**0.019**
w Transformer body (MTL)	**0.945 (0.896 - 0.995)**	**0.266**	**0.893 (0.833 - 0.954)**	**0.768**	**0.938 (0.903 - 0.974)**	**0.010**

Model Sizes & Communicative Benefit

Parameter numbers & model sizes of sub-networks

Task	Head		Body		Tail	
	Parameters	Size	Parameters	Size	Parameters	Size
Classification	13.313 M	54.1 MB	66.367 M	265.5 MB	0.002 M	11.7 KB
Segmentation	15.041 M	60.2 MB			7.387 M	29.6 MB
Detection	27.085 M	108.8 MB			19.773 M	79.1 MB

Note: Model sizes were estimated by parameter numbers and file sizes of saved weights.

- **Enables efficient communication under limited transmission speed and network capacities**

Model Sizes & Communicative Benefit

Communication cost per 1 Federated Averaging (FedAvg)

	Total transmission	Feature and gradient transmission	Network parameter transmission
Classification			
Federated learning	159.365M	-	159.365M
Split learning	78.950M	78.950M	-
FESTA	105.580M	78.950M	26.630M
Segmentation			
Federated learning	177.592M	-	177.592M
Split learning	78.950M	78.950M	-
FESTA	123.808M	78.950M	44.858M
Detection			
Federated learning	226.450M	-	226.450M
Split learning	78.950M	78.950M	-
FESTA	172.665M	78.950M	93.715M

- When period between FedAvg is k , transmission of features, gradients and network parameters are F , G and P respectively, total transmission from Server to Client T can be represented as follows:

$$T = k \times (F + G) + P$$

- T for each strategies can be formulated as follows:

$$T_{FL} = P_h + P_b + P_t$$
$$T_{SL} = 100 \times (F + G)$$
$$T_{FESTA} = 100 \times (F + G) + (P_h + P_t)$$

- If the transmission from Server to Client T and that from Client to Server $T_{C \to S}$ are assumed to be equal ($T_{C \to S} = T$), total transmission T' is as follows:

$$T' = 2T$$

Summary

- A novel Federated Split Task-Agnostic (FeSTA) framework suitable to leverage the benefit of ViT to process multiple CXR tasks are proposed.
- We showed that the proposed method outperforms the existing distributed learning methods, showing comparable performance to data-centralized method even under the extremely skewed data distribution.
- Our framework alongside with clients to process multiple related tasks also improves the performances of individual task.

Thank you!

Sangjoon Park*
Ph.D. Student
depecher@kaist.ac.kr

Gwanghyun Kim*
M.S. Student
gwang.kim@kaist.ac.kr

Jeongsol Kim
M.S. Student
wjdthf3927@kaist.ac.kr

Boah Kim
Ph.D. Student
boahkim@kaist.ac.kr

Jong Chul Ye
Professor
jong.ye@kaist.ac.kr