Deep Learning for Computer Vision

Fall 2022

https://cool.ntu.edu.tw/courses/189345 (NTU COOL)

http://vllab.ee.ntu.edu.tw/dlcv.html (Public website)

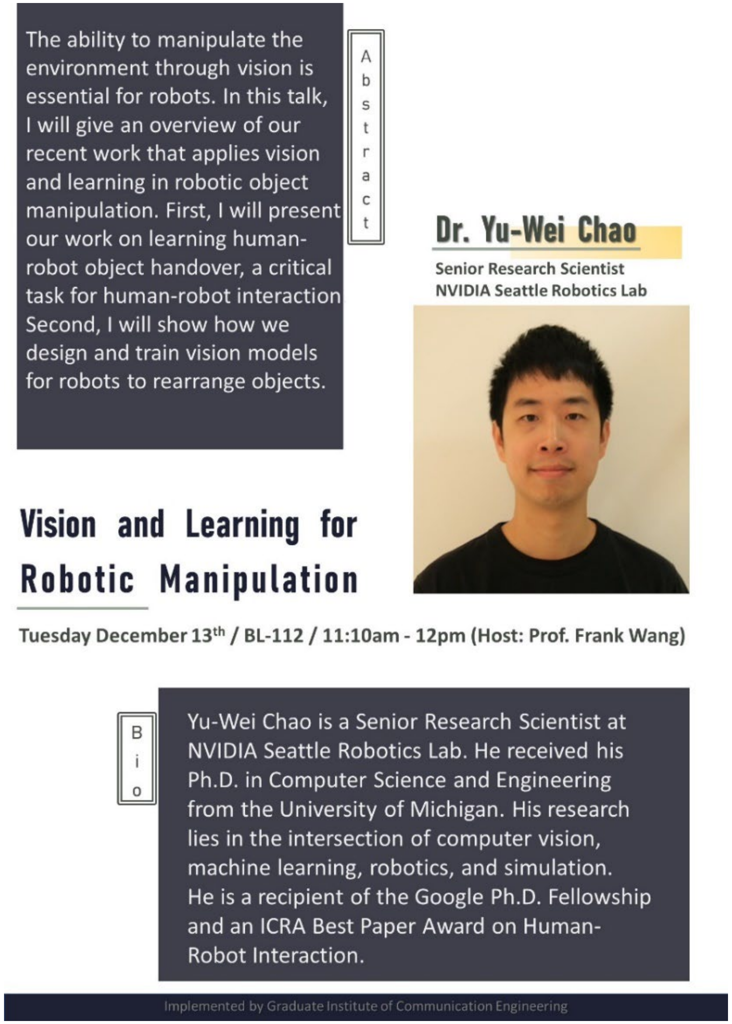
Yu-Chiang Frank Wang 王鈺強, Professor

Dept. Electrical Engineering, National Taiwan University

2022/12/13

What to Cover Today…

- Self-Supervised Learning (SSL)
  - SSL Beyond Images
- Domain Generalization
- Federated Learning
- Invited Talk
  - Vision and Learning for Robotic Manipulation
  - Dr. Yu-Wei Chao
    Sr. Research Scientist
    NVDIA Seattle Robotics Lab



Remarks

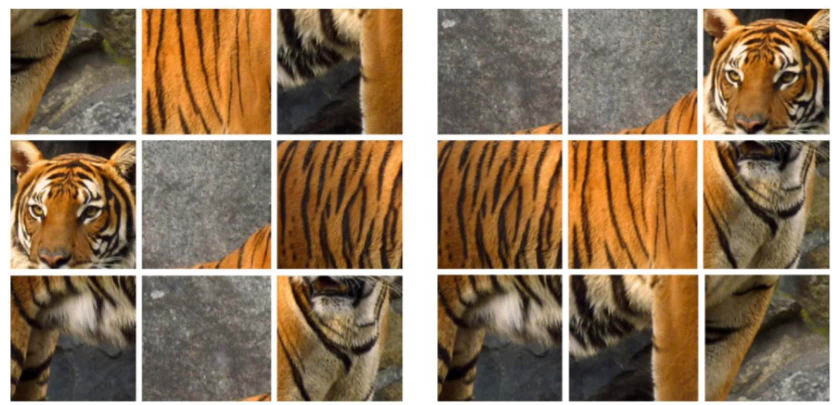
- Final Challenge
  - Date: Thursday, Dec. 29th
  - Location: TBD
  - Cash Prize: NTD $10K/5K/3K for the top 3 teams
  - Snack boxes will be provided
- 期末教學意見調查
  - ePo學習歷程檔 https://if163.aca.ntu.edu.tw/eportfolio/
  - 期末教學意見調查 https://investea.aca.ntu.edu.tw/opinion/login.asp

Self-Supervised Learning (SSL)

- Learning discriminative representations from **unlabeled** data
- Create self-supervised tasks via **data augmentation**

Colorization

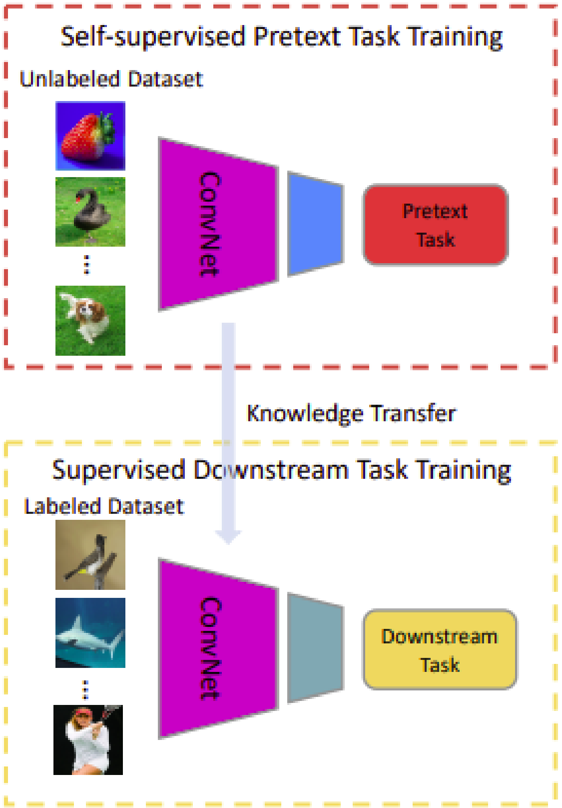
Rotation



Jigsaw Puzzle

Self-Supervised Learning (SSL)

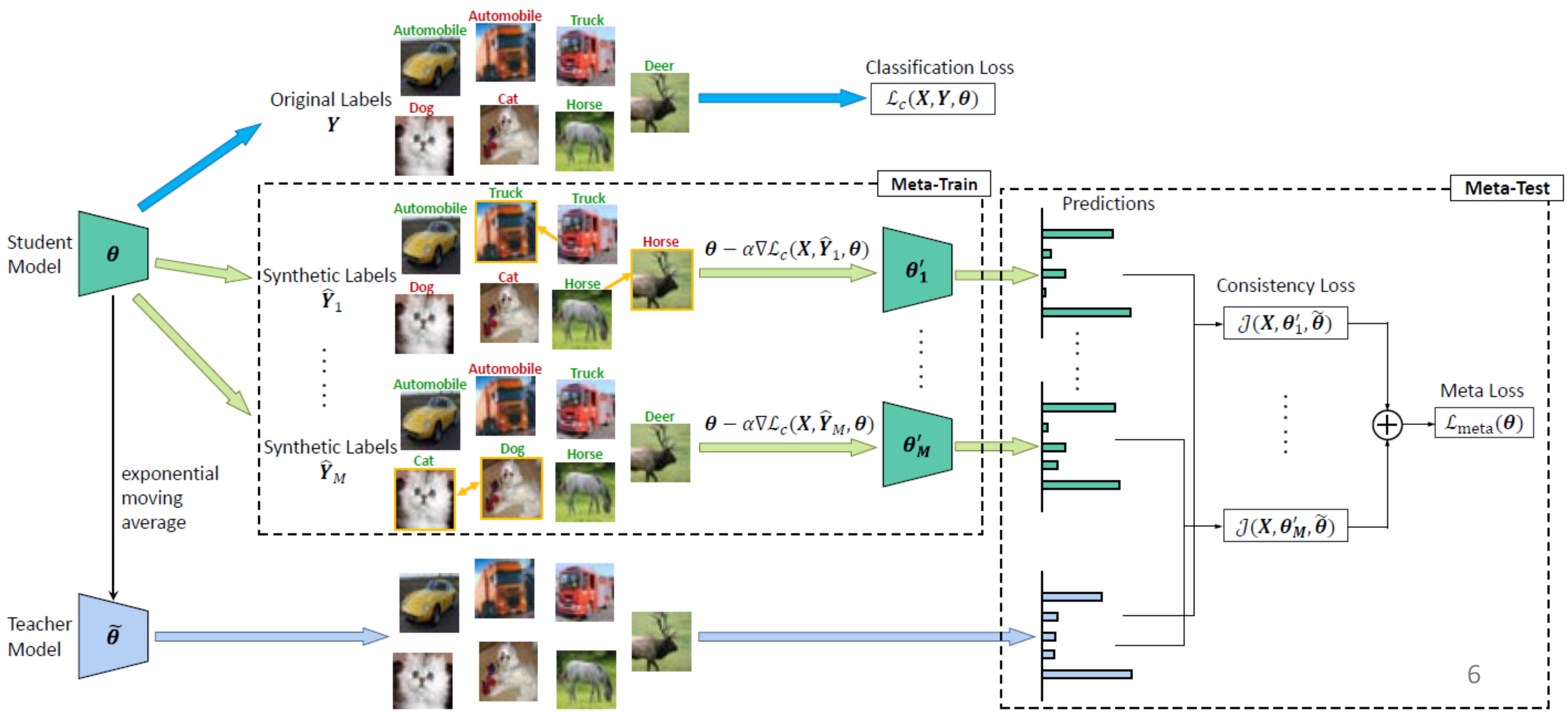
- Self-Supervised Pretraining (e.g., pretext training or contrastive learning)
    - Pretext Tasks
        - Jigsaw (ECCV’16)
        - RotNet (ICLR’18)
    - Contrastive Learning
        - CPC (ICML’20)
        - SimCLR (ICML’20)
    - Learning w/o negative samples
        - BYOL (NeurIPS’20)
        - Barlow Twins (ICML’21)
- Supervised Fine-tuning



SSL Beyond Image Data

- What about videos?

- What about noisy data? J. Li et al., Learning to Learn from Noisy Labeled Data, *CVPR* 2019

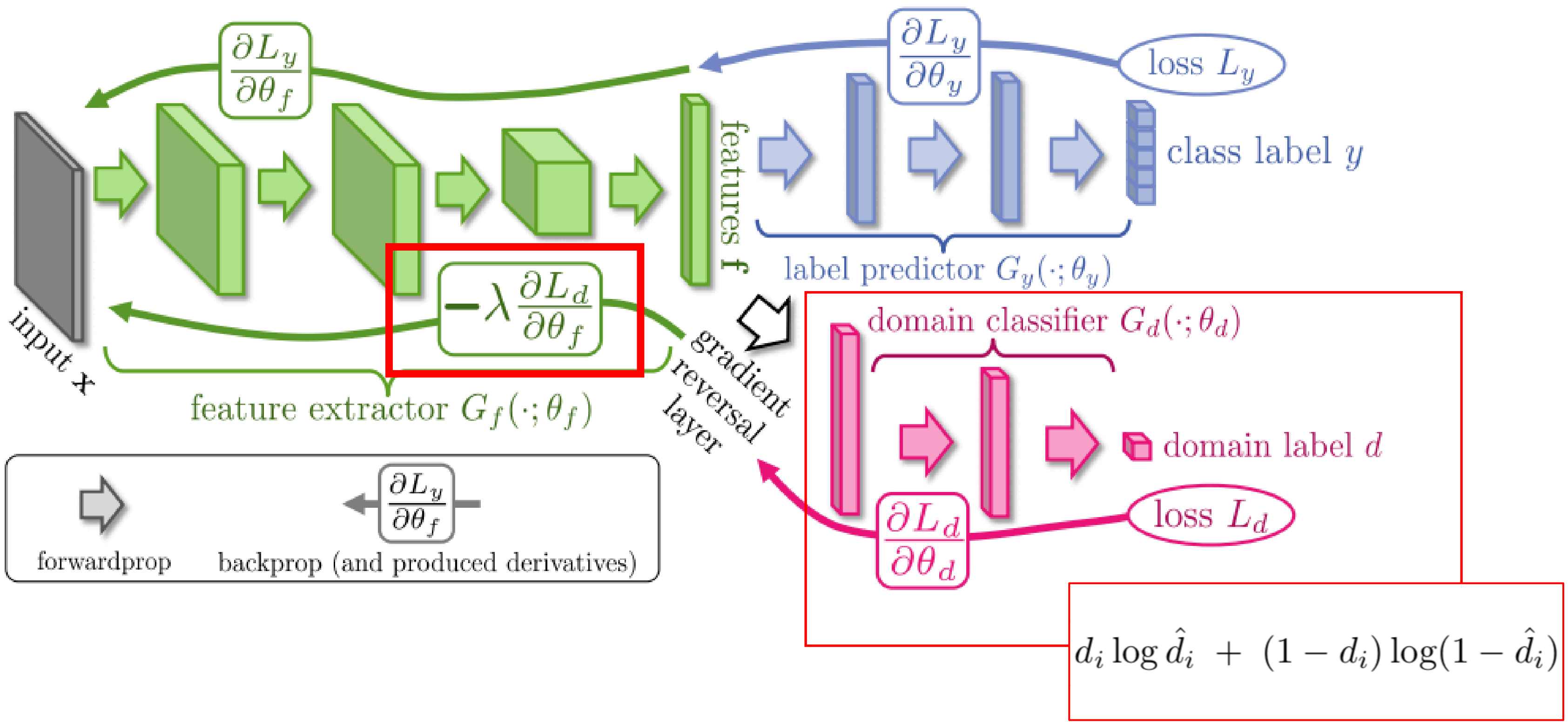


What to Cover Today…

- Self-Supervised Learning (SSL)
  - SSL Beyond Images
- Domain Generalization
- Federated Learning
- Invited Talk
  - Vision & Learning
    for Robotic Manipulation
  - Dr. Yu-Wei Chao
    Sr. Research Scientist, NVIDIA

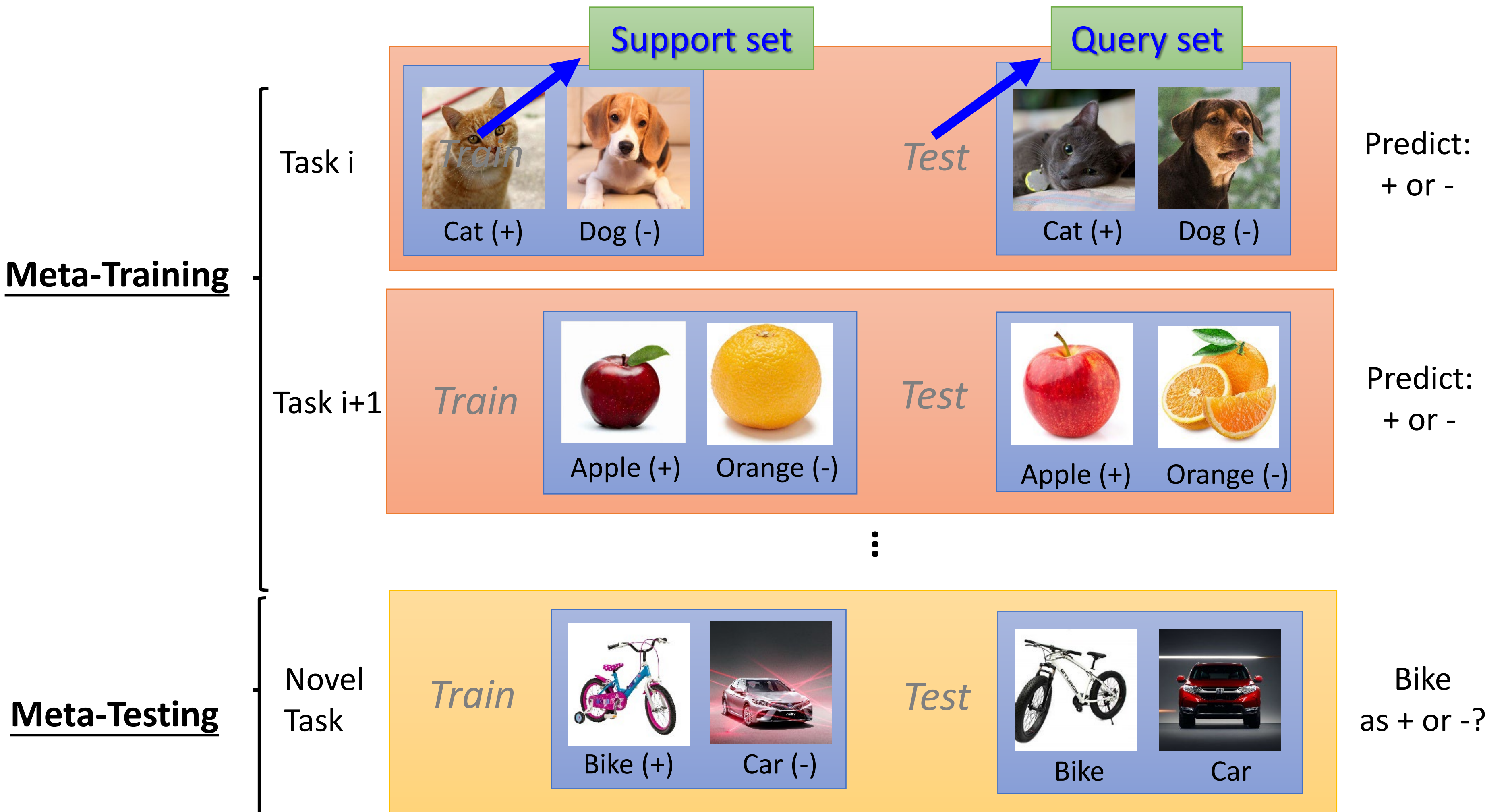
Recap: Domain Adaptation

- Domain-Adversarial Training of Neural Networks (DANN)
  - Y. Ganin et al., ICML 2015
  - Maximize domain confusion = maximize domain classification loss
  - Minimize source-domain data classification loss
  - The derived feature f can be viewed as a <u>disentangled & domain-invariant feature</u>.



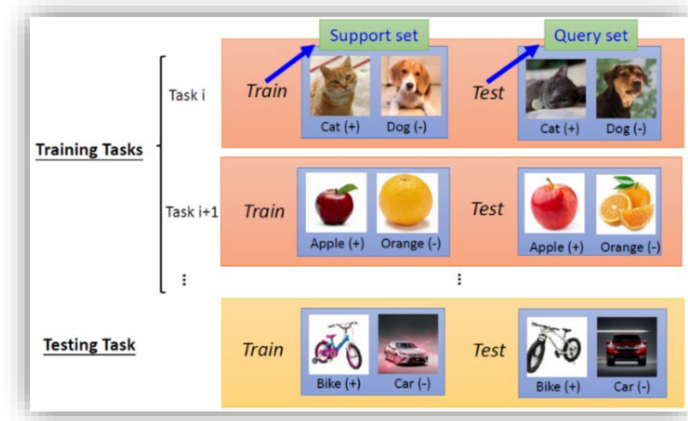
Recap: Meta Learning = Learning to Learn

- A powerful solution for learning from few-shot data
- Let's consider the following "2-way 1-shot" learning scheme:

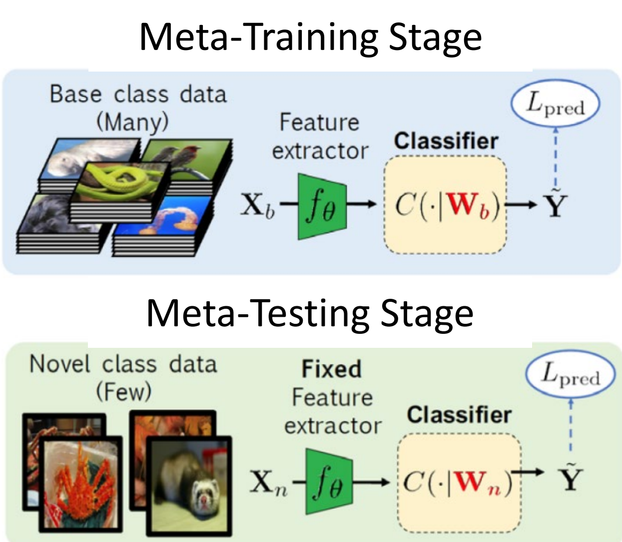


Slide credit: H.-Y. Lee

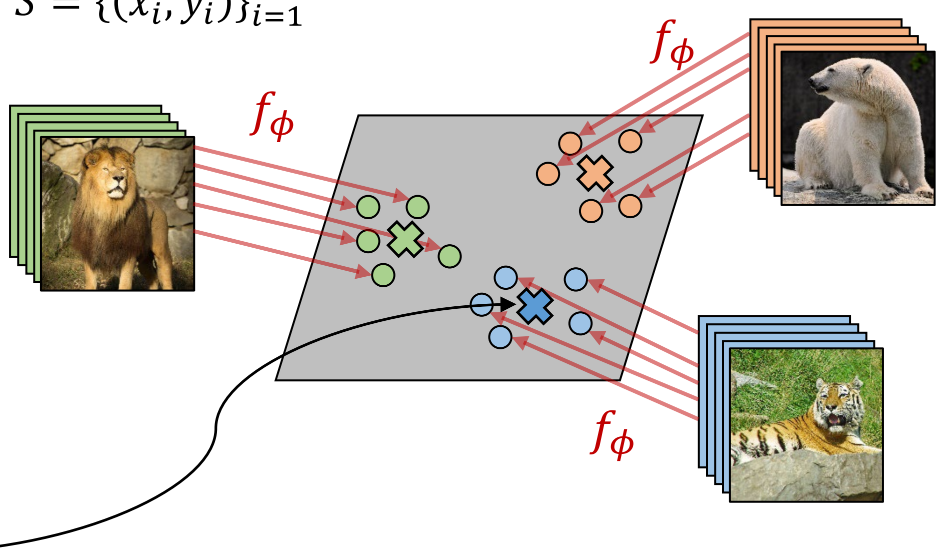
Recap:
Learn to Compare
with the Representative Ones!



- Prototypical Networks
  - Learn a model which properly describes data in terms of intra/inter-class info.
  - It learns a prototype for each class, with data similarity/separation guarantees. For DL version, the learned feature space is derived by a non-linear mapping f_θ and the representatives (i.e., prototypes) of each class is the **mean feature vector $\mathbf{c}_k$**.



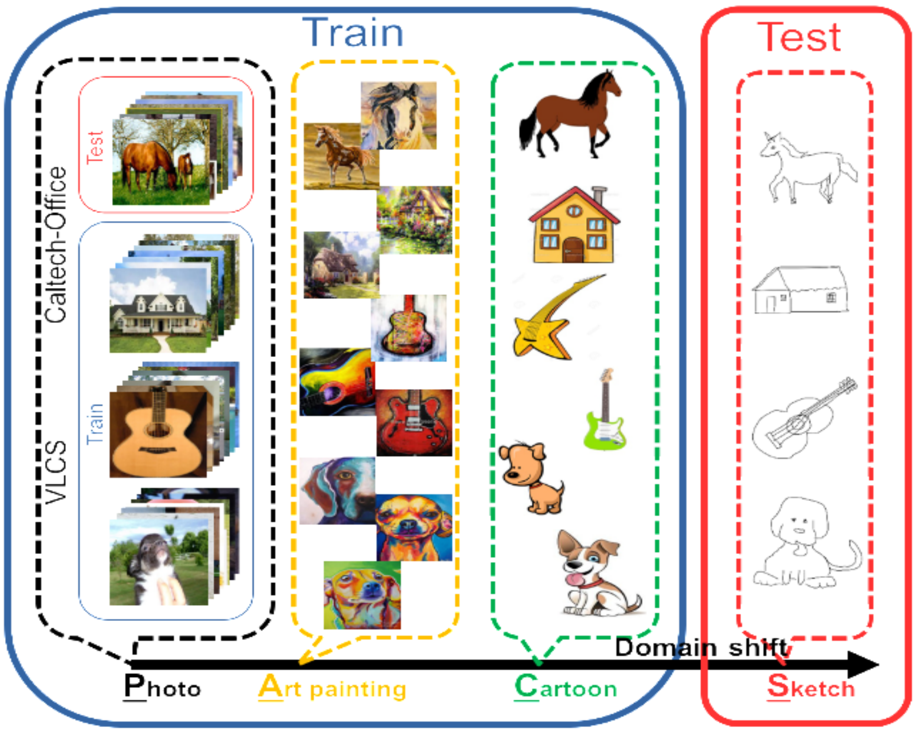
support set
$S = \{(x_i, y_i)\}_{i=1}^{k}$



$$\mathbf{c}_k = \frac{1}{|S_k|} \sum_{(\mathbf{x}_i, y_i) \in S_k} f_\theta(\mathbf{x}_i)$$, where $S_k \subset S$ indicates features of class k from support set S

Domain Generalization

- Input: Images and labels from multiple source domains
- Output: A well-generalized model for unseen target domains

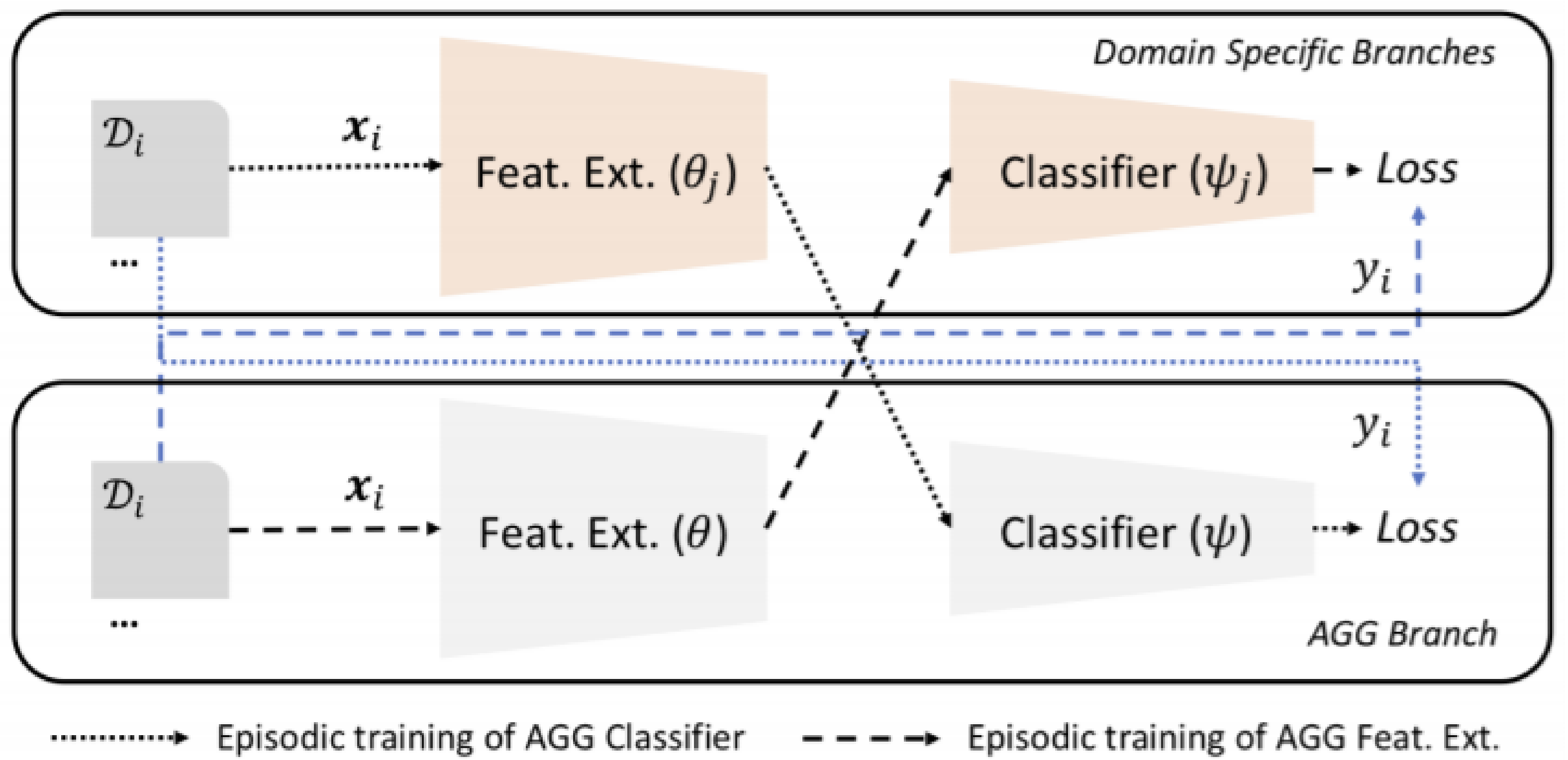


D_S = {Photo, Painting, Cartoon}

D_T = {Sketch}

Strategy of Episodic Training

- **Episodic training for domain generalization** (ICCV'19)
- Generalize across domains via Meta-Learning

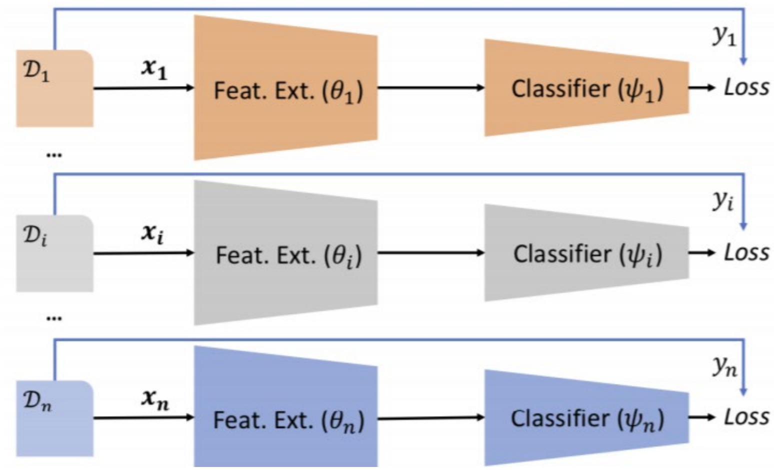


Zhang et al. : Episodic training for domain generalization. In ICCV (2019)

Episodic Training (cont'd)

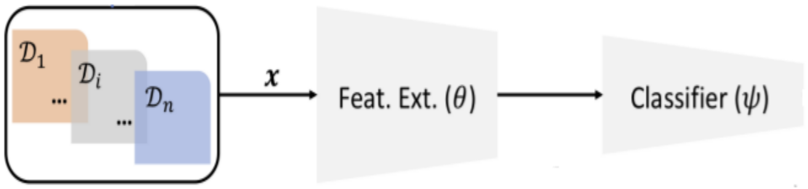
- Motivation

Domain Specific Models



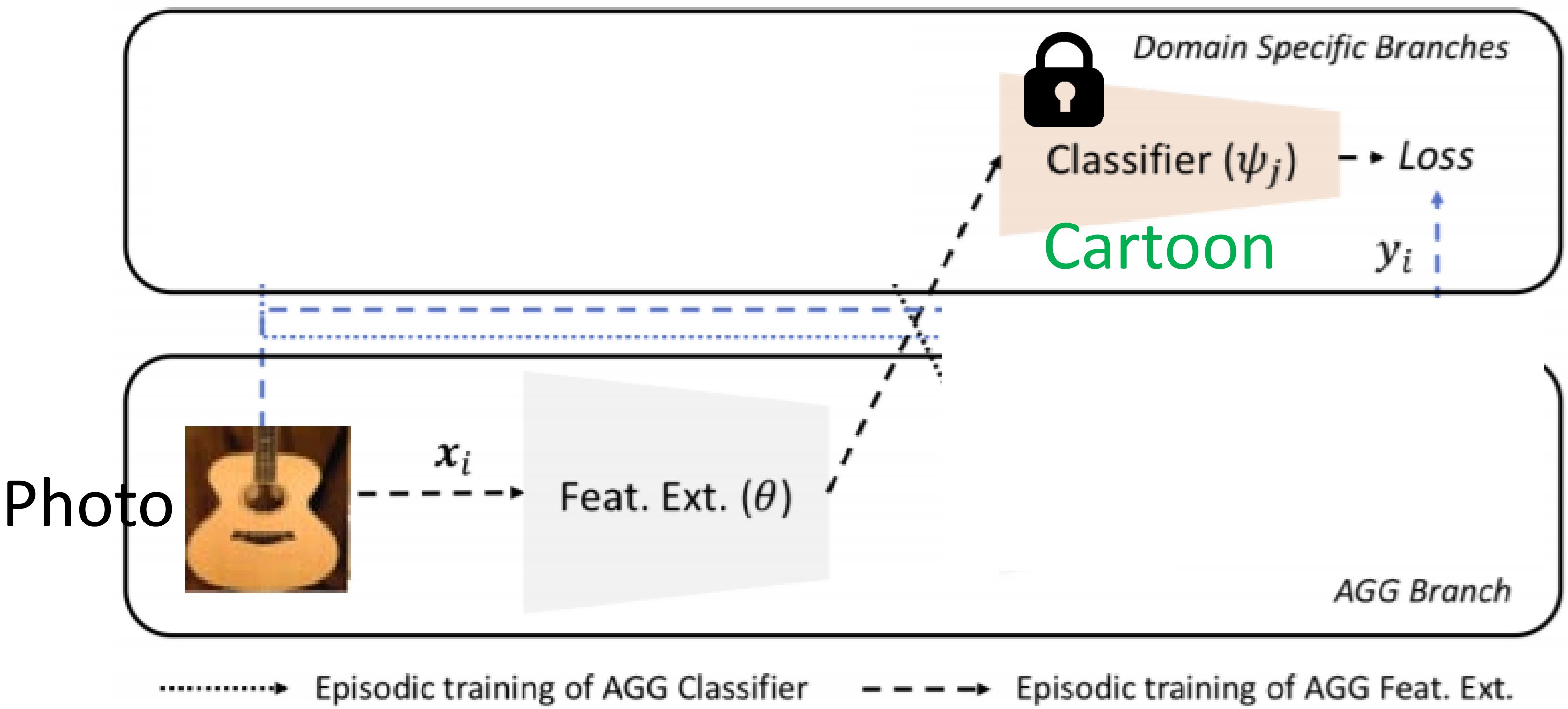
Episodic training

Aggregated Model



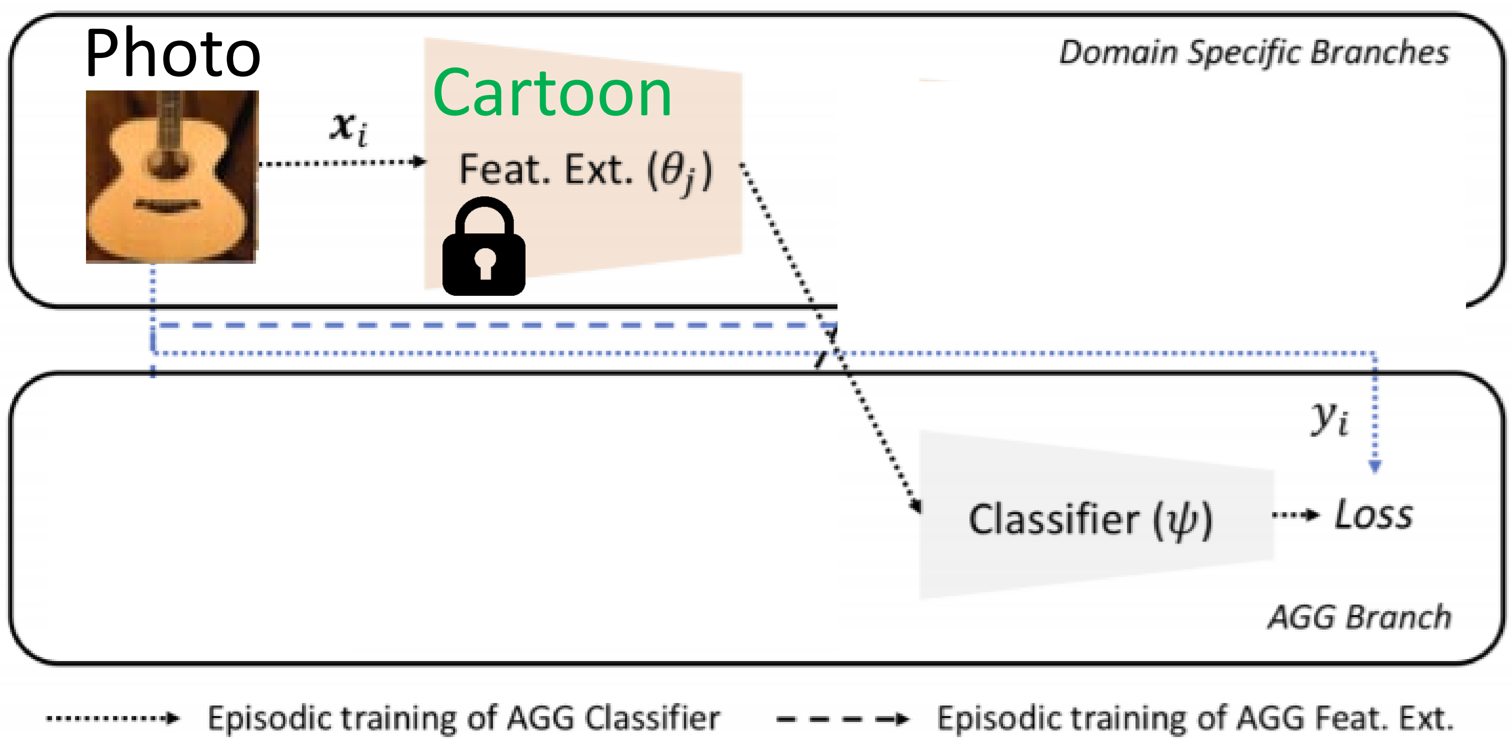
Episodic Training (cont'd)

- Random sample two domains, e.g., Photo and Cartoon

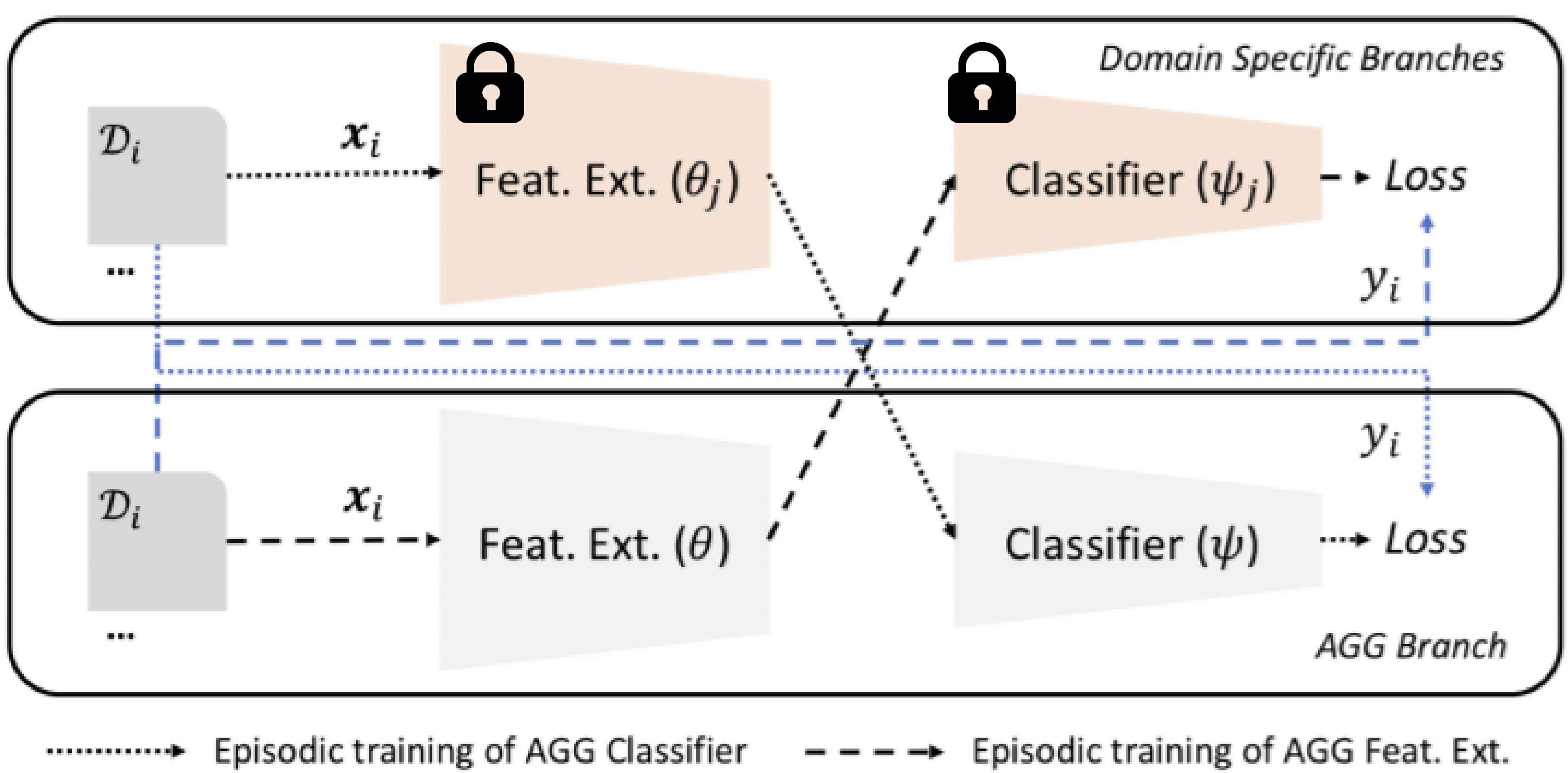


Episodic Training (cont'd)

- Random sample two domains, e.g., Photo and Cartoon

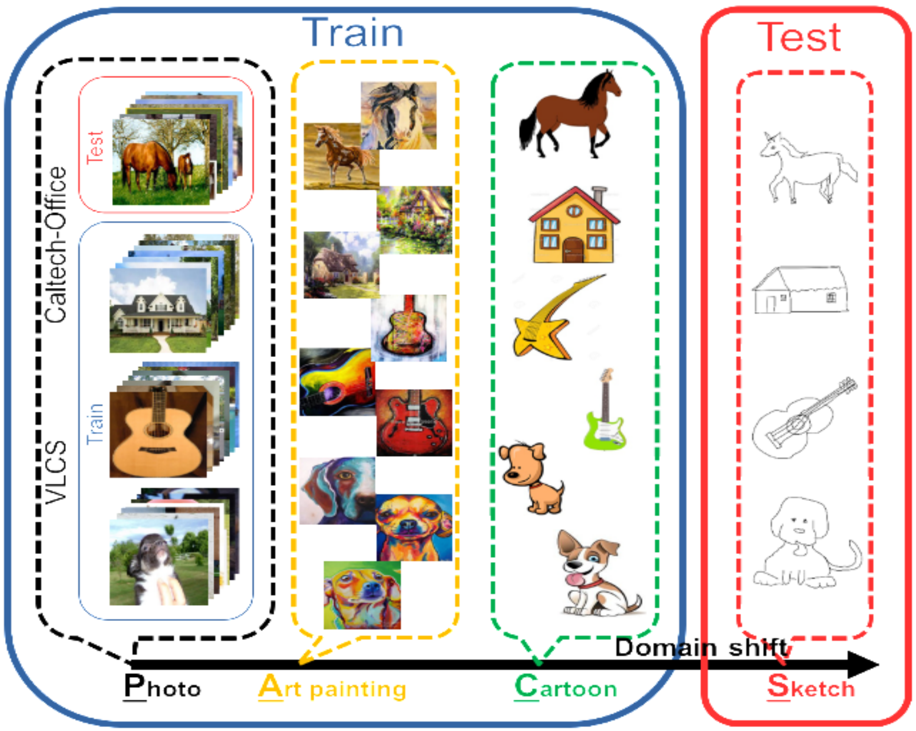


Episodic Training (cont'd)



Experiments

- Input: Images and labels from multiple source domains
- Output: A well-generalized model for unseen target domains



D_S = {Photo, Painting, Cartoon}

D_T = {Sketch}

Experiments (cont'd)

- Domain Generalized Classification

Source	Target	DICA [26]	LRE-SVM [38]	D-MTAE [12]	CCSA [25]	MMD-AAE [20]	DANN[11]	MLDG [18]	CrossGrad [32]	MetaReg [1]	AGG	Epi-FCR
0,1,2,3	4	61.5	75.8	78.0	75.8	**79.1**	75.0	70.7	71.6	74.2	73.1	76.9
0,1,2,4	3	72.5	86.9	92.3	92.3	94.5	94.1	93.6	93.8	94.0	94.2	**94.8**
0,1,3,4	2	74.7	84.5	91.2	94.5	95.6	97.3	97.5	95.7	96.9	95.7	**99.0**
0,2,3,4	1	67.0	83.4	90.1	91.2	93.4	95.4	95.4	94.2	97.0	95.7	**98.0**
1,2,3,4	0	71.4	92.3	93.4	**96.7**	**96.7**	95.7	93.6	94.0	94.7	94.4	96.3
Ave.		69.4	84.6	87.0	90.1	91.9	91.5	90.2	89.9	91.4	90.6	**93.0**

Table 1: Cross-view action recognition results (accuracy. %) on IXMAS dataset. Best result in bold.

Source	Target	DICA [26]	LRE-SVM [38]	D-MTAE [12]	CCSA [25]	MMD-AAE[20]	DANN [11]	MLDG [18]	CrossGrad [32]	MetaReg [1]	AGG	Epi-FCR
L,C,S	V	63.7	60.6	63.9	67.1	**67.7**	66.4	**67.7**	65.5	65.0	65.4	67.1
V,C,S	L	58.2	59.7	60.1	62.1	62.6	64.0	61.3	60.0	60.2	60.6	**64.3**
V,L,S	C	79.7	88.1	89.1	92.3	**94.4**	92.6	**94.4**	92.0	92.3	93.1	94.1
V,L,C	S	61.0	54.9	61.3	59.1	64.4	63.6	**65.9**	64.7	64.2	65.8	**65.9**
Ave.		65.7	65.8	68.6	70.2	72.3	71.7	72.3	70.5	70.4	71.2	**72.9**

Table 2: Cross-dataset object recognition results (accuracy. %) on VLCS benchmark. Best in bold.

Zhang et al.: Episodic training for domain generalization. In ICCV (2019)

What to Cover Today…

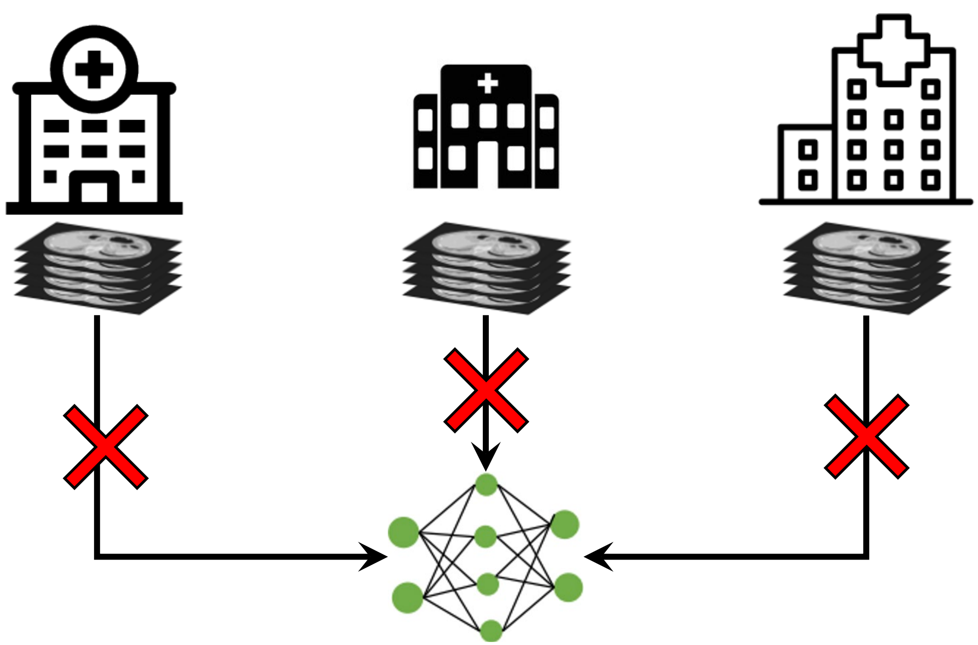
- Self-Supervised Learning (SSL)
  - SSL Beyond Images
- Domain Generalization
- Federated Learning
- Invited Talk
  - Vision & Learning
    for Robotic Manipulation
  - Dr. Yu-Wei Chao
    Sr. Research Scientist, NVIDIA

Outline

- Introduction to Federated Learning
- Federated Learning on Non-IID Data Silos
- Beyond Supervised Federated Learning
  - Semi-supervised
  - Self-supervised
- Personalized Federated Learning

Why Federated Learning?

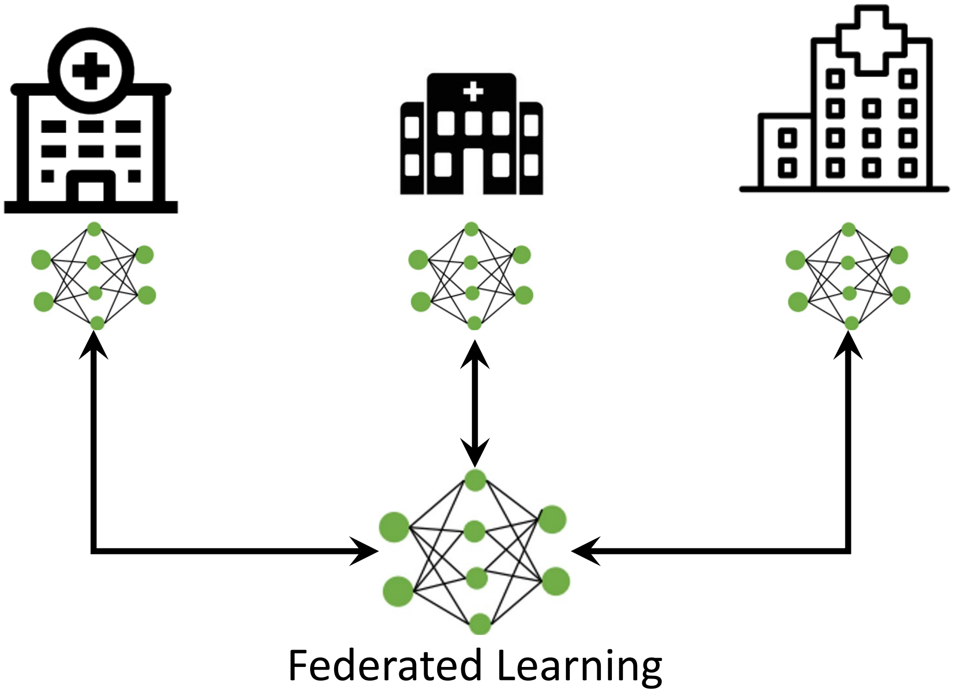
- Data privacy issue becomes a growing concern in modern AI services
- Regulations like CCPA (California) or GDPR (Europe) restrict data transmission across different data sources



Centralized Learning

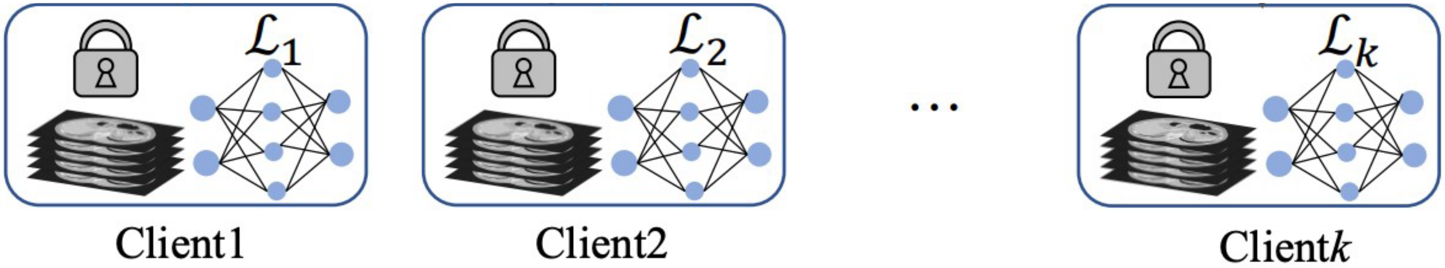
Federated Learning

- **Collaborative learning** without centralizing data
- Share **model weights** instead of **raw data (or features)**!
- Model training occurs locally at each participant/client



Federated Learning (cont'd)

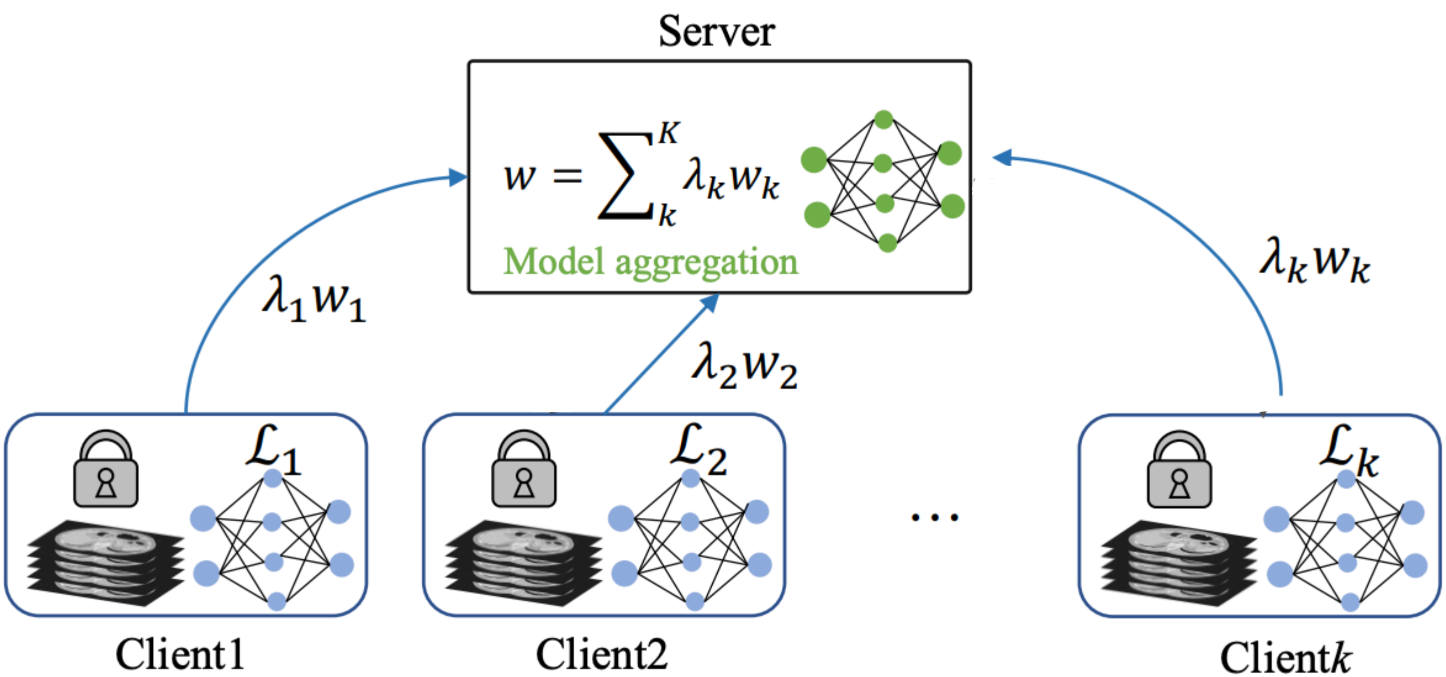
- Training models collaboratively without sharing the raw data
- FedAvg:
  - Local client training using private data



Communication-Efficient Learning of Deep Networks from Decentralized Data, AISTATS 2017

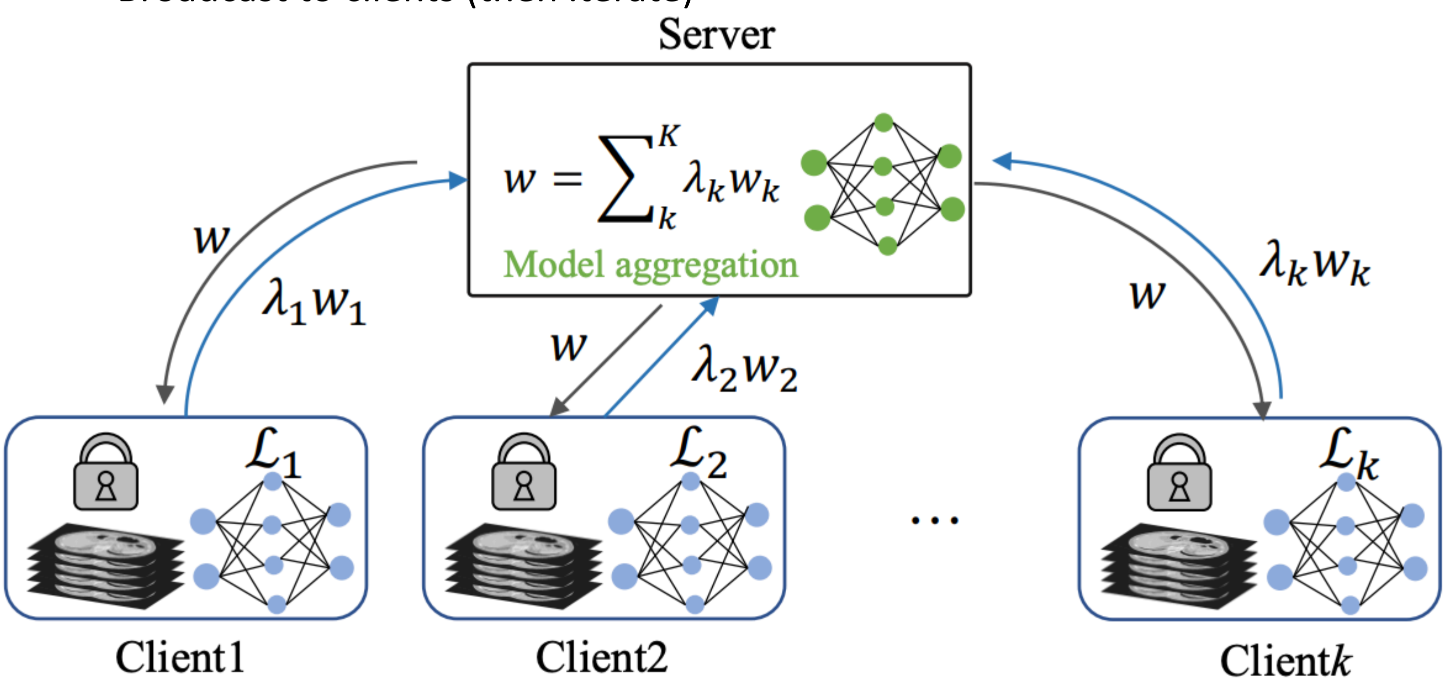
Federated Learning (cont'd)

- Training models collaboratively without sharing the raw data
- FedAvg:
    - Local client training using private data --> Server aggregation (i.e., averaging)



Federated Learning (cont'd)

- Training models collaboratively without sharing the raw data
- FedAvg:
  - Local client training using private data --> Server aggregation (Averaging) --> Broadcast to clients (then iterate)



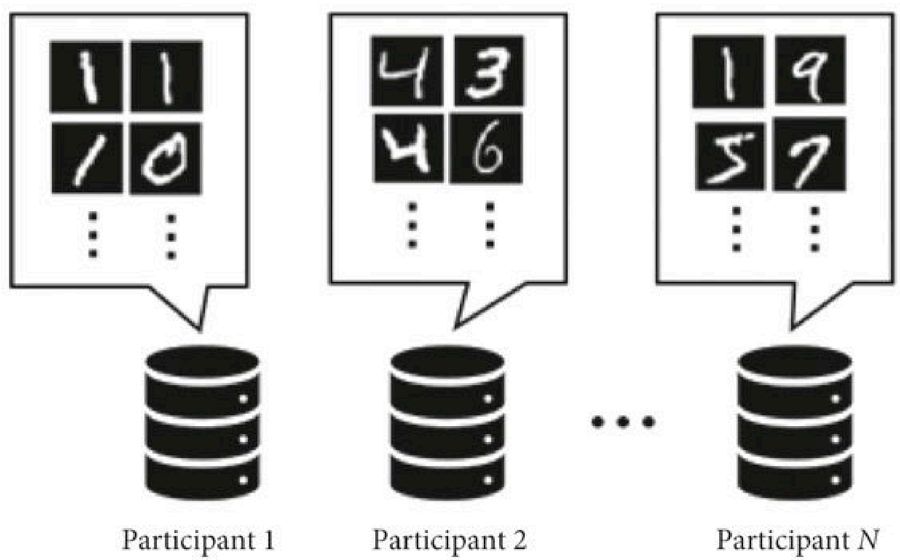
Communication-Efficient Learning of Deep Networks from Decentralized Data, AISTATS 2017

Outline

- Introduction to Federated Learning
- Federated Learning on Non-IID Data Silos
- Beyond Supervised Federated Learning
    - Semi-supervised
    - Self-supervised
- Personalized Federated Learning

Non-IID Data

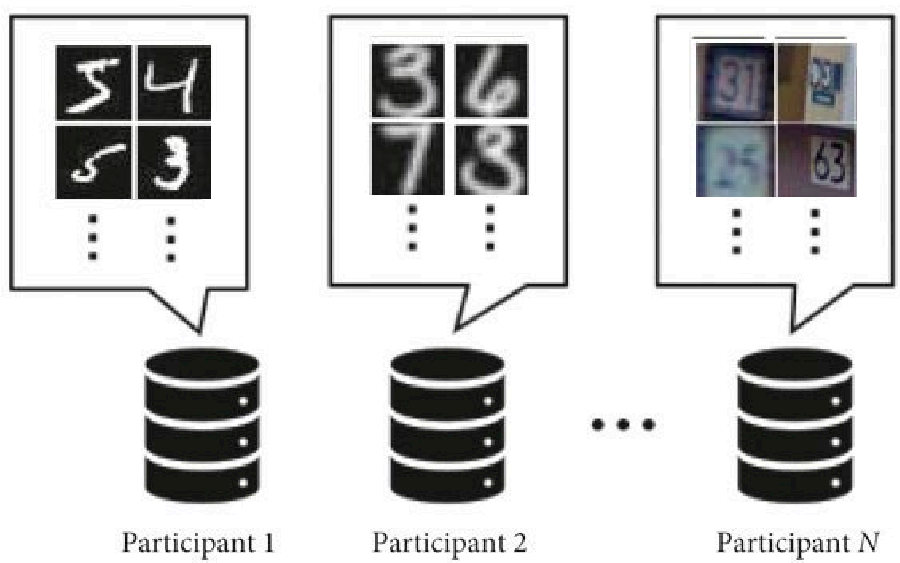
- In real-world FL applications, data distributions among different clients are usually **N**on-**I**ndependently and **I**dentically **D**istributed (**non-IID**)
- For example:
  - Class/label distribution skew



Reference: Utility Optimization of Federated Learning with Differential Privacy, DDNS 2021

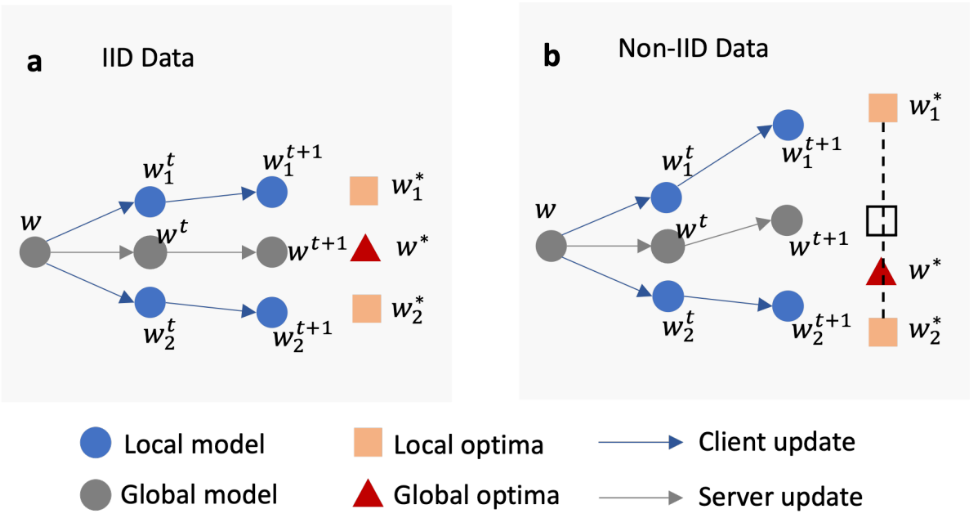
Non-IID Data (cont'd)

- In real-world FL applications, data distributions among different clients are usually **N**on-**I**ndependently and **I**dentically **D**istributed (**non-IID**)
- For example:
  - Label distribution skew
  - Domain shift



Non-IID Data (cont'd)

- In real-world FL applications, data distributions among different clients are usually **N**on-**I**ndependently and **I**dentically **D**istributed (**non-IID**)
- For example:
  - Label distribution skew
  - Domain shift
- Models trained on such data are hard to achieve global optima



Reference: Towards Personalized Federated Learning, TNNLS 2022

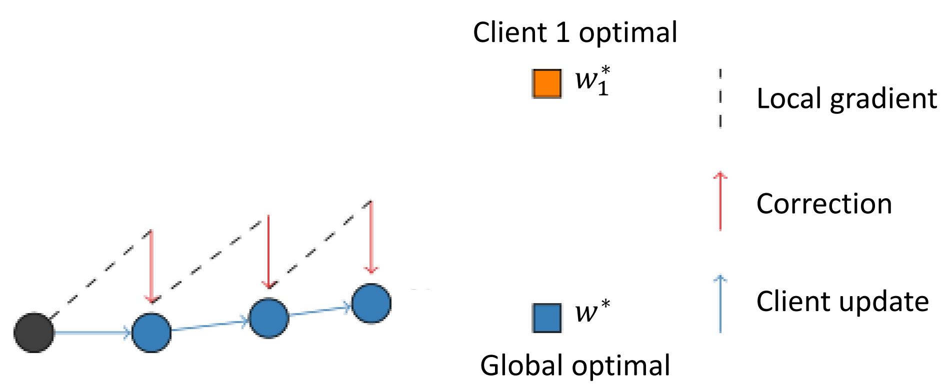
Tackling Non-IID Data (cont’d)

- Limiting the impacts of heterogeneous local updates
  - FedProx:
    Add a **proximal term** to force the local model to be closed to the global model
    - $\min_{w} h_k(w;\ w^t) = F_k(w) + \frac{\mu}{2}\|w - w^t\|^2$ l model weight *w* that satisfy:

      Proximal term

Tackling Non-IID Data (cont'd)

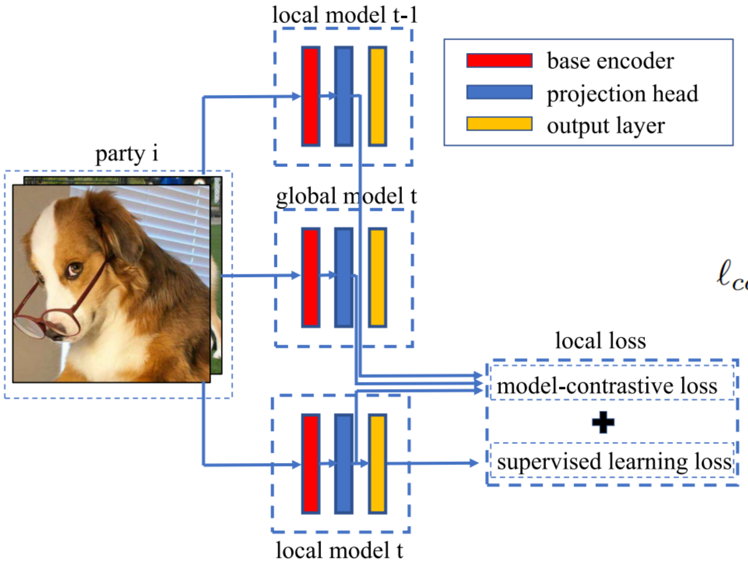
- Limiting the impacts of heterogeneous local updates
  - FedProx
  - SCAFFOLD: Correcting local gradient to avoid client drift



SCAFFOLD: Stochastic Controlled Averaging for Federated Learning, ICML 2020

Tackling Non-IID Data (cont'd)

- Limiting the impacts of heterogeneous local updates
  - FedProx
  - SCAFFOLD
  - MOON: Enforce local features to be similar to global features
    - (local model t, global model t) --> positive
    - (local model t, local model $t-1$) --> negative



$$\ell_{con} = -\log \frac{\exp(\mathrm{sim}(z, z_{glob})/\tau)}{\exp(\mathrm{sim}(z, z_{glob})/\tau) + \exp(\mathrm{sim}(z, z_{prev})/\tau)}$$

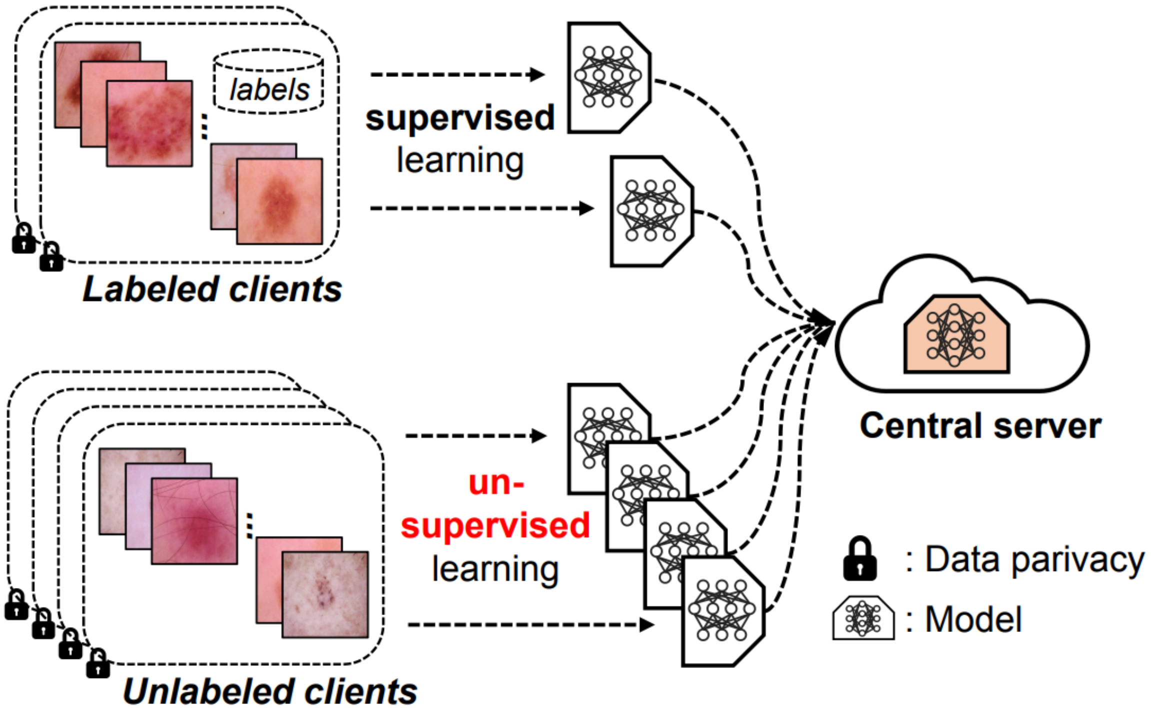
Model-Contrastive Federated Learning, CVPR 2021

Outline

- Introduction to Federated Learning
- Federated Learning on Non-IID Data Silos
- Beyond Supervised Federated Learning
  - Semi-supervised
  - Self-supervised
- Personalized Federated Learning

Federated Semi-Supervised Learning (FSSL)

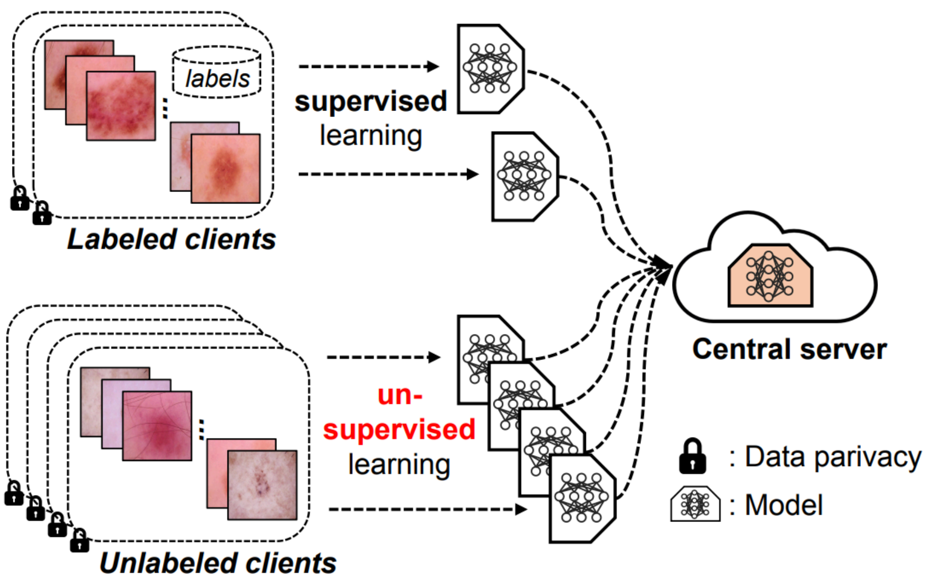
- Some labeled clients, and other unlabeled clients



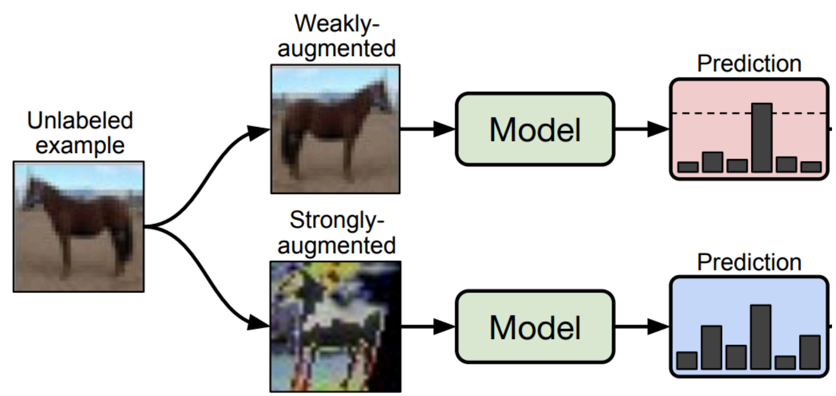
Federated Semi-supervised Medical Image Classification via Inter-client Relation Matching, MICCAI 2021

FSSL Baseline Method

- ***Labeled clients:*** use standard cross-entropy loss
- ***Unlabeled clients:*** use consistency loss



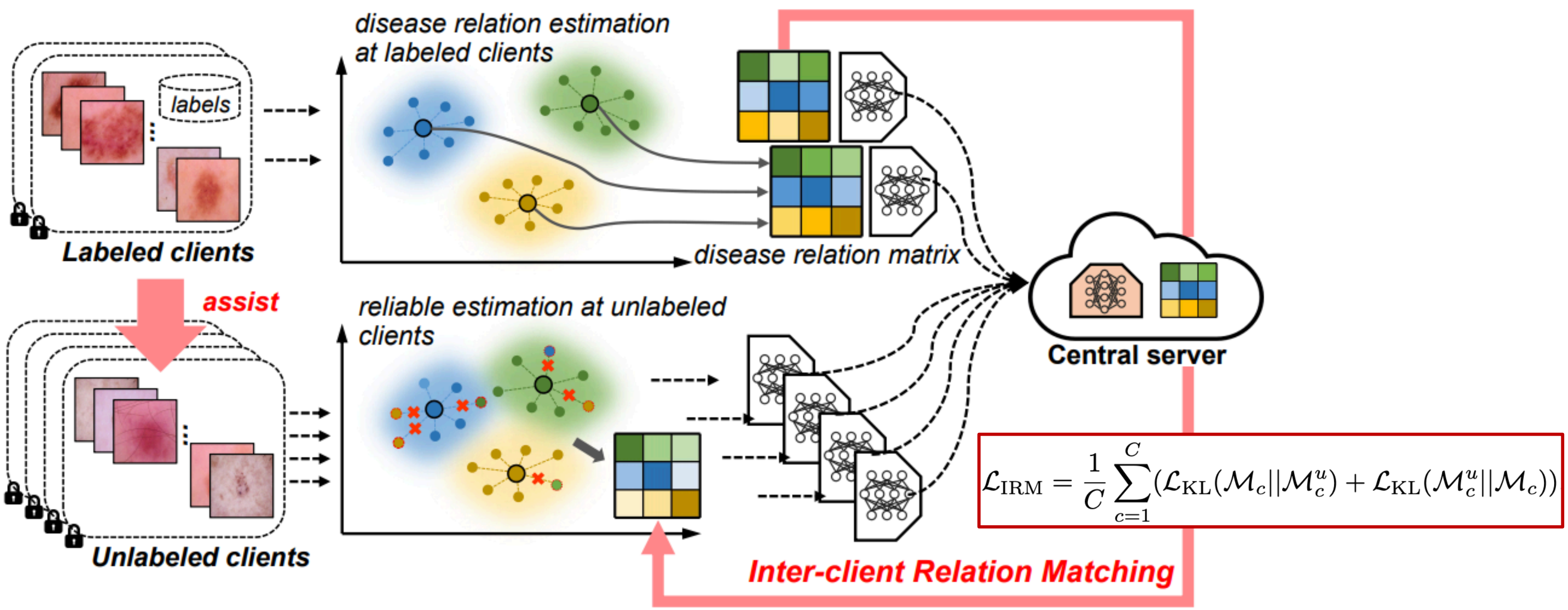
Unlabeled clients



$$\mathcal{L}_c(\mathcal{D}^u, \theta^u) = \sum_{i=1}^{N^u} \mathbb{E}_{\xi,\xi'} || f_{\theta^u}(x_i^u, \xi), f_{\theta^u}(x_i^u, \xi') ||_2^2$$

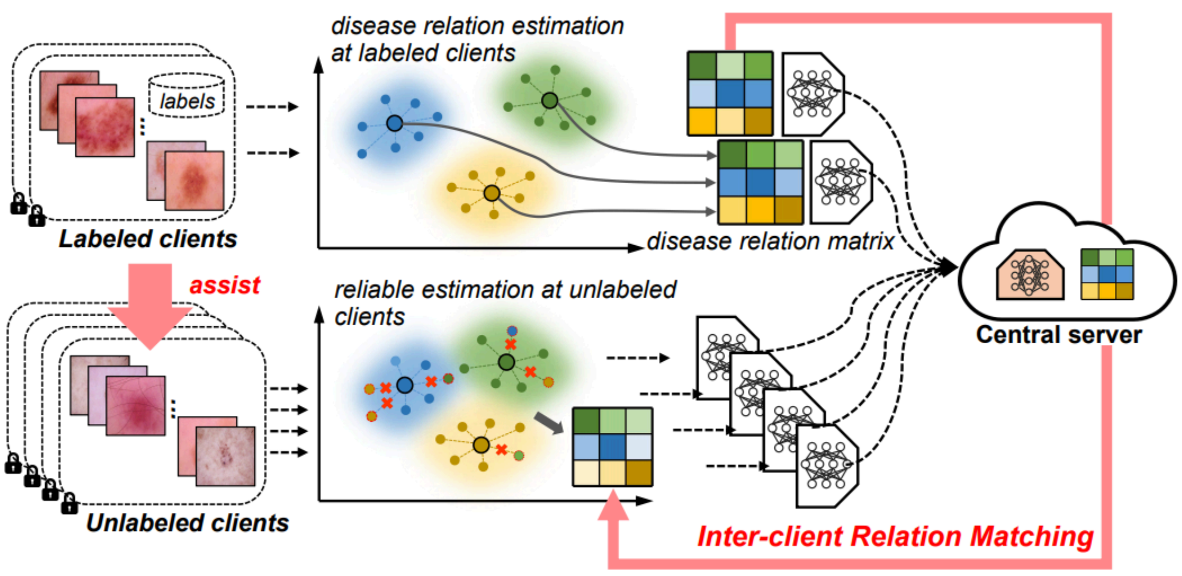
FedIRM

- Labeled clients use *class-correlation matrix* to guide the learning of unlabeled clients



FedIRM (cont'd)

- Labeled clients use *class-correlation matrix* to guide the learning of unlabeled clients



- **Per-category mean feature**

$$\mathbf{v}_c^l = \frac{1}{N_c^l} \sum_{i=1}^{N^l} \mathbb{1}_{[y_i^l = c]} \hat{f}_{\theta^l}(x_i^l) \quad \mathbf{v}_c^l \in \mathbb{R}^C$$

- **Soft label distribution**

$$\mathbf{s}_c^l = \mathrm{softmax}(\mathbf{v}_c^l / \tau)$$

- **Class confusion matrix**

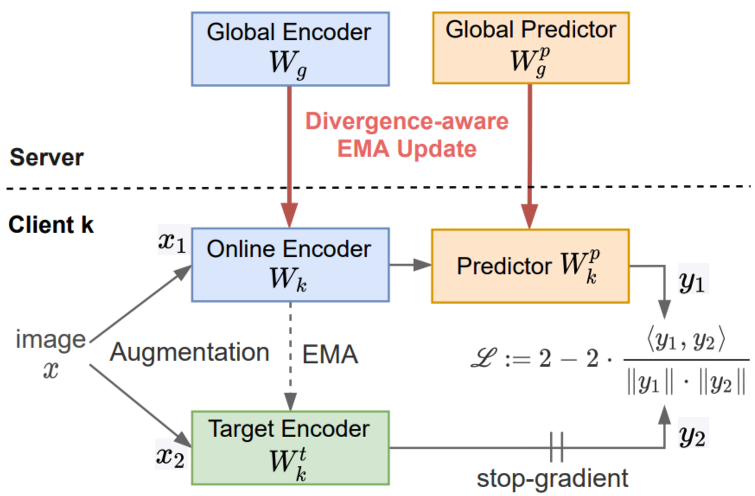
$$\mathcal{M}^l = [\mathbf{s}_1^l, \ldots, \mathbf{s}_C^l]$$

- **Inter-client relation matching**

$$\mathcal{L}_{\mathrm{IRM}} = \frac{1}{C} \sum_{c=1}^{C} (\mathcal{L}_{\mathrm{KL}}(\mathcal{M}_c || \mathcal{M}_c^u) + \mathcal{L}_{\mathrm{KL}}(\mathcal{M}_c^u || \mathcal{M}_c))$$

Federated Self-Supervised Learning

- Learn useful representation from **distributed unlabeled** data
- FedEMA
  - Local training: apply BYOL for local training
  - Update online network (student) with divergence-aware EMA
    - If W_k is similar to W_g, update W_k
    - Otherwise, keep W_k unchanged for retaining local knowledge



$$W_k^r = \mu W_k^{r-1} + (1-\mu) W_g^r,$$

$$W_k^{p,r} = \mu W_k^{p,r-1} + (1-\mu) W_g^{p,r},$$

$$\mu = \min(\lambda \left\| W_g^r - W_k^{r-1} \right\|, 1),$$

Outline

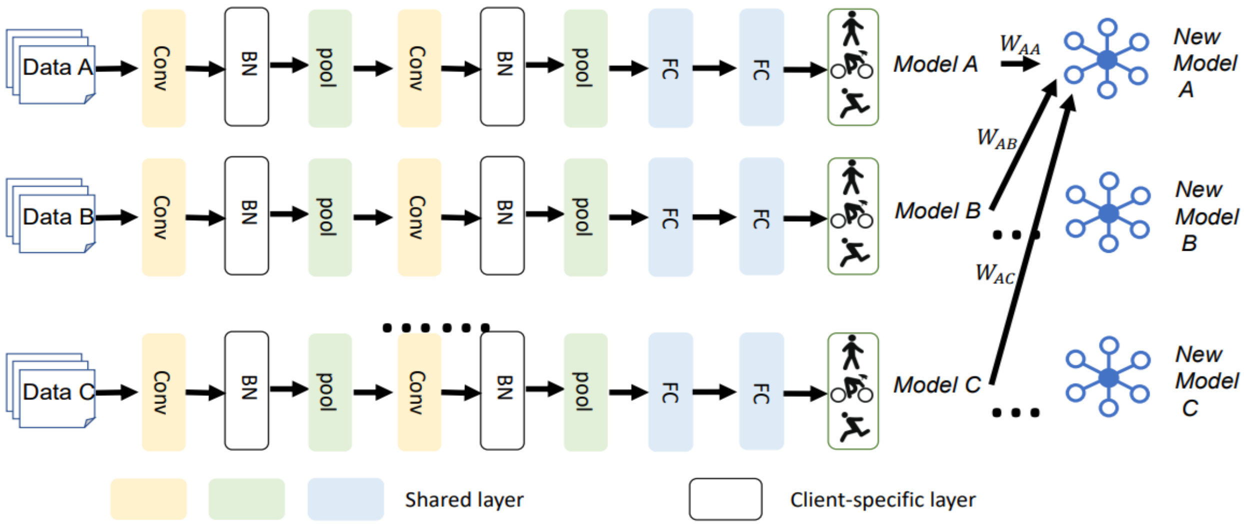
- Introduction to Federated Learning
- Federated Learning on Non-IID Data Silos
- Beyond Supervised Federated Learning
  - Semi-supervised
  - Self-supervised
- Personalized Federated Learning

Personalized FL (PFL)

- In real-world scenarios, a customized model for each client would be desirable
    - E.g., advertising recommendation system customized for different users
- What to personalize in FL?
    - Personalized aggregation strategy
    - Personalized layers
    - …

Personalized Aggregation Strategy

- BN layers (white) describe client-specific data distribution (μ, r)
- Shared layers (colored) are uploaded for weighted aggregation
  - Clients with similar distribution would contribute more

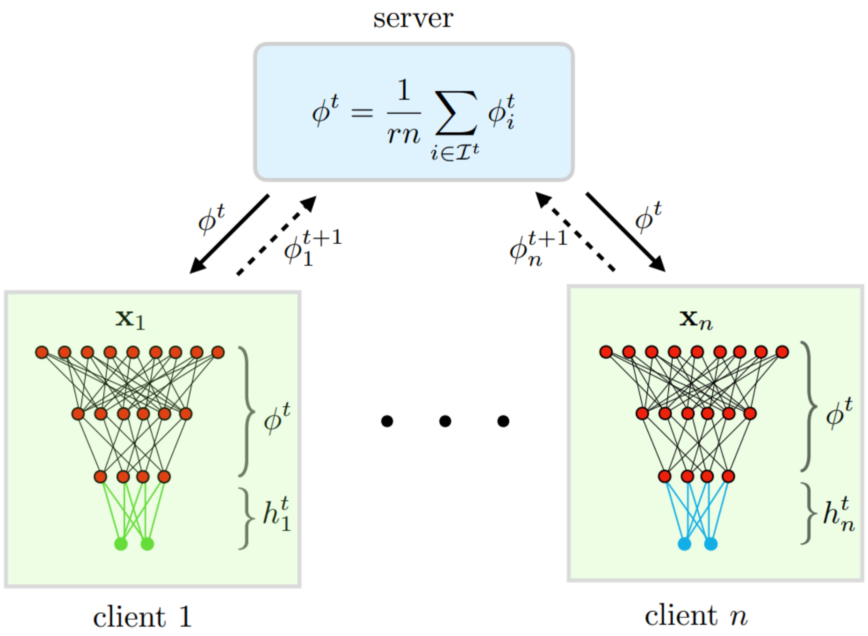


$$w_{i,j} = 1/d_{i,j}$$

$$d_{i,j} = \sum_{l=1}^{L} (||\boldsymbol{\mu}^{i,l} - \boldsymbol{\mu}^{j,l}||^2 + ||\sqrt{\mathbf{r}^{i,l}} - \sqrt{\mathbf{r}^{j,l}}||_2^2)^{1/2}$$

Personalized Layers

- FedRep
- Each personalized model contains
  - Shared global feature extractor φ: $R^d \to R^k$
  - Personalized classification head h: $R^k \to y$
- Local update for client i:
  1. Fix φ^t, train h_i^t for τ epochs
  2. Fix h_i^t, train φ_i^t for 1 epoch
- Server aggregation:
  - Collect $\varphi_1^t, \ldots, \varphi_n^t$ from clients
  - $\varphi^{t+1} = Avg(\varphi_1^t, \ldots, \varphi_n^t)$



What We've Covered This Semester…

- NN & CNN
- Object Detection & Semantic Segmentation
- Generative Model & GAN
- Diffusion Model
- Transfer Learning (Domain Adaptation & Generalization)
- RNN & Transformer
- Vision & Language
- Few-Shot Learning
- 3D Vision
- Federated Learning