PHILIPS
www.philips.com

MIDL 2020

Continual Learning
for Domain Adaptation in Chest X-ray Classification

Matthias Lenga, Heinrich Schulz, Axel Saalbach
Philips Research Hamburg

arxiv.org/abs/2001.05922

innovation you

Domain shift in CXR classification

- For Chest X-ray classification DL performance is on par to radiologists [Majkowska et al., 2019]

- Performance degradations were reported, when applied to data from a (unseen) target domain [Zhang et al., 2019; Yao et al., 2019]

- **Example:** DenseNet121 (ChestX-ray14 → MIMIC-CXR)
  PTX: 0.86 → 0.77 mean AUC
  CMG: 0.88 → 0.76 mean AUC

- **Domain shift:** Data distributions of source and target domain differ
  – hospital specific protocols
  – operator preferences
  – different scanners
  – changing class frequencies
  – errors in labelling

ChestX-ray14

Label	Initial
Atelectasis*	.7730±.0019
Cardiomegaly*	.8806±.0021
Consolidation*	.7468±.0011
Edema*	.8490±.0010
Effusion*	.8308±.0009
Emphysema	.9122±.0054
Fibrosis	.8273±.0044
Hernia	.8918±.0174
Infiltration	.6978±.0011
Mass	.8203±.0041
Nodule	.7587±.0035
Pl. thickening	.7741±.0045
Pneumonia*	.7233±.0024
Pneumothorax*	.8620±.0026
Average	.8106

ChestX-ray14 model on ChestX-ray14 test dataset

MIMIC-CXR

Label	Initial
Airspace opacity	
Atelectasis*	.7797±.0019
Cardiomegaly*	.7603±.0050
Consolidation*	.7654±.0025
Edema*	.8343±.0027
Effusion*	.8913±.0019
Enl. cardio.	
Fracture	
Lung lesion	
No finding	
Pleural other	
Pneumonia*	.6830±.0031
Pneumothorax*	.7691±.0064
Support devices	
Average	.7833

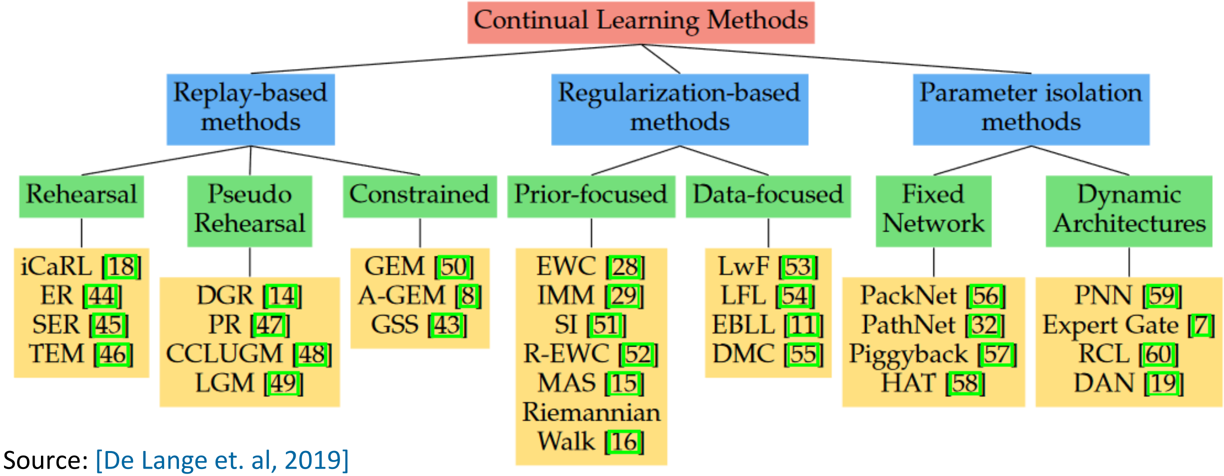
ChestX-ray14 model on MIMIC-CXR test dataset

PHILIPS

Continual Learning

Studies the problem of **learning from a stream of data**:

- **Sequential learning process**: Only small portion of input data from one (or a few) tasks is available at once
- Gradually extend acquired knowledge
- **Learn without catastrophic forgetting**: Preservation of certain model characteristics might be required due to regulatory considerations



Source: [De Lange et. al, 2019]

Regularization-based CL for CXR classification

- Feasibility study focusing on regularization-based methods **EWC** and **LWF**
- These methods **do not** require any data from the source domain (e.g. containing sensitive PHI)

EWC: Assumes a prior distribution on the network weights [Kirkpatrick et al., 2017]

$$\theta_i = \underset{\theta}{\operatorname{argmin}} \underbrace{L(\theta, T_i)}_{\text{current task's loss}} + \underbrace{\lambda(\theta - \theta_{i-1})^\top \Sigma_{i-1}^{-1}(\theta - \theta_{i-1})}_{\text{LL of prior (on NN weights) related to previous task}}$$

$$\Sigma_{i-1}^{-1} = \operatorname{diag}(F_{i-1})$$

$$F_{i-1} := \frac{1}{N_{i-1}} \sum_{j=1,\ldots,N_{i-1}} \nabla_\theta \log p(y_{i-1,j}|\theta_{i-1}, x_{i-1,j}) \, \nabla_\theta \log p(y_{i-1,j}|\theta_{i-1}, x_{i-1,j})^\top,$$

(empirical Fisher matrix of LL related to previous task)

LWF: Adds soft-target regularization to training loss which reflects the behavior of the model associated to the previous task on current task data [Li and Hoiem, 2017]

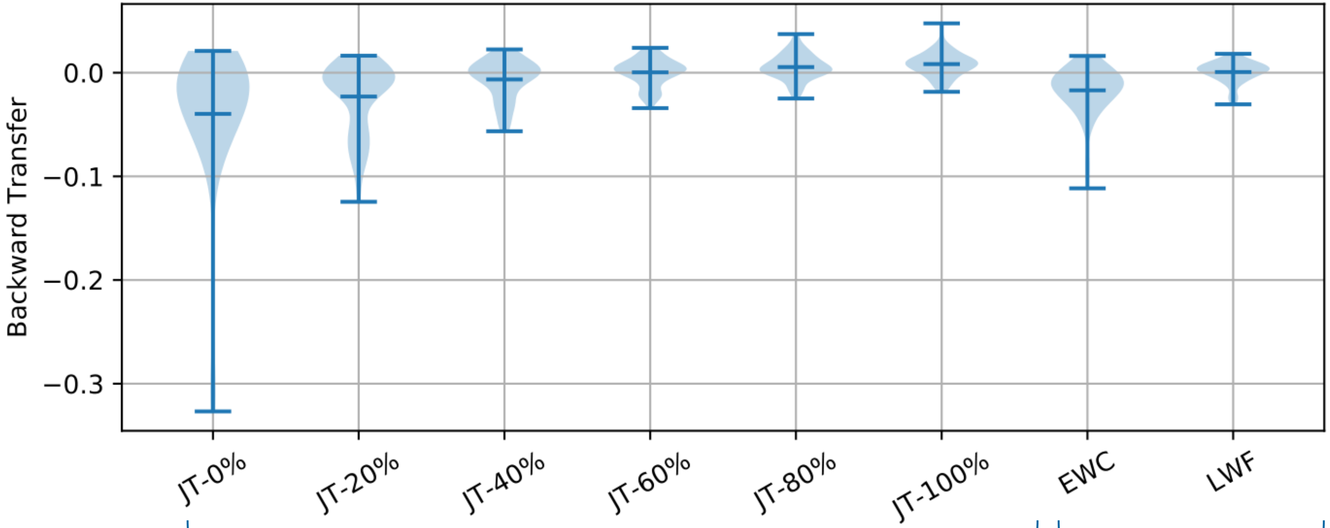
$$\theta_i = \underset{\theta}{\operatorname{argmin}} \sum_{j=1,\ldots,N_i} \underbrace{- \log p(y_{i,j}|\theta, x_{i,j})}_{\text{current task's loss}} - \underbrace{\lambda \log p(\hat{y}_{i,j}|\theta, x_{i,j})}_{\text{soft target regularization}} \qquad \hat{y}_{i,j} := \underbrace{M_{\theta_{i-1}}}_{\substack{\text{Model related to previous task} \\ \text{(e.g. model trained on source domain)}}}(x_{i,j})$$

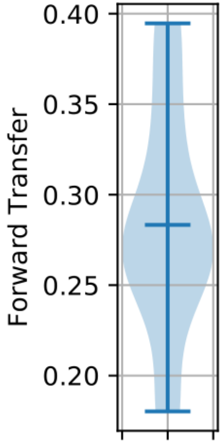
PHILIPS

Quantitative results: Forward & Backward Transfer

<u>**Evaluation:**</u> Joint Training (JT) baseline vs. EWC vs. LWF

- **Setup:** DenseNet121, ChestX-ray14 (source domain) → MIMIC-CXR (target domain)
- Mean AUC after adaption to targeted domain : **JT-k% / EWC / LWF ≈ 0.82**
- **FTW:** measures how good the model generalizes to target domain
- **BWT:** measures model performance on source domain after adaptation to target domain [Lopez-Paz and Ranzato, 2017]





k% of the data from source domain required

No data from source domain required

PHILIPS

Conclusion

- Shifts in the distribution of medical image data across different sites
  - Pre-trained models are often not directly applicable as a result of performance degradations
  - On-site retraining desired but potentially constrained owing to regulatory guidelines

- Investigated the applicability of different **Continual Learning methods for domain adaptation in CXR classification**
  - Adapt to target domain data
  - Preserve source domain performance (avoid “Catastrophic Forgetting”)

- Selected ChestX-ray14 and MIMIC-CXR as distinct domains in order to **simulate a realistic domain shift**
  - Discussion of regularization based CL methods EWC and LWF
  - Continual learning without image / gradient / … information related to source domain (privacy compliant)

- **Quantitative evaluation**: EWC vs. LWF vs. JT, measuring FWT and BWT

- Continual Learning methods for Medical image classification:
  - Provide effective means in order to overcome performance degradations resulting from a domain shift
  - For ChestX-ray14/MIMIC-CXR a **positive Backward Transfer** was obtained using LWF (on average)

arxiv.org/abs/2001.05922

PHILIPS

PHILIPS