SURVEY

Federated Learning in Computer Vision

DONALD SHENAJ, (Student Member, IEEE), GIULIA RIZZOLI, (Student Member, IEEE), AND PIETRO ZANUTTIGH, (Member, IEEE)

Department of Information Engineering, University of Padova, 35131 Padua, Italy

Corresponding author: Pietro Zanuttigh (zanuttigh@dei.unipd.it)

ABSTRACT Federated Learning (FL) has recently emerged as a novel machine learning paradigm allowing to preserve privacy and to account for the distributed nature of the learning process in many real-world settings. Computer vision tasks deal with huge datasets often with critical privacy issues, therefore many federated learning approaches have been presented to exploit its distributed and privacy-preserving nature. Firstly, this paper introduces the different FL settings used in computer vision and the main challenges that need to be tackled. Then, it provides a comprehensive overview of the different strategies used for FL in vision applications and presents several different approaches for image classification, object detection, semantic segmentation and for focused settings in face recognition and medical imaging. For the various approaches the considered FL setting, the employed data and methodologies and the achieved results are thoroughly discussed.

INDEX TERMS Federated learning, computer vision, scene understanding, survey.

I. INTRODUCTION

Large-scale machine learning models deployed in real-world scenarios often require training data sharing on a centralized server, where the actual model optimization takes place. Federated Learning (FL) was initially introduced in [155] as an alternative approach to train a global model across multiple devices while preserving the privacy and decentralization of their respective data.

In particular, machine learning methods for computer vision heavily rely on collecting and storing a huge amount of annotated image data on a central server. Centralizing such data necessitates the transfer of a significant volume of information, resulting in substantial communication overhead. Moreover, centralized data storage poses risks to user privacy and confidentiality, and recent regulations on data privacy prohibit the uploading of sensitive local data to centralized data centers [214].

Federated learning has emerged as a promising solution to address these challenges, as it enables on-device training of visual models. In FL, data remains localized on

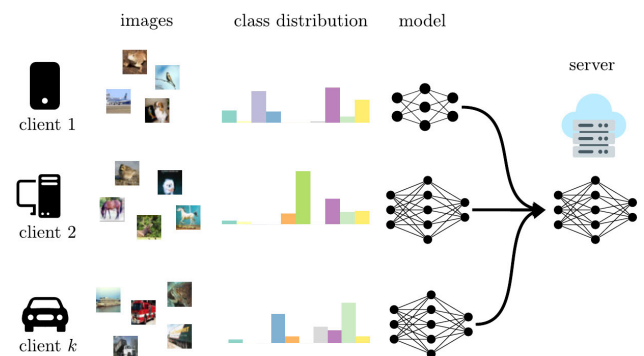The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwanggil.

FIGURE 1. Standard federated learning setting for vision applications, whereas devices are heterogeneous among each other in terms of computational capacity, number of image samples and statistical distribution of data.

individual devices, and the collaborative training process involves exchanging model parameters instead of raw data. This approach opens up practical applications and opportunities for privacy preservation and effective management of sensitive data, such as medical images or facial pictures [103], [127], [258]. The fundamental framework,

initially introduced by McMahan et al. [155] and depicted in Fig. 1, has been extended in various directions and explored in many works (refer to Section III for more details). Most studies on federated learning focus on its theoretical and communication aspects [2], [105], [128], [145]. Nevertheless, Federated Learning has recently attracted a wide interest when applied to computer vision tasks, ranging from image classification [26], [88], [269] to semantic segmentation [24], [62], [161], and object detection [96], [147], [200].

As already pointed out, real-world computer vision settings typically deal with huge amounts of data and often with critical privacy issues, and the distributed and privacy-preserving nature of FL makes it an extremely good candidate to solve these problems. As an example FL sparkled a large interest in human-centric tasks like face recognition [3], [157] and medical imaging [18], [129], [192], [236], [238], where privacy is a key requirement.

After discussing the challenges that arise in Section II, we overview the basic FL theoretical frameworks in Section III, while the different settings for FL in computer vision will be detailed in Section IV. The main challenges can be tackled using different techniques, e.g., knowledge distillation, representation and prototype learning and different aggregation strategies. Section V introduces the main insights and ideas that allow to efficiently tackle vision tasks in a federated learning environment. This section empowers readers to focus on the methodologies that align with their interests and encourages further exploration of specific works in the subsequent sections. Many different approaches apply these ideas to well-known computer vision problems like image classification, object detection, semantic segmentation and face recognition and it is widely used also in the medical imaging field. The most relevant approaches are presented in Section VI, while some benchmark comparisons can be found in Section VII. In Section VIII, we provide an overview of potential current and future trends in Federated Learning and finally in Section IX we draw the conclusions.

II. MAIN CHALLENGES IN FL

Compared to standard supervised learning, the Federated setting introduces new challenges: the clients usually have different hardware capabilities (*system heterogeneity*), a different amount of samples to process (*data imbalance*), and their data statistics may be different (*statistical heterogeneity*). Furthermore, the system should be efficient in terms of communication. Finally, any proper FL algorithm must not break the *privacy preservation* of clients' data.

Summarizing, the main challenges are:
- *Statistical heterogeneity*: clients' data is highly non-IID, i.e., their statistics may not be indicative of the global distribution as they reflect the specific clients' usage.
- *Model heterogeneity*: clients can have different models.
- *Communication cost*: transmitting model weights from clients to server determines a latency in training and sharing a huge quantity of information could cause
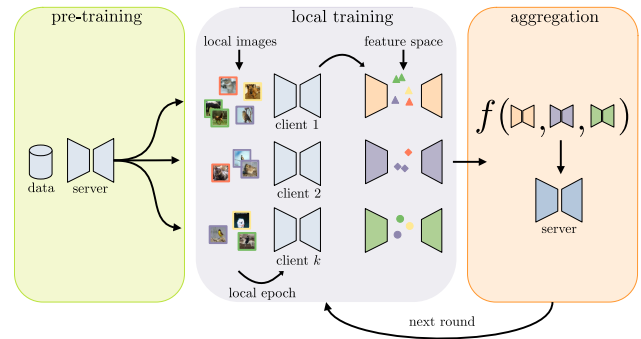

FIGURE 2. Pipeline of a generic FL system for computer vision tasks.

network overloading [74]. The communication bottleneck issue can also be enhanced by the devices' limited or intermittent connectivity due to battery consumption constraints, faults, or data unavailability.
- *Convergence time*: reducing the time required to complete the training is a key target.
- *Privacy and Security*: client-server communications should not contain sensitive information and FL systems should prevent the server from accessing clients' local data [21], [127].
- *Catastrophic forgetting*: inconsistent predictions can arise between subsequent training rounds [117].
- *Unlabeled data at clients*: in some settings assuming that clients have access to ground truth data is not realistic.

III. OVERVIEW OF FEDERATED LEARNING

The first approach for FL, that represent also the baseline algorithm, is Federated Averaging (FedAvg) [155]. FedAvg uses a client-server architecture to perform collaborative learning in synchronous rounds. The server (or aggregator) broadcasts the global model's current parameters to some of the clients at the start of each round.

Each participant locally trains the model on its own private data and sends back the updated model parameters to the server. The server collects these updates and combines them using a specified strategy, *i.e.*, a weighted average based on the amount of local data each participant has. The combined updates are applied to the global model as a "pseudo-gradient" [180]. This process can be repeated for multiple rounds of FL by distributing the updated global model to participants. A summary of the pipeline is shown in Fig. 2.

Despite proving solid empirical results in IID and balanced settings, FedAvg performances degrade when dropping this assumption [269]. Subsequent research works took into consideration that heterogeneous data characterize real-world federated learning and simulate this scenario by the use of realistic per-user data splits [88], [269]. To this extent, they address two types of distribution shift: non-identical class distribution, where the visual distribution of classes differs by device, and imbalanced client data sizes, where the number of samples available for training varies for each client.

FedProx [128] can be viewed as a generalization and re-parametrization of FedAvg. Theoretically, FedProx provides convergence guarantee when learning over data from non-identical distributions (statistical heterogeneity), and while adhering to device-level systems constraints by allowing each participating device to perform a variable amount of work (systems heterogeneity). Different devices in federated networks often have different resource constraints in terms of computing hardware, network connections, and battery levels. Therefore, it is unrealistic to force each device to perform a uniform amount of work (i.e., running the same number of local epochs), as in FedAvg. FedProx allows for variable amounts of work to be performed locally across devices based on their available systems resources and then aggregates the partial solutions sent from the stragglers (as compared to discarding updates from these devices). Moreover, a proximal term has been added to the local subproblem to effectively limit the impact of local variable updates.

SCAFFOLD [105] aims for a faster convergence and for a reduction of the so-called ''client-drift'' in local updates. SCAFFOLD estimates the update direction both for the server model and for each client and the difference between the two is then an estimate of the client-drift which is used to correct the local update. It can be seen as an improved version of [190], introduced for distributed parallel optimization.

MIME [104], extends SCAFFOLD to all types of functions and applies global momentum locally since it proved to be more effective than using server-only-momentum strategies. While MIME has demonstrated good performance, it can be detrimental to training efficiency as it requires computing the gradient twice at each local step [220].

AdaBest [212], proposes an adaptive algorithm that estimates drift across clients, using less storage and communication bandwidth, as well as lower compute costs. Additionally, it improves stability by constraining the norm of estimates for client drifts, making it more practical for large-scale FL.

FedNova [219] provides a general framework to analyze the convergence of heterogeneous federated optimization algorithms. The authors focus on understanding the solution bias and the convergence slowdown due to objective inconsistency. Moreover, FedNova presents a normalized averaging method that aims to eliminate objective inconsistency while preserving fast error convergence. They claim that sophisticated approaches such as FedProx [128] and SCAFFOLD [105], designed to handle non-IID local datasets, can be used to reduce (not eliminate) objective inconsistency to some extent, but these methods either result in slower convergence or require additional communication and memory resources. In FedNova the locally normalized updates (that are just re-scaled versions of cumulative local changes) are averaged instead of the local changes.

The insights introduced by these works have then been applied in many papers targeting federated learning for different vision tasks, ranging from image classification to object detection and semantic segmentation, and again to more human-centric tasks such as face recognition and medical imaging. The various FL approaches for vision applications are detailed in Section VI.

IV. FEDERATED LEARNING SETTINGS

This section overviews the main FL settings of interest in computer vision.

- **Standard FL**: the standard setting of FL with non-IID data scattered across different clients [269], with the objective of training a single global model.
- **Heterogeneous FL**: clients can have different models, different computation resources and different communication capabilities [48].
- **Personalized FL**: in standard FL client data are used to train collaboratively a global model. In personalized FL instead, each client aims at optimizing a specific model to be deployed on its own data [76].
- **Clustered FL**: in this setting subsets of clients share some common characteristics, *i.e.*, their data belong to the same subdomain, and they can be grouped with the objective to build a personalized model for the cluster that works well in the corresponding subdomain [25], [60], [64], [69], [188], [193].
- **Continual FL**: Continual learning enables a model to learn from a never-ending stream of data, without the need to retrain the model from scratch every time new data or new tasks become available. Class-incremental continual learning, allowing clients to learn from non-stationary data and learn new tasks over time has been explored in the FL setting [50], [51], [90], [149], [247], [252]. An asynchronous federated continual learning (AFCL) setting has been introduced in [194], where the learning of multiple tasks happens with different orderings and in asynchronous time slots.
- **Federated Domain Adaptation**: Domain adaptation (DA) deals with the statistical heterogeneity between a source dataset, used for training, and a target dataset to which the model should be adapted. It aims at transferring the learning performed on source data to the target set. Interesting DA settings are UDA (Unsupervised DA), where no supervision on the target dataset is available and SFDA (source-free DA) where source data is not available during adaptation.

  Within the context of FL, DA has been approached from diverse perspectives to target various applications, e.g. face recognition [276]. Examples include its integration with adversarial learning techniques [175], the consideration of individual clients as distinct target domains, denoted as multi-target DA [245], the use of each client as a source domain with an additional target one [98], and finally in the SFDA setting [193].

V. METHODS

The purpose of this section is to present diverse insights and ideas by emphasizing their broad utility within the context of Federated Learning. We aim to offer readers a comprehensive

grasp of the prevailing ideas and concepts, laying the foundation for the subsequent task-specific discussion in Section VI. Since FL needs to solve a wide range of challenges in many different settings, a large number of different strategies have been proposed both for the local training procedures on the clients and for the model aggregation at server side. We present the main families of strategies and we briefly discuss how the different provisions are implemented into FL approaches. In Figure 3 an overview of the family of strategies is presented, grouping them into client and server-side techniques.

Pre-training is a transfer learning technique that has been widely used to reduce training time and improve final accuracy in large-scale deep learning. Even though the standard FL setup [155] does not consider an initial pre-training step at server side, the nature of FL makes it the perfect candidate for such strategies. Starting from a pre-trained model significantly reduces the impact of heterogeneity of data (data is IID during pre-training while clients have non-IID data), thus allowing to train clients with more local epochs (and less rounds) since the client drift is more limited [170]. This enables the learned global models under different clients' data conditions to converge to the same loss basin and makes global aggregation more stable.

The first systematic study on pre-training for FL is [34], which uses five different image datasets. They consider two cases, when the server has a pre-trained model or data for pretraining from real-world dataset (e.g., Imagenet), and when the server has no data, resorting to image generation techniques such as random generative models or fractals [7], [106]. Fractal pre-training is also used in [194].

Another approach is to force clients to jointly learn to fuse the representations generated by multiple fixed pre-trained models rather than training a large-scale model from scratch [204]. Pre-training on a supervised source dataset at the server side can also be used to tackle real-world setups, where clients have unsupervised data and require adaptation [193], [245].

Image Augmentation and Style Transfer techniques have been used to improve domain generalization in FL [35], [62], [142]. Data augmentation techniques can be used to improve generalization thus mitigating issues due to data heterogeneity on the clients, and in addition, allow to improve accuracy on new unseen clients [46]. Furthermore, in clustered FL clients can be clustered according to their style to improve performances [193].

In a practical FL application, the models trained with local datasets are likely to establish decision rules on biased attributes (*i.e.*, fur color for animals), which hinders the aggregated model's ability to learn a suitable representation for classification. Hence, [239] propose to learn a Bias-Eliminating Augmenter at each client, able to generate bias-conflicting samples thus reducing the bias in local updates.

Knowledge Distillation (KD) was introduced in [84] for model compression: it allows to transfer knowledge from a larger network (teacher) to a smaller one (student). It has
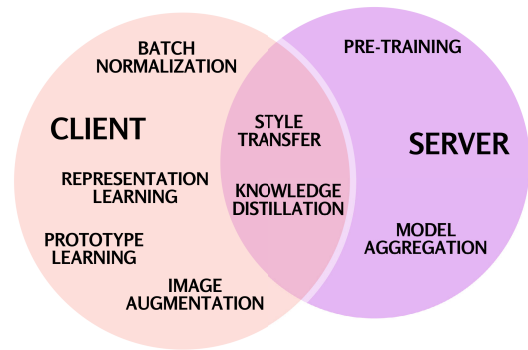

FIGURE 3. Overview of the main strategies employed at client and server side for Federated Learning.

been widely used in continual learning and recently has been increasingly employed in FL algorithms [162] to reduce catastrophic forgetting, tackle data heterogeneity, and enable model heterogeneity.

At client side, KD can be used to tackle catastrophic forgetting, by forcing the prediction of the current local model to be consistent with the global model of the previous step, preserving inter-class semantic consistency across different incremental tasks [50], and balancing knowledge from others while boosting both inter and intra-domain performance [90].

KD methods can address data heterogeneity, both at client and server sides. At the client side, global knowledge is used to control the client drift via on-device regularizers [80], [81], [117], [246] or using synthetically-generated data [275]. On the server side, instead, the global model can be rectified via ensemble distillation of a proxy dataset [32], [138], [187] or using a generator network [215], [262], [263].

Finally, KD can be used without resorting to a public dataset or a generative model at the server, by applying it on averaged data representation and soft predictions (referred to as ''hyperknowledge'') to improve both personalized and global model performances [31].

Representation Learning techniques aim at improving a downstream task by enforcing meaningful distribution of the training data representations (features).

A first set of approaches aims at aligning features across clients during local training to address data heterogeneity on clients [271] and domain generalization [261]. Contrastive Learning is used in [204] to assist local training and achieve higher model performances [124], [166], [259].

The permutation invariance property of neural networks leads to neuron misalignment across local models, therefore [133] binds neurons in positions and pre-aligns parameters for better coordinate-wise parameter averaging, while [218] matches the neurons of client models before averaging them, and permits global model size adaptation.

A different approach is proposed in [52], where the non-IID issue is tackled by constraining learned representations of data points to be on a unit hypersphere shared by clients.

A feature-oriented model structure adaptation method is exploited in [254] to ensure explicit feature allocation. Applying the structure adaptation to collaborative models, matchable structures with similar feature information can be initialized at the very early training stage. Then, during the federated learning process, a feature-paired averaging scheme is used to guarantee aligned feature distributions and avoid feature fusion conflicts under either IID or non-IID scenarios.

Prototype Learning techniques aim to learn a compact representation of features, called prototypes, which can be used as representatives of target classes. Hence the large adoption of these methods for continual learning. In FL, a first possible strategy is the correction of the client drift by computing client deviations using margins of prototypical representations learned on distributed data. These margins can be exploited to drive the federated optimization, via an attention mechanism [161], to address system and statistical heterogeneity. Prototypes can also be transmitted instead of model weights [203], to reduce communication cost, to allow clients to learn a more customized local model and to be more robust to gradient-based attacks [28], [273] since high-level statistic information (prototypes) are more privacy-compliant than raw features. Prototype-based contrastive losses can be used to make local objectives consistent with the global optima and tackle the non-IID data issue [166]. Finally, since global prototypes could be biased towards the dominant domain distribution in the presence of numerous domains (domain shift), a combination of cluster prototypes and averaged prototypes can be employed [91].

Batch Normalization (BN) is a key tool in deep learning. However, standard batch normalization in FL is not very effective (according to [54] this is due to the fact that statistics of channels change significantly across clients). Various customized versions of BN layers have been proposed for federated learning empirically showing better performance.

Firstly, it is possible to update the client batch-norm layers locally without communicating them to the server, in order to achieve personalized FL [132]. Local-statistic batch normalization (BN) layers can be exploited, resulting in collaboratively-trained, yet center-specific models [8]. This strategy improves robustness to data heterogeneity while also reducing the potential for information leaks by not sharing the center-specific layer activation statistics. Group normalization (GroupNorm) can improve the convergence of FL [86]. Nevertheless, GroupNorm is an instance-based normalization scheme, which is highly sensitive to the noise on data samples. The privacy concern can be tackled by not tracking running estimates and simply normalizing batch data [48].

Model Aggregation: variations of the standard FedAvg [155] aggregation have been introduced for a variety of reasons. With FairAvg [160] each user contributes equally to the aggregated model (simple mean), increasing accuracy and convergence rate. Layered-wise aggregation schemes are used to enable personalized [153] and clustered FL [193]. A selective model aggregation scheme has been used to reduce the influence of varying image quality and computation capabilities in vehicular clients [248]. The work in [29] presents elastic aggregation as a novel approach in federated learning for addressing gradient dissimilarity in heterogeneous scenarios. By leveraging parameter sensitivity, this technique improves convergence behavior and enhances the effectiveness of federated learning. In [264], to enable domain generalization to unseen domains, the global model dynamically calibrates the aggregation weights minimizing the variance of the generalization gap. In [248] FL is applied to Vehicular Edge Computing, and a selective model aggregation approach is proposed to reduce the influence of the diversity of image quality and computation capability of the different vehicular clients. Finally, FedFusion [55] introduces a variational autoencoder method for learning the optimal parameters of distribution fusion components based on observed information. These parameters are then utilized to optimize the federated model aggregation in the presence of non-IID data. Notice that, despite the server not having direct access to private data, there are statistical characteristics embedded within the received model parameters (normalization layers) from which the server can infer the local distributions.

Privacy and Security are implemented in FL by enabling local training without the need for the exchange of critical data between the server and the clients. This safeguards the clients' data from potential eavesdropping by hidden adversaries. However, it is important to note that adversaries may still be able to collect private information by analyzing the differences in trained network weights or other parameters transmitted by the clients [150], [196], [226]. To mitigate the risk of information leakage, a first solution is to perturb in some way the transmitted data to make harder the eavesdropping task without affecting too much the model performances. Techniques for this task include differential privacy (DP) [56], additive perturbation and multiplicative perturbation [251]. As an example, the authors of [227] propose a framework based on differential privacy in which each client adds noise to its locally trained parameters before uploading them to the server for aggregation. An alternative is to use cryptographic techniques such as homomorphic encryption [183], secret sharing, and secure multi-party computation. For example in [151], the model updates undergo encryption using an aggregated public key before being shared with the server for aggregation. To decrypt the updates, collaboration among all participating devices is necessary. This approach demonstrates resilience against potential attacks from the participants and collusion attempts between the participating devices and the server. Many other approaches for this task have been proposed, for comprehensive surveys focusing on privacy and security in FL see [165] and [251].
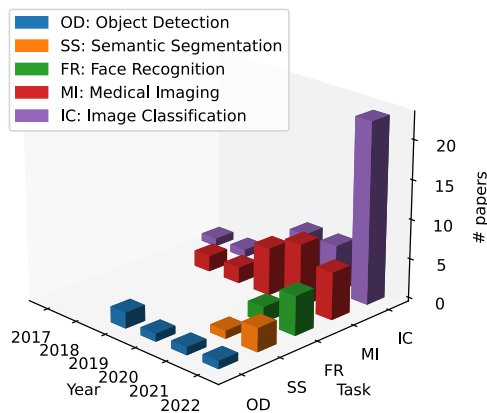
FIGURE 4. A timeline that illustrates increasing research focus on these vision tasks each year. The plot shows the number of papers in each year for each of the 5 tasks in Section VI. Notice that we assigned papers tackling face recognition and medical imaging to these tasks independently of the underlying computer vision task.

VI. FL IN COMPUTER VISION: TASKS AND APPROACHES

FL has shown great potential in computer vision applications, particularly when dealing with large and sensitive datasets, and it has been applied to a variety of computer vision tasks. However, each of these tasks poses unique challenges when it comes to federated learning. For instance, semantic segmentation requires pixel-level annotations, which are usually scarce and expensive to obtain. Additionally, the distribution of the data across different clients may be heterogeneous, which can lead to performance degradation if not addressed appropriately. Despite these challenges, recent works have demonstrated promising results in applying federated learning to computer vision tasks [24], [62], [161], [193], and ongoing research aims at improving the performance and scalability of FL in this domain. An overview of the distribution of the papers on the different tasks is shown in Fig. 4.

Throughout this section, we address the different tasks (i.e., the three main image understanding tasks - Classification, Object Detection and Semantic Segmentation - plus two widely considered application fields, Face Recognition and Medical Imaging) and present the main approaches developed to tackle each of them. This structure allows to focus on specific application domains and facilitates an in-depth exploration of the strategies presented across different research papers. We decided to subdivide the works on the basis of the considered task since, although many FL techniques can be applied in principle to different tasks, the majority of current research works focus on a single vision task. In the remainder of this section, we present the various approaches, while some performance comparisons will be shown in Section VII.

A. IMAGE CLASSIFICATION

Image classification, *i.e.*, classifying an entire image into one of the possible semantic classes, is a fundamental task in computer vision. Several backbone models have been used

for federated image classification tasks, from the widely used ResNet [79] model to lightweight architectures. As an example, MobileNet [186] runs efficiently on mobile devices with limited computational power, thus being well suited for real-world FL settings. Simple datasets have been widely used such as the 10 classes color images CIFAR-10 [112] dataset, its extended version with 100 classes (CIFAR-100) or the grayscale images of handwritten numbers from MNIST [116] or EMNIST [42]. The data inside these datasets need to be split in order to simulate the non-IID distribution across clients, which is a key characteristic of FL data. To this aim, FL researchers have proposed new data splits to simulate real-world scenarios. As an example, Zhao et al. [269] suggest two partitions of MNIST and CIFAR-10: in the first case each client receives data belonging only to a class, while in the other each client is randomly assigned 2 partitions from 2 classes.

1) TACKLING STATISTICAL HETEROGENEITY

As for most tasks, the impact of non-IID data at the clients is one of the key issues to be tackled by approaches targeting image classification in the federated setting. The work of [269] shows that the accuracy is reduced significantly when data is non-IID, which can be explained by the model weights divergence at each device. As a solution, they propose sharing a small subset of image data across all devices. A simple modification of FedAVG, called FedAvgM [87] addresses the non-IID issue by introducing momentum at the server side. The work of [182] introduces a FL framework robust to affine distribution shifts. They propose a fast and efficient optimization method and provide convergence and performance guarantees via a gradient Descent Ascent (GDA) method. In [124] the authors address the non-IID issue observing that the global model trained on a whole dataset is able to learn a better representation than the local model trained on a skewed subset. They propose model-contrastive learning, to force the representation learned by the local and global models to be aligned. A probabilistic FL framework is considered in [257], where the global model is built with a Bayesian non-parametric strategy that allows the local parameters to match existing global ones, otherwise new global parameters are created if existing ones are poor matches. FedAlign [156] studies the data heterogeneity challenge of FL focusing on local learning generality. They resort to a distillation-based regularization method to align the Lipschitz constants, promoting smooth optimization and consistency within the model. A general multi-stage FL framework (FedCorr) is proposed in [237] to tackle data heterogeneity, with respect to both local label quality and local data statistics. More specifically, an adaptive local proximal regularization term based on estimated local noise levels is introduced. The work of [133] focuses on solving the non-IID challenge in FL limiting the misalignment across local models. FedPVR [120] starts from the observation that there is a stronger misalignment between models in the last few layers of the network

TABLE 1. Comparison of Federated Learning approaches for image classification. The considered datasets are: MNIST [116], EMNIST [42], FEMNIST/CelebA [26], FashionMnist [230], BelgiumTSC [209], Natural Disaster/Waste [4], CIFAR-10/100 [112], TinyImageNet [115], ISIC 2018/HAM1000 [210], Mixed Digits [132], SVHN [167], GTSRB [199], CINIC-10 [45], PACS [121], Office-Home [213], Camelyon17 [15], iNaturalist [211], Clothing1M [232], Commercial Image Sources [95], DomainNet [174], TerraInc [17], Colored MNIST [10], Corrupted CIFAR-10 [83], Collage CIFAR-10 [207], Office-Caltech [70], SST-2 [198], AGNews [266], SUN397 [231], MedMNIST [243].

Paper	Model	Dataset	# Clients	# Rounds
[155]	2-layer NN, 2-layer CNN	MNIST, CIFAR-10	100	1000, 3000
[128]	log. regression	MNIST, FEMNIST, FMNIST	200, 1000	100, 200
[105]	2-layer NN, log. regression	EMNIST	100	70
[136]	LeNet	MNIST, TinyImageNet	8	100
[87]	2-layer CNN	CIFAR-10	100	10k
[104]	log. regression, 2-layer CNN	EMNIST	3400	1000
[219]	VGG-11	CIFAR-10	16	200
[269]	2-layer NN, CNN	MNIST, CIFAR-10	10	500
[257]	1/2 layer NN	MNIST, CIFAR-10	-	-
[101]	CNN	FEMNIST-62	3400	500
[248]	2-layer CNN	MNIST, BelgiumTSC	-	-
[218]	LeNet, VGG-9	MNIST, CIFAR-10	16	16
[257]	ResNet-18	CIFAR-10	100	1000
[4]	ResNet-101	Natural Disaster, Waste	-	-
[182]	AlexNet, Inceptionet, mini-ResNet	MNIST, CIFAR-10	10	100
[260]	1-layer CNN	MNIST	4	100, 300, 1000
[25]	2-layer CNN	CelebA, FEMNIST	9343, 3550	100, 1000
[71]	ResNet-8	CIFAR-10/100	-	-
[124]	ResNet-50, 2-layer NN, CNN	CIFAR-10/100, TinyImageNet	50, 100	100, 20
[261]	2-layer NN, 2-layer CNN	MNIST, CIFAR-10	100	-
[254]	VGG9, VGG16, MobileNet	CIFAR-10/100	10-100	200
[166]	ResNet-50, CNN	CIFAR-10/100, TInyImageNet	50, 100	100
[177]	AlexNet, VGG16, Inception	Skin Cancer MNIST: HAM1000	4	10, 20, 30
[271]	CNN, ResNet-18	EMNIST, FMNIST, CIFAR-100, Mixed Digits	10	200
[203]	2-layer CNN, ResNet-18	MNIST, FEMNIST, CIFAR-10	20	100
[160]	2-layer CNN	MNIST, FEMNIST	1000, 200	200, 400
[161]	2-layer CNN, 4-layer CNN	MNIST, FEMNIST, CelebA	200, 400	
[133]	x-layer NN, VGG11, ResNet-20	MNIST, FEMNIST, SVHN, GTSRB, CIFAR-10/100 , Cinic10	355	2000
[212]	2-layer NN, 2-layer CNN	EMNIST, CIFAR-10/100	100, 1000	1200, 2000
[35]	ResNet, DenseNet121	PACS, Office-Home, Camelyon17	domains ↔ clients	100, 500
[206]	3-layer NN, 3-layer CNN	FMNIST, CIFAR-10	100	500, 2000
[24]	LeNet, ResNet-18	CIFAR-10/100	100	5000
[153]	CNN	CIFAR-10/100, EMNIST, FMNIST	10, 100	600, 2500
[135]	2-layer CNN, ResNet-18	SVHN, CIFAR-100, ISIC2018	10	1000
[33]	CNN, VGG11, ResNet-8/20	CIFAR-10/100, FMNIST, EMNIST	20, 100, 2185	100
[34]	ResNet (18,20,32,44,56)	CIFAR-10, TinyImageNet, iNaturalist	10, 100, 136	100
[156]	ResNet-56	CIFAR-100	16	25

TABLE 1. *(Continued.)* Comparison of Federated Learning approaches for image classification. The considered datasets are: MNIST [116], EMNIST [42], FEMNIST/CelebA [26], FashionMnist [230], BelgiumTSC [209], Natural Disaster/Waste [4], CIFAR-10/100 [112], TinyImageNet [115], ISIC 2018/HAM1000 [210], Mixed Digits [132], SVHN [167], GTSRB [199], CINIC-10 [45], PACS [121], Office-Home [213], Camelyon17 [15], iNaturalist [211], Clothing1M [232], Commercial Image Sources [95], DomainNet [174], TerraInc [17], Colored MNIST [10], Corrupted CIFAR-10 [207], Office-Caltech [70], SST-2 [198], AGNews [266], SUN397 [231], MedMNIST [243].

Paper	Model	Dataset	# Clients	# Rounds
[170]	SqueezeNet, ResNet-18	CIFAR-10, FEMNIST	100, 3150	1000, 1082
[201]	2-layer CNN	MNIST, EMNIST, CIFAR-10	20, 40, 60	1000, 2000
[61]	ResNet-10/12, ShuffleNet, MobileNetV2	CIFAR-10/100	4	40
[237]	ResNet-18/34/50	CIFAR-10/100, Clothing1M	50, 100, 500	955, 910, 102
[67]	CNN, ResNet-18	EMNIST, FMNIST, CIFAR-10/100, TinyImageNet	100, 500, 20	35, 86, 43, 53, 39
[182]	AlexNet, InceptionNet, mini-ResNet	MNIST, CIFAR-10	10	100
[94]	ResNet-18	CIFAR-10/100	10, 50, 100	200
[95]	VGG-19	Commercial Image Sources	8	50, 150
[194]	ResNet-18	CIFAR-100	50, 100, 500	5000
[118]	ResNet-18	CIFAR-10/100, FEMNIST	100	4000
[29]	2-layer CNN, ResNet-20,	MNIST, FEMNIST, CIFAR-10/100, CINIC-10	500, 1000, 34000	20, 200, 1000, 4000
[264]	ResNet-18/50, AlexNet	PACS, Office-Home, DomainNet, TerraInc	4, 6	40
[235]	3-layer CNN	MNIST, CIFAR-10, CIFAR-100	10	20
[239]	LeNet, ResNet-18	Colored MNIST, Corrupted CIFAR-10, Collage CIFAR-10	10	100
[148]	LeNet-5, VGG-11, ResNet-20	MNIST, CIFAR-10/100, TinyImagenet	100	500, 1000, 2000
[123]	ResNet-18, ViT-Tiny	SVHN, CIFAR-10/100, FMNIST, ISIC 2018	10	100, 200, 400
[178]	2-layer CNN, ResNet-18	CIFAR-10/100	30, 50	200
[91]	ResNet-10	Office Caltech, MNIST, USPS, SVHN, SYN	10, 20	100
[215]	ResNet-11	MNIST, EMNIST, FMNIST, SVHN	20	60
[120]	ResNet-8, VGG-11	CIFAR-10/100	10	80
[137]	VGG-9, LeNet-5	CIFAR-10, FMNIST, SVHN	20, 100	2000
[92]	EfficientNet-B4, ResNet-101	CIFAR-10/100, ImageNet, SST-2, AGNews	50, 100	90, 400
[195]	ResNet-18	EMNIST, CIFAR-10/100	100, 200, 300, 500, 700	200
[272]	2-layer NN, LeNet, 6-layer CNN	FMNIST, CIFAR-10/100, SUN397	50, 100, 200	100
[108]	4-layer CNN	CIFAR-10, SVHN, MedMNIST	10	100
[55]	ResNet18 LeNet, DenseNet-121, MobileNetV2, BiLSTM	MNIST, FMNIST, CIFAR-10	20, 50, 100	100
[179]	ResNet-18, WideResNet28-10, ResNet-50	CIFAR-10/100, TinyImageNet	20, 100	1000

compared to the rest of it. To address this issue, the authors proposed a partial variance reduction that focuses on aligning the local models specifically in the last layers (differently from SCAFFOLD which performs it on all layers).

2) TACKLING MODEL HETEROGENEITY

In real-world scenarios, clients exhibit variations in their computational capabilities, leading to discrepancies in the complexity of their models. Therefore, several studies have

been conducted to enable FL in the context of such heterogeneous models. Lin et al [138] propose an ensemble distillation for model fusion (FedDF). The server collects the class scores of the public dataset on each client model and calculates the average value as the updated consensus.

A scenario with noisy and heterogeneous clients is considered in [61]. As in the previous work, public data is used to deal with model heterogeneity, and a correction loss computes the optimal weighted combination of the selected client outputs, reducing the contribution of the noisy ones.

Diao et al. [48] present HeteroFL, a method to dynamically allocate a subset of global model parameters as local model parameters, considering the capabilities of the local clients. Ilhan et al. [92] introduce ScaleFL, a framework that adjusts the size of a deep neural network based on available resources. It achieves this by allowing the model to make predictions at different stages with early exit points. This enables the selection of the best-fit models for efficient training on distributed client devices. It also incorporates self-distillation, which involves leveraging predictions from different early exits to improve knowledge transfer among subnetworks.

3) PERSONALIZED FL

Some works instead of aiming at the construction of a unique global model, focus on the construction of personalized models optimized for the data of a specific client or set of clients. The work of [101] draws a connection between two widely used FL models and model agnostic meta-learning algorithms (MAML), and interpret existing FL algorithm in the light of existing MAML algorithms. They show empirically that FedAvg is already a meta-learning algorithm, optimizing for personalized performance, as opposed to the quality of the global model. A model trained using a standard centralized optimization method is harder to personalize, compared to one trained using Federated Averaging, supporting the first claim. However, concentrating solely on personalization might lead to a biased personalized FL (pFL) result, where clients with lower performances suffer from the large client deviation. For this reason, in [179] pFL is addressed by dividing the layers into personalized and universal, where the personalized layers extract personalized attributes and the universal layer universal information. Unlike most of the previous work, in [33] the performance is improved simultaneously for pFL and standard FL, showing that strong personalized models emerge from the local training of generic FL algorithms due to implicit regularization. Furthermore, when clients have non-IID distributions, class-balanced objectives can further improve FL performances. In [218] a layer-wise federated learning algorithm is designed to account for permutation invariance of the neurons and permits global model size adaptation. Ma et al. [153] propose a layer-wise aggregation policy to enable pFL among heterogeneous clients. Gao et al. [67] introduce lightweight modifications in the training phase to decouple the global model from clients' local models using the local drift, improving the robustness and speed of model convergence. Liang et al. [135]

present a novel federated semi-supervised method, to address the uneven reliability of non-IID local clients. Instead of aggregating directly local clients, they propose the concept of updating the global model via aggregating multiple sub-consensus models. Inside a sub-consensus model, they propose a novel distance-reweighted model aggregation (DMA) module, which dynamically adjusts the weights of each sampled local client to the sub-consensus model.

In [272], the authors present a hierarchical Bayesian modeling and variational inference algorithm that offers a closed-form estimation of a confidence value. This confidence value takes into account the uncertainty of clients' parameters and the local model deviations from the global model. During the aggregation stage, the confidence value is used to weigh clients' parameters and adjust the regularization effect of the global model. Similarly, [178] estimate the uncertainty according to the performance of each client and perform aggregation by selecting highly reliable clients.

4) REPRESENTATION'S ALIGNMENT

A widely used strategy to align the learned representations across the different clients is to introduce additional losses and constraints that work on the internal feature representations trying to regularize them across multiple clients. In [261], an adversary module is proposed to reduce the divergence in feature representation among different clients and two consensus losses are proposed to reduce the inconsistency in optimization objectives from two perspectives. In the pursuit of a structure-feature alignment across the collaborative model, [254] designs a feature-oriented model structure adaptation method to ensure explicit feature allocation in different neural network structures. Then, they propose a feature-paired averaging scheme to guarantee aligned feature distribution and maintain no feature fusion conflicts. Zhou et al. [271] demonstrate how data heterogeneity leads to a vicious cycle between classifier divergence and feature inconsistency across client models. To break it, they leverage feature anchors to align features and classifiers across clients. All client models are updated in a uniform feature space with corresponding classifiers.

5) OPTIMIZING COMMUNICATION COST

Another set of works focuses instead on the communication aspects of FL and aims at reducing the network communication resources usage and at the same time enforcing privacy by avoiding transmitting sensitive data. A new FL framework exploiting knowledge distillation that can preserve privacy by design, while also consuming substantially less network communication resources w.r.t. baseline methods is proposed in [71]. An extensive empirical study is performed in [170] by comparing 12 variations of federated optimization methods on three commonly-used FL benchmark dataset, with the objective of understanding how the initialization impact the behavior of federated optimization methods. They found that a pre-trained solution can close the gap between training on IID and non-IID data and that initializing FL with a

pre-trained model can increase final model accuracy and reduce the number of rounds required to achieve a target accuracy. Tan et al. [203] present a novel FL method that improves communication efficiency in the heterogeneous setting by transmitting prototype representations instead of model weights, further improving privacy. They propose also a novel prototype-based aggregation. Isik et al. [94] propose a FL framework where instead of model weights, clients train a stochastic binary mask used to sparsify the dense network with random weights, yielding a communication efficient solution, faster convergence and higher accuracy. Similarly, in [137] sparsity-aware training on clients is used to reduce both communication and computational costs. Zhang and Hanzo [260] propose a FL-aided multi-UAV system for classification optimizing the communication cost.

An image classification dataset designed for FL is presented in [95], which is close to a real-world scenario where each client has a unique dataset under domain shift. They propose two algorithms dealing with two topologies (cycle and star) which result to be more communication efficient. Instead of sending model weights to the server for aggregation, [235] propose generating synthetic data on each client to mimic the loss landscape of the original data through distribution matching. This approach decreases the number of communication rounds and enhances model quality by transmitting more informative and smaller data.

6) FASTER CONVERGENCE
Another key aspect is the time required for the convergence of the model. Hsu et al. [88] propose a framework to speed up convergence for imbalanced clients. The idea is to conceptually split large clients into multiple smaller ones, and repeat these small clients multiple times such that all virtual clients are of similar sizes. In addition, to avoid underutilizing training examples from large clients, the probability that any client is selected for a round is proportional to its data size. Tang et al. [206] formulate the goal of accelerating the convergence of FL as optimization problems that maximize the posterior expectation of loss decrease. They utilize Gaussian Processes to solve the optimization problem and obtain an effective client selection strategy for heterogeneous FL. Sun et al. [201] design a synchronization scheduler that can reduce the wasted waiting time of the server and improves the time efficiency in realistic settings. Moreover, they mitigate the accuracy drop by applying the semi-asynchronous protocol and enable extremely lagging devices to contribute to the global model training.

7) ENSURING PRIVACY AND SECURITY
Local adaptation techniques (fine-tuning, freeze base, multi-task learning, knowledge distillation) that an individual participant can use to mitigate the damage from privacy and robustness mechanisms are investigated in [256]. An hybridization of meta-heuristic methods with FL is proposed in [177], and analyzed in terms of the accuracy of the general model as well as for security against poisoning

attacks. Shi et al. [195] observe that the loss landscapes are sharper in the presence of differential privacy. Therefore, they propose to use Sharpness Aware Minimization in FL like in [24], in combination with differential privacy to generate local flat models with better stability and weight perturbation robustness. Other works focusing on attacks and security are presented in [122], [125], [126], [134], [202], [216], and [268].

8) CATASTROPHIC FORGETTING
The phenomenon of catastrophic forgetting (CF), typical of continual learning, is manifested also in federated learning from the presence of diverse data and incomplete participation, ultimately undermining its performance [148]. In particular, [118] formalizes the concept of "local client forgetting", showing that when a client performs local updates, it risks to overly optimize his local objective and to forget knowledge about the other subsets of data. They address this problem at client side, by re-weighting the softmax logits before computing the loss.

Another approach GradMA [148], addresses CF both at the client and server side, by correcting the update directions of the server and clients simultaneously. They resort to quadratic programming and memorize updates to correct the update directions. They propose also a memory reduction strategy for a practical FL scenario with a large number of clients.

9) NON-STANDARD SETTINGS
Other works focus on non-standard settings where data is unlabeled or made available in an incremental way. Huang et al. [90] make use of unlabeled public data and adopt self-supervised learning, to enable heterogeneous models communication and learn a generalizable representation. To alleviate catastrophic forgetting in FL, inter and intra-domain knowledge distillation is used.

Li et al. [123] focus on federated semi-supervised learning, assuming that only a subset of clients is fully supervised. They propose a pseudo-labeling strategy to handle the catastrophic forgetting problem and a class-balanced adaptive threshold selection to refine them.

Among the various approaches considering unlabeled data, the federated active learning framework has emerged as a promising solution. In [4], a simple active learning-based FL framework is presented to utilize unlabeled samples at clients for training local models in two real-world applications: natural disaster and waste classification. Whilst, [108] introduced a novel query sample strategy that utilizes local-only and global models to ensure both sides of inter-class diversity.

In [50] the authors address the problem of federated class incremental learning, *i.e.*, a global class-incremental model is trained in the FL setting. They identify the catastrophic forgetting in this scenario to be distinguished in local and global. A class-aware gradient compensation loss and a class-semantic relation distillation loss are used to address local forgetting, while a class-semantic relation distillation on the local clients is used to compensate for global forgetting.

TABLE 2. Comparison of object detection papers. The considered datasets are: Street-5/20 [147], KITTI [68], SODA10M [75], NuScenes [23], BDD100K [253], Pascal VOC2007 [58], INRIA Person [44].

Paper	Model	Dataset	# Clients	# Rounds
[147]	YOLOv3, Faster R-CNN	Street-5, Street-20	5, 20	∼[30, 189], ∼[83, 448]
[143]	YOLOv3	Private	-	-
[96]	YOLOv3	KITTI	4	15
[200]	RetinaNet	BDD100K, SODA10M, NuScenes	4	3
[40]	Faster R-CNN	Pascal VOC2007, INRIA Person	100	200

TABLE 3. Comparison of semantic segmentation papers. The considered datasets are: Pascal-VOC2012 [59], Citysacpes [43], IDDA [5], CrossCity [39], Mapillary [168], ADE20k [270], CamVID [22].

Paper	Model	Dataset	# Clients	# Rounds
[161]	DeepLabV3+-MobilenetV2	Pascal VOC2012	100	400
[62]	BiSeNet V2	Cityscapes, IDDA	[146,144], [90,69]	1600
[24]	BiSeNet V2	Cityscapes, IDDA	[146,144], [90,69]	1600
[245]	DeepLabV3-MobilenetV2	CrossCity	4	-
[193]	DeepLabV3-MobilenetV2	Cityscapes, CrossCity, Mapillary	144, 476, 357	300, 1000, 100
[34]	DeepLabV3-MobileNetV2	Cityscapes	18	100
[51]	DeepLabV3-ResNet-101	Pascal-VOC 2012, ADE20k	10	
[159]	BiSeNetv2	Cityscapes, CamVID, Pascal-VOC2012, ADE20k	19, 11, 20, 150	

Moreover, a prototype gradient-based mechanism is implemented to protect communication.

B. OBJECT DETECTION

Object detection consists in localizing - typically with a bounding box - and labeling every single object of interest into an image. Luo et al. [147] introduce a real-world dataset collected from street cameras that reflects the characteristics of the federated settings, namely that it is non-IID and imbalanced. They propose two configurations: in the first, each camera is a client (Street-20); in the latter, cameras are clustered in nearby areas (Street-5). Finally, they test their benchmark using a modified version of FedAvg. Liu et al. [143] developed a platform (FedVision) that enables end-to-end collaborative training using personalized and locally stored datasets from different clients. They implement FedYOLOv3, which fits the YOLOv3 [181] object detection model into the FedAvg framework. The local parameters are sent in a compressed and encrypted manner to the server. A similar strategy is implemented in [96], considering an autonomous driving FL scenario, where they tested their framework on four clients using the KITTI dataset [68]. In FedOD [200], a multi-teacher distillation and a weighted bounding boxes fusion scheme are exploited to provide each client with both global and personalized models. In the FedOD setup, two pairs of clients take data respectively from SODA10M [75] and NuScenes [23] datasets, while the server data comes from BDD100K [253], which is the largest one in order to mimic statistical heterogeneity. Chow et al. [40] present a systematic framework for protecting FL from attacks. The authors propose a spatio-temporal signature analysis to mitigate network failures resulting from errors inherent in spatial clustering of gradients.

C. SEMANTIC SEGMENTATION

Semantic Segmentation involves assigning a label to every single pixel in an image. State-of-the-art models utilize encoder-decoder architectures based on CNNs or transformers [37], [130], [144], [233]. These models have typically a huge number of parameters but lightweight architectures with smaller model sizes and lower computational complexity [77], [85], [186] have also been proposed. As for the other

tasks distributed training solutions must be considered due to privacy and efficiency constraints. FedMargin [161] provides a distributed framework for both image classification and semantic segmentation. It estimates client deviations from the margin of class-conditional representations, and uses this information to drive the federated optimization by means of an attention mechanism. FedDrive [62] explores the application of federated learning to the problem of autonomous driving using data from multiple vehicles.

Federated Multi-target Domain Adaptation (FMTDA) [245] addresses the challenge of dealing with a limited number of clients having unlabeled target local datasets with dissimilar distributions. The approach leverages a labeled source dataset that is accessible on the server side, while simultaneously handling the aforementioned challenge in a federated setting. The authors pigeonhole their setting *Multi-Target* to highlight how the distributions of the target datasets of the clients are affected by statistical heterogeneity. Experimental evaluation is performed on four clients, where each client has all the images belonging to one among the four cities of CrossCity [39] or to one of the four domains of BDD100K [253]. Similarly, the Federated source-Free Domain Adaptation (FFREEDA), introduced in [193], considers a more realistic scenario with a large number of unsupervised clients, while the server has access to a source labeled dataset only for pre-training. Instead, the task of federated incremental learning for semantic segmentation is considered in [51]. Differently from previous works which focused on domain heterogeneity, [159] addresses the challenging class heterogeneity problem for semantic segmentation. They propose a modified cross entropy loss and a pixel contrastive loss to mitigate inconsistencies and client drift during local adaptation.

D. FACE RECOGNITION

Face recognition (FR) is a biometric task that compares and analyzes patterns to uniquely identify the face of a person. A worldwide discussion on AI ethics has been sparked

by commercial applications based on face recognition and facial analysis techniques, leading to the release of governance guidelines and recommendations from various countries (e.g., EU's General Data Protection Regulation [214]). Therefore, in most cases, personal devices are not authorized to access large amounts of facial data, and data sharing is not permitted. Even providing only internal network representations, as in other vision problems addressing FL, might result in privacy leakage since high-fidelity face pictures can be generated using ad-hoc ML techniques like DeepInversion [250]. The dataset themselves are subject to privacy violations: as an example, MS-Celeb-1M [73], a large-scale dataset containing 100K individuals, was taken off the Internet due to privacy concerns. Examples of publicly available datasets for this task are Labeled faces in the wild (LFW) [89], IARPA Janus Benchmark-A (IJB-A) [109], IJB-B [229], IJB-C [154] and Facial Faces in the Wild (RFW) [224]. As in other vision tasks, most of the approaches rely on pre-trained networks such as CosFace [217] and MobileFaceNet [38], which constitute efficient backbones for FR. Summarizing, face recognition requires model training and distribution techniques able to ensure privacy protection in light of the rising social agreement on data privacy and FL is a very valuable tool to achieve this goal.

The first work to introduce federated learning into a FR task has been FedFace [14]. The framework consists of a server, 3 trainers, and 2 validators. The server maintains a global momentum, which is evenly applied to each step of a training round. In addition, the aggregated models are validated on several parties, each of which has a private validation dataset, in order to dynamically discover the nearly optimal weightings for aggregation. FedFace [3] addresses the scenario in which each of the participating clients has photos belonging to a single identity. The server only communicates the parameters of the feature extractor and the class embedding, rather than the classification matrix which contains sensitive information. FedGC [171] aims to guarantee that each client holds private class embeddings. They propose a softmax-based regularizer to correct the gradients of the embeddings by injecting a cross-client gradient term. Without data sharing, FedFR [276] seeks to learn a model for the unlabeled target domain by adapting from a source domain. The source dataset is used to pre-train one client while on the other two clients, a hierarchical clustering algorithm is used to generate pseudo labels for the target data. FedFR [141] employs a globally shared dataset to regularize the local model training. Only "hard" global samples, *i.e.*, those with cosine similarity greater than a threshold to any of the local data for training, are supplied in order to improve local personalization without increasing the computational load. In practice, a larger threshold corresponds to less global data. To identify a trade-off between personalizing and maintaining the local models' similarity to the global model, contrastive regularization and feature customization are exploited simultaneously. In PrivacyFace [157], privacy-agnostic class prototypes are generated

TABLE 4. Comparison of Face Recognition papers. The considered datasets are: MS-Celeb-1M [73], LFW [89], CPLFW/CALFW/SLLFW [47], CFP-FP [189], AgeDB [164], MegaFace [107], VGGFace2 [27], IJB-A [109], IJB-B [229], IJB-C [154], RFW [224], CASIAWebFace [249], BUPT-BalancedFace [223].

Paper	Model	Dataset	# Clients	# Rounds
[14]	ResNet-18	MS-Celeb-1M, Asian-DeepGlint, VGGFace2, LFW, CPLFW, CALFW, CFP-FP, AgeDB, MegaFace, IJB-B/C	3, 12	26
[3]	CosFace	CASIAWebFace, IJB-A/C, LFW	4, 8, 16, 64, 10000	200
[171]	ResNet-34	CASIAWebFace, VGGFace2, LFW, SLLFW, CPLFW, CALFW, CFP-FP, AgeDB, MegaFace, IJB-B/C	36	
[276]	ResNet-34	MS-Celeb-1M, RFW	2, 5	
[141]	CosFace	MS-Celeb-1M, IJB-C	40	30
[157]	ResNet-18	CASIAWebFace, BUPT-Balancedface, RFW, IJB-B/C	4	10
[49]	MobileFaceNet	MS-Celeb-1M, Private		20
[191]	ResNet-50, MobileFaceNet	MS-Celeb-1M, CASIAWebFace, LFW, CFP-FP, AgeDB	10, 50, 500, 2500	80

to prevent any specific individual in the cluster to be learned. Moreover, during the local optimization, a consensus-aware loss forces clients not to embed samples in inappropriate feature spaces (*i.e.*, private clusters). Ding et al. [49] use transfer learning to speed up federated training on devices. They present an architecture in which a private projector helps to secure shared gradients without involving additional memory consumption or computational costs. Shang et al. [191] test the use of various loss functions. In particular, using pair-based methods such as Multi-Similarity Loss can be more accurate and communication-efficient when there are few classes for each client. However, when there are many classes on each client, using classification-based loss functions such as the CosFace [217] loss can improve the global model faster and with less communication cost.

E. MEDICAL IMAGING

Privacy issues are even more critical in the medical imaging field, where the data could disclose very sensitive information about the patient's pathologies and medical history. In addition, although institutions and hospitals may gather the same kind of medical data, those data may have varying characteristics due to collection methods, standards and protocols, institutional policies and priorities, different patient population and different privacy regulations. The bottleneck of FL in dealing with multi-source decentralized medical imaging is the cross-client variance issue, which is exacerbated by the restricted amount of training data that is typically accessible for each client and the constrained number of clients.

1) DISEASE CLASSIFICATION

A FL framework based on FedAvg for detecting lung nodules in CT scans is proposed in [13]. In this work the task is modeled as a classification one. A FL method for medical imaging

classification, that uses multiple domains of siloed data, *i.e.*, data from different sites with different features and labels, is proposed in [8]. It exploits generative models for data synthesis, which allows the training of a deep learning model on the synthesized data from all domains. Roth et al. [185] investigate using FedAvg for a breast density classification task on real-world multi-institutional data. Tan et al. [205] use transfer learning to improve performances over the centralized approach. The method is evaluated on a dataset of mammograms for breast cancer diagnosis. Jimenez et al. [102] test their framework on breast cancer datasets as well. They exploit an ensemble of deep learning models and use a memory mechanism that allows the models to learn from the past experience of other models. Additionally, Li et al. investigate multi-site fMRI classification for the Autism Brain Imaging Data Exchange (ABIDE) dataset using federated learning [131].

2) SEGMENTATION FOR DISEASES DETECTION

Yan et al. [241] aim to reduce client-to-client differences by converting each client's raw image data into a common image space using image-to-image translation techniques while still respecting FL's privacy settings. The algorithm is validated on PROSTATEx benchmark [11] for prostate cancer detection. Fedcross [238] targets four public datasets (MSD [9], NCI-ISBI [140], PROMISE12 [20], and PROSTATEx [11]) gathered in various clinics with different MRIs. They train a global model sequentially across clients rather than performing aggregation. FedDG [142] learns a federated model from multiple remote source domains in order to directly generalize to unknown target domains. The data distribution information is shared across clients via an efficient continuous frequency space interpolation approach. Comprehensive studies on two common medical image segmentation tasks (retinal fundus image segmentation and prostate MRI segmentation) are performed. The narrowing of the generalization gap is also addressed in FedSM [236] which exploits personalized models that match different data distributions. A model selector then determines the closest model/data distribution for every test sample. The model has been evaluated on retinal disc/cup and prostate segmentation. Segmentation is addressed in FedDM [274] using bounding boxes as a form of weak supervision. Together with a collaborative annotation strategy, a hierarchical aggregation scheme is employed to mitigate local drift. The method is evaluated on polyp and prostate segmentation, using both magnetic resonance and endoscopic images. FedCE [99] provides an approach to compute client contributions from both the gradient and the data space and then recommends a fair global aggregation based on those estimates.

MRI brain tumor segmentation is a challenging task due to several factors, including differences in tumor size, shape, location, and appearance, as well as variations in imaging protocols and equipment across different institutions. These variations can result in differences in image quality, noise, and artifacts, making it difficult to accurately segment brain tumors using traditional machine-learning approaches. By training models on data from multiple sources, federated learning can help to account for variability in imaging protocols and equipment, as well as capture a broader range of tumor characteristics. Additionally, federated learning can help to reduce the risk of data breaches and protect sensitive patient information. Sheller et al. [192] are the first to test FedAvg on a multi-institutional dataset, the brain tumor segmentation dataset (BraTS) [158]. Li et al. use a similar method and added a privacy-preserving feature extraction technique to improve the model's robustness [129]. A distillation-based strategy aimed at reducing communication resources requirements is used in [71].

3) MRI RECONSTRUCTION

Guo et al. propose a multi-center MRI reconstruction method, FL-MRCM, using a federated learning approach that maintained the privacy of the local MRI data [72]. They help the adaptation of the local models through adversarial alignment by sending global latent representations to a domain discriminator. In FedMRI [64] the MRI reconstruction model is divided into two parts: a shared encoder and a site-specific decoder to mitigate domain shift. In FedGIMP [57], each site generates a synthetic image to enforce MRI reconstruction: the local generators follow a federated scheme to produce synthetic priors, which are combined with global priors as in adversarial models. FedPR [63] addresses the communication bottlenecks and catastrophic forgetting by learning prompts with minimal trainable parameters and only updating local prompts in the approximate null space of global prompts. In this way it efficiently tackles catastrophic forgetting ensuring that knowledge beyond the local distribution is not overwritten. Wang et al. [221] deal with the problem of misaligned unpaired neuroimaging data. They used significantly distorted photos to train a federated-based framework on brain image synthesis without breaking the hospital's privacy regulations.

4) COVID-19 DETECTION

Privacy protection is also required for contributing to breakthrough medical discoveries that necessitate data sharing across borders. The identification and segmentation of COVID-19 lesions from medical images have been targeted using deep learning with the aim to provide a useful tool for doctors. Yan et al. [240] investigate the use of FL for COVID-19 detection from chest X-ray images. They conducted experiments using different backbones and demonstrated the effectiveness of federated learning for training using distributed data without compromising privacy. Dou et al. [53] present a study that shows the effectiveness of FL for detecting COVID-19 lung abnormalities in CT scans. They trained their framework with private data from 3 different medical institutions in Hong Kong and validate their model on 4 external datasets, which consisted of CT

TABLE 5. Comparison of Medical imaging papers. Prostate segmentation: PROSTATEx [11], MSD [9], NCI-ISBI [20], PROMISE12 [140], I2CVB [119]; Retinal Segmentation: DRISHTI-GS1 [197], RIM-ONE [66], REFUGE [172], RIGA [6]; Polyp Segmentation: CVC-DB (ClinicDB, ColonDB) [19], Kvasir-seg [97]; Breast cancer classification: INBreast [163], DMIST [176], DDSM [82], CheXpert [93], ChestX-ray8 [225], Camelyon 16 and 17 [139]; MRI Segmentation: BraTS [158], IXI [1], fastMRI [110], FeTS [173], HPKS [100]; Covid detection: COVIDx [222], COVID-19-CT-Seg [152], CC-19 [113], LIDC [12], COVID-CT-Dataset [267], COVID-QU-EX [41].

Paper	Model	Dataset	# Clients	# Rounds
[241]	8-layer CNN	PROSTATEx, Private	2	100
[238]	3DUNet	PROSTATEx, MSD, NCI-ISBI, PROMISE12	4	25, 50, 100, 200, 400
[142]	2D Unet-like	DRISHTI-GS1, RIM-ONE, REFUGE, NCI-ISBI, PROMISE12, I2CVB	4, 6	100
[236]	UNet	DRISHTIGS1, RIM-ONE, REFUGE, NCBI-ISBI, PROMISE12, PROSTATEx	6	150
[274]	UNet	NCI-ISBI, PROMISE12, I2CVB, CVC-DB, Kvasir-seg	6	200
[99]	UNet	DRISHTIGS1, RIM-ONE, REFUGE, RIGA, NCI-ISBI, PROMISE12, I2CVB	6	200
[192]	UNet	BraTS	$\sim$[4, 32]	
[129]	ResNet101	BraTS	13	300
[71]	PSPNet-ResNet101	CheXpert, ChestX-ray8, BraTS	2, 3	
[131]	2-layer NN	ABIDE	4	50
[72]	UNet	IXI, BraTS, fastMRI, HPKS	4	50
[64]	UNet	BraTS, fastMRI, Private	4	50
[57]	8 FC layers	IXI, BraTS, fastMRI, Private	3	100
[221]	CycleGAN	IXI, BraTS	4	3
[63]	Swin, ConvNet	fastMRI, FeTS, IXI, Private	15	80
[240]	ResNeXt, ResNet-18, MobileNet-v2	COVIDx	5	100
[53]	RetinaNet	COVID-19-CT-Seg, Private	3	
[113]	SegCaps	CC-19	3	75
[242]	AH-Net	LIDC, Private	2	500, 1000
[265]	GhostNet, ResNet-50/101	COVID-CT-Dataset, COVID-QU-Ex, COVIDx	3	30
[13]	VNet	LIDC	3	$\sim$[3,5]
[8]	UNet	Camelyon-16/17	5	
[185]	DenseNet-121	DMIST, Private	7	300
[205]	MobileNet, DenseNet-121, ResNet-50, Xception	DDSM	4	60
[102]	ResNet-22	INBreast, Private	3	50

images from different countries. Kumar et al. [113] propose a blockchain-federated learning approach for COVID-19 detection from CT scans to facilitate privacy-preserving learning. The global model, like blockchains, is maintained as a distributed ledger with expanding lists of data records that are privacy-ensured by cryptographic hashes. Yang et al. [242], propose a federated semi-supervised learning method for segmenting COVID-19 regions in chest CT images using data from China, Italy, and Japan. They used a combination of labelled and unlabeled data from the participating sites to train the model. Finally, Zhang et al. [265]

suggest a dynamic-fusion-based federated learning approach for COVID-19 detection which exploits both X-ray and CT scan images. The study presented a method for dynamically fusing the model updates from different clients.

VII. BENCHMARKS

Comparing performances of different federated learning schemes is a challenging task since it is a relatively new area of research and there is still some lack of standardization in its application. In particular, although several efforts have been made to introduce standard benchmarks, the community has yet to agree on a common setting for FL.

The lack of a consistent data distribution across the clients and of a common choice of parameters (e.g., number of clients, local epochs, rounds, etc.) makes often the obtained performances not directly comparable. For this reason, Tables 1,2,3, 4 and 5 include also the various considered settings for the various approaches tackling the different tasks.

Without a standardized setting, it is hard to determine whether a particular algorithm or model is really better than another, or if the difference in performance is simply due to differences in the experimental setup. Nevertheless, in this section, we introduce the main efforts in outlining benchmarks and then we present some results, trying to organize them into groups of comparable outcomes.

LEAF [26] offers a benchmark for federated learning that focuses on edge devices with limited computational resources. It includes benchmark models and several datasets, including a federated split for EMNIST [42], the so-called FEMNIST. FedML [78] proposes a benchmark suited for federated learning that includes both vision and non-vision datasets. The benchmark comprehends baseline models (e.g., FedAvg [155] and FedProx [128]) and federated splits for CIFAR-10 [112] and CIFAR-100. FedScale [114] presents a benchmark for federated learning that focuses on scaling up to large numbers of devices. It includes some vision datasets, among which OpenImage [111] and Google Landmark [228], as well as baseline models and evaluation metrics. These benchmarks [26], [78], [114] present experiments with various settings, including different numbers of devices, data distribution, and organizations of the communication rounds.

For the two tasks of image classification and semantic segmentation we present a comparison of recent papers on the most common dataset respectively in Tables 6 and 7, while for the other tasks the great variability of considered datasets and settings make the comparison less significant.

Starting from image classification we focus on the CIFAR-10/100 dataset, which is the most common choice for works tackling this task. Following the previous discussion, we divided Table 6 into blocks where each block corresponds to a set of experiments in the same conditions: results are comparable inside the same block but otherwise refer to different settings. As a general result, it is possible to see how state-of-the-art techniques provide a relevant improvement with respect to baseline approaches like FedAvg or FedProx.

Results for semantic segmentation are in Table 7. Here there is a much more limited number of comparable approaches and two most common choices for the dataset: Cityscapes and CamVID. Again, ad-hoc approaches for the task achieve gains of $5-10\%$ with respect to the baseline FedAvg scheme.

VIII. NEW TRENDS

Traditional machine learning algorithms are commonly trained on specific datasets to address particular tasks. However, this approach encounters limitations when confronted with novel tasks or domains lacking sufficient labeled data. Meta-learning offers a solution to this challenge by learning from a diverse set of tasks, facilitating adaptation to new tasks efficiently. Additionally, it could contribute to FL by assisting in selecting relevant tasks or clients in each round, which enhances overall efficiency. Consequently, it would allow for personalized FL applications, making it easier to deploy personalized models. Although considerable research has been conducted on this subject [30], [101], there is a need for future studies to delve into meta-learning techniques and facilitate the emergence of personalized federated learning [169], [244].

In real-world computer vision settings, especially when the clients are users' mobile devices it is also not too realistic to assume the availability of labeled data at client side. Thus, federated unsupervised domain adaptation approaches are another intriguing research direction [193], [245].

Another salient aspect to consider is that computer vision primarily focused on single-modality data, i.e, RGB images, and single objective tasks. However, real-world FL applications often involve multiple data types, including images, textual descriptions, 3D or depth data, and other data coming from different sensors. Learning a joint representation from multiple modalities, typically leads to deeper and more meaningful understandings of the provided information [16], [184]. Nevertheless, multimodal data introduces additional challenges such as modality incongruities (e.g., different noise and distribution shifts) and missing modalities [65] that are even more challenging in FL settings where they can appear in a client-dependent way. Recent works extend FL concepts to tackle these problems in vision-language tasks [36], [65], [234], [255]. To this end, we expect more research on multimodal FL, by allowing the aggregation of knowledge from different data modalities in order to train a more comprehensive and robust model.

Additionally, there is a rising demand for more extensive research regarding the use of foundation models in FL. These models can be employed on the server-side to substantially enhance the performance and convergence speed of computer vision tasks. The pre-training of the models from scratch is a costly operation, both in terms of computational resources and training data. The utilization of foundation models can help overcome this issue. Therefore, researchers have started to explore novel approaches, taking inspiration from

TABLE 6. Performance comparison for recent papers on image classification using the CIFAR 10/100 datasets. C_N^x stands for CIFAR-N ($N=10$ or $N=100$) non-i.i.d. partitioned using a Dirichlet distribution parametrized by $\alpha=x$. In the parameters c is the total number of clients while the percentage of clients per round is behind the parenthesis and r is the total number of rounds. Results taken from [24], [67], [120], [118], and [148].

Method	Model	Params.	Accuracy	
			$C_{10}^{0.1}$	$C_{100}^{0.1}$
FedAvg [155]	ResNet-8	10 c (100%), 80 r	64.9	38.8
FedProx [128]	ResNet-8	10 c (100%), 80 r	66.1	42.0
SCAFFOLD [105]	ResNet-8	10 c (100%), 80 r	66.6	43.8
FedDyn [2]	ResNet-8	10 c (100%), 80 r	63.8	36.4
FedPVR [120]	ResNet-8	10 c (100%), 80 r	69.3	43.5
FedAvg [155]	VGG-11	10 c (100%), 80 r	69.3	34.3
FedProx [128]	VGG-11	10 c (100%), 80 r	72.1	35.0
SCAFFOLD [105]	VGG-11	10 c (100%), 80 r	74.1	43.4
FedDyn [2]	VGG-11	10 c (100%), 80 r	77.4	43.8
FedPVR [120]	VGG-11	10 c (100%), 80 r	78.2	43.5
			$C_{10}^{0.3}$	$C_{100}^{0.3}$
FedAvg [155]	ResNet-18	100 c (15%), 1000 r	79.8	41.8
FedProx [128]	ResNet-18	100 c (15%), 1000 r	79.8	41.8
SCAFFOLD [105]	ResNet-18	100 c (15%), 1000 r	82.5	50.3
FedDyn [2]	ResNet-18	100 c (15%), 1000 r	82.3	51.1
FedDC [67]	ResNet-18	100 c (15%), 1000 r	84.6	53.9
FedAvg [155]	ResNet-18	100 c (100%), 1000 r	79.1	40.1
FedProx [128]	ResNet-18	100 c (100%), 1000 r	78.9	40.9
SCAFFOLD [105]	ResNet-18	100 c (100%), 1000 r	83.0	50.4
FedDyn [2]	ResNet-18	100 c (100%), 1000 r	84.1	51.9
FedDC [67]	ResNet-18	100 c (100%), 1000 r	84.3	54.9
			C_{10}^{0}	C_{100}^{0}
FedAvg [155]	ResNet-18	100 c (5%), 10k/20k r	65.0	30.3
FedSAM [24]	ResNet-18	100 c (5%), 10k/20k r	70.2	31.0
FedASAM [24]	ResNet-18	100 c (5%), 10k/20k r	73.7	36.0
FedAvgM [87]	ResNet-18	100 c (5%), 10k/20k r	10.0	1.0
FedProx [128]	ResNet-18	100 c (5%), 10k/20k r	62.7	31.2
SCAFFOLD [105]	ResNet-18	100 c (5%), 10k/20k r	32.3	1.0
FedDyn [2]	ResNet-18	100 c (5%), 10k/20k r	67.7	1.0
AdaBest [212]	ResNet-18	100 c (5%), 10k/20k r	66.8	29.9
FedAvgM [87] + ASAM	ResNet-18	100 c (5%), 10k/20k r	77.3	1.0
FedProx [128] + ASAM	ResNet-18	100 c (5%), 10k/20k r	73.7	36.1
SCAFFOLD [105] + ASAM	ResNet-18	100 c (5%), 10k/20k r	77.8	43.7
FedDyn [2] + ASAM	ResNet-18	100 c (5%), 10k/20k r	77.4	22.2
AdaBest [212] + ASAM	ResNet-18	100 c (5%), 10k/20k r	77.5	39.8
FedAvg [155]	ResNet-18	100 c (20%), 10k/20k r	68.5	38.6
FedSAM [24]	ResNet-18	100 c (20%), 10k/20k r	72.9	28.6
FedASAM [24]	ResNet-18	100 c (20%), 10k/20k r	76.1	40.8
FedAvgM [87]	ResNet-18	100 c (20%), 10k/20k r	10.0	40.6
FedProx [128]	ResNet-18	100 c (20%), 10k/20k r	68.4	38.6
SCAFFOLD [105]	ResNet-18	100 c (20%), 10k/20k r	15.6	1.0
FedDyn [2]	ResNet-18	100 c (20%), 10k/20k r	73.8	1.4
AdaBest [212]	ResNet-18	100 c (20%), 10k/20k r	72.3	39.1
FedAvgM [87] + ASAM	ResNet-18	100 c (20%), 10k/20k r	84.9	39.6
FedProx [128] + ASAM	ResNet-18	100 c (20%), 10k/20k r	75.8	40.9
SCAFFOLD [105] + ASAM	ResNet-18	100 c (20%), 10k/20k r	77.9	42.6
FedDyn [2] + ASAM	ResNet-18	100 c (20%), 10k/20k r	81.0	23.5
AdaBest [212] + ASAM	ResNet-18	100 c (20%), 10k/20k r	78.4	45.0
			$C_{10}^{0.1}$	$C_{100}^{0.1}$
FedAvg [155]	ResNet-18	100 c (10/2%), 4000 r	80.0	49.4
FedProx [128]	ResNet-18	100 c (10/2%), 4000 r	77.6	48.7
SCAFFOLD [105]	ResNet-18	100 c (10/2%), 4000 r	79.4	53.2
FedNova [219]	ResNet-18	100 c (10/2%), 4000 r	78.0	50.0
FedAvg [155] + WSM [118]	ResNet-18	100 c (10/2%), 4000 r	86.4	64.0
FedProx [128] + WSM [118]	ResNet-18	100 c (10/2%), 4000 r	81.8	56.2
SCAFFOLD [105] + WSM [118]	ResNet-18	100 c (10/2%), 4000 r	81.2	57.2
FedNova [219] + WSM [118]	ResNet-18	100 c (10/2%), 4000 r	78.0	50.0

TABLE 6. *(Continued.)* Performance comparison for recent papers on image classification using the CIFAR 10/100 datasets. C_N^x stands for CIFAR-N ($N = 10$ or $N = 100$) non-i.i.d. partitioned using a Dirichlet distribution parametrized by $\alpha = x$. In the parameters c is the total number of clients while the percentage of clients per round is behind the parenthesis and r is the total number of rounds. Results taken from [24], [67], [118], [120], and [148].

Method	Model	Params.	Accuracy		
			C_{10}^1	$C_{10}^{0.1}$	$C_{10}^{0.01}$
FedAvg [155]	LeNet-5	100 c (10%), 1000 r	69.5	47.9	21.0
FedProx [128]	LeNet-5	100 c (10%), 1000 r	71.6	48.6	20.4
MOON [124]	LeNet-5	100 c (10%), 1000 r	71.1	48.8	19.4
SCAFFOLD [105]	LeNet-5	100 c (10%), 1000 r	53.3	35.9	15.6
FedAvgM [87]	LeNet-5	100 c (10%), 1000 r	71.0	51.9	21.0
FedDyn [2]	LeNet-5	100 c (10%), 1000 r	74.9	41.9	17.9
MimeLite [104]	LeNet-5	100 c (10%), 1000 r	78.0	53.3	20.7
GradMA [148]	LeNet-5	100 c (10%), 1000 r	79.9	65.9	30.8
			$C_{100}^{0.1}$ (5/10/50%)		
FedAvg [155]	VGG-11	100 c (5/10/50%), 1000 r	56.0	61.2	64.8
FedProx [128]	VGG-11	100 c (5/10/50%), 1000 r	55.9	61.3	64.7
MOON [124]	VGG-11	100 c (5/10/50%), 1000 r	55.4	60.6	64.5
SCAFFOLD [105]	VGG-11	100 c (5/10/50%), 1000 r	32.2	34.7	45.7
FedAvgM [87]	VGG-11	100 c (5/10/50%), 1000 r	55.9	61.3	64.9
FedDyn [2]	VGG-11	100 c (5/10/50%), 1000 r	53.0	58.5	61.7
MimeLite [104]	VGG-11	100 c (5/10/50%), 1000 r	58.0	63.3	64.7
GradMA [148]	VGG-11	100 c (5/10/50%), 1000 r	59.5	63.5	65.7

TABLE 7. Performance comparison for semantic segmentation on the Cityscapes and CamVID datasets). c is the total number of clients while the number of clients per round is between parenthesis and r is the total number of rounds. Results taken from [24], [159].

Method	Model	Params.	mIoU	
			Cityscapes	CamVID
FedAvg [155]	BiSeNetv2	160/900 r	45.1	58.4
FedProx [128]	BiSeNetv2	160/900 r	44.9	58.3
FedDyn [2]	BiSeNetv2	160/900 r	45.2	59.4
MOON [124]	BiSeNetv2	160/900 r	45.8	58.9
FedSeg [159]	BiSeNetv2	160/900 r	50.2	63.5
			Cityscapes	
FedAvg [155]	BiSeNetv2	144 c (5),1500 r	38.7	
FedSAM [24]	BiSeNetv2	144 c (5),1500 r	41.2	
FedASAM [24]	BiSeNetv2	144 c (5),1500 r	42.3	
SiloBN [8]	BiSeNetv2	144 c (5),1500 r	46.0	
SiloBN [8] + SAM	BiSeNetv2	144 c (5),1500 r	49.1	
SiloBN [8] + ASAM	BiSeNetv2	144 c (5),1500 r	49.8	

traditional transfer learning methods. These techniques aim to facilitate the seamless integration of large-scale pre-trained models, such as BERT [208] or CLIP [146], into the existing systems.

IX. CONCLUSION

Federated learning research has grown significantly over the past years, and computer vision has been one of the driving fields. The large size of training data and the privacy issues in some application fields like face recognition or medical imaging make federated learning a valuable tool for vision applications. Consequently, many approaches have been developed and this survey highlights the most significant ones.

With respect to standard deep learning approaches for computer vision, FL research has not evolved in a straightforward comparison of performances on benchmark datasets. Instead, most works try to address various practical problems by also designing the considered setting. On one side, this diversity of research directions stimulated the development of a variety of new techniques and ideas. On the other hand, it makes difficult to perform comparisons across approaches and consequently the understanding of which are the best building blocks to be used in real applications. To this aim, in this paper, we presented an overview of the most relevant tasks, problems and methods, in order to gain a better view of the whole picture. Despite these challenges, there are ongoing efforts to standardize federated learning settings. As the field continues to grow, it is likely that a more standardized approach to federated learning will emerge, enabling researchers to better compare results and build upon previous work.

Another limitation regards the target application: most of the techniques focus on one or a few vision tasks, although the algorithms could technically address other tasks. Following this observation, we firstly presented the employed methodologies in a task-agnostic way in Section V and then introduced the various approaches organized according to the considered task in Section VI. We think that our categorization of the main methods could be beneficial for new research as different methods and strategies can be combined for better performances.

REFERENCES

[1] *IXI Dataset; Brain Development*. Accessed: May 23, 2023. [Online]. Available: http://brain-development.org/ixi-dataset/

[2] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, and V. Saligrama, "Federated learning based on dynamic regularization," 2021, *arXiv:2111.04263*.

[3] D. Aggarwal, J. Zhou, and A. K. Jain, "FedFace: Collaborative learning of face recognition model," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Aug. 2021, pp. 1–8.

[4] L. Ahmed, K. Ahmad, N. Said, B. Qolomany, J. Qadir, and A. Al-Fuqaha, "Active learning based federated learning for waste and natural disaster image classification," *IEEE Access*, vol. 8, pp. 208518–208531, 2020.

[5] E. Alberti, A. Tavera, C. Masone, and B. Caputo, "IDDA: A large-scale multi-domain dataset for autonomous driving," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5526–5533, Oct. 2020.

[6] A. Almazroa, S. Alodhayb, E. Osman, E. Ramadan, M. Hummadi, M. Dlaim, M. Alkatee, K. Raahemifar, and V. Lakshminarayanan, "Retinal fundus images for glaucoma analysis: The RIGA dataset," *Proc. SPIE*, vol. 10579, pp. 55–62, Jan. 2018.

[7] C. Anderson and R. Farrell, "Improving fractal pre-training," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 2412–2421.

[8] M. Andreux, J. O. D. Terrail, C. Beguier, and E. W. Tramel, "Siloed federated learning for multi-centric histopathology datasets," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*. Cham, Switzerland: Springer, 2020, pp. 129–139.

[9] M. Antonelli, A. Reinke, and S. Bakas, "The medical segmentation decathlon," *Nature Commun.*, vol. 13, p. 4128, Jul. 2022.

[10] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," 2019, *arXiv:1907.02893*.

[11] S. G. Armato, H. Huisman, K. Drukker, L. Hadjiiski, J. S. Kirby, N. Petrick, G. Redmond, M. L. Giger, K. Cha, and A. Mamonov, "PROSTATEx challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images," *J. Med. Imag.*, vol. 5, no. 4, 2018, Art. no. 044501.

[12] S. G. Armato et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Jan. 2011.

[13] P. Baheti, M. Sikka, K. Arya, and R. Rajesh, "Federated learning on distributed medical records for detection of lung nodules," in *Proc. 15th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2020, pp. 445–451.

[14] F. Bai, J. Wu, P. Shen, S. Li, and S. Zhou, "Federated face recognition," 2021, *arXiv:2105.02501*.

[15] P. Bándi et al., "From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 550–560, Feb. 2019.

[16] K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, "A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets," *Vis. Comput.*, vol. 10, pp. 1–32, Jan. 2021.

[17] S. Beery, G. Van Horn, and P. Perona, "Recognition in Terra incognita," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 456–473.

[18] C. I. Bercea, B. Wiestler, D. Rueckert, and S. Albarqouni, "FedDis: Disentangled federated learning for unsupervised brain pathology segmentation," 2021, *arXiv:2103.03705*.

[19] J. Bernal, J. Sánchez, and F. Vilarino, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognit.*, vol. 45, no. 9, pp. 3166–3182, 2012.

[20] N. Bloch, A. Madabhushi, H. Huisman, J. Freymann, J. Kirby, M. Grauer, A. Enquobahrie, C. Jaffe, L. Clarke, and K. Farahani, "NCI-ISBI 2013 challenge: Automated segmentation of prostate structures," *Cancer Imag. Arch.*, vol. 370, no. 6, p. 5, 2015.

[21] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, and C. Kiddon, "Towards federated learning at scale: System design," in *Proc. Mach. Learn. Syst.*, vol. 1, 2019, pp. 374–388.

[22] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 88–97, Jan. 2009.

[23] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "NuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11618–11628.

[24] D. Caldarola, B. Caputo, and M. Ciccone, "Improving generalization in federated learning by seeking flat minima," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 1–12.

[25] D. Caldarola, M. Mancini, F. Galasso, M. Ciccone, E. Rodolà, and B. Caputo, "Cluster-driven graph federated learning over multiple domains," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 2743–2752.

[26] S. Caldas, S. M. K. Duddu, P. Wu, T. Li, J. Konečný, H. B. McMahan, V. Smith, and A. Talwalkar, "LEAF: A benchmark for federated settings," 2018, *arXiv:1812.01097*.

[27] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 67–74.

[28] C. Chen, L. Lyu, H. Yu, and G. Chen, "Practical attribute reconstruction attack against federated learning," *IEEE Trans. Big Data*, early access, Mar. 15, 2022, doi: 10.1109/TBDATA.2022.3159236.

[29] D. Chen, J. Hu, V. J. Tan, X. Wei, and E. Wu, "Elastic aggregation for federated optimization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 12187–12197.

[30] F. Chen, M. Luo, Z. Dong, Z. Li, and X. He, "Federated meta-learning with fast convergence and efficient communication," 2018, *arXiv:1802.07876*.

[31] H. Chen, W. Johnny, and H. Vikalo, "The best of both worlds: Accurate global and personalized models through federated learning with data-free hyper-knowledge distillation," 2023, *arXiv:2301.08968*.

[32] H.-Y. Chen and W.-L. Chao, "FedBE: Making Bayesian model ensemble applicable to federated learning," 2020, *arXiv:2009.01974*.

[33] H.-Y. Chen and W.-L. Chao, "On bridging generic and personalized federated learning for image classification," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–12.

[34] H.-Y. Chen, C.-H. Tu, Z. Li, H. W. Shen, and W.-L. Chao, "On the importance and applicability of pre-training for federated learning," in *Proc. 11th Int. Conf. Learn. Represent.*, 2023, pp. 1–15.

[35] J. Chen, M. Jiang, Q. Dou, and Q. Chen, "Federated domain generalization for image recognition via cross-client style transfer," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 361–370.

[36] J. Chen and A. Zhang, "FedMSplit: Correlation-adaptive federated multi-task learning across multimodal split networks," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2022, pp. 87–96.

[37] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[38] S. Chen, Y. Liu, X. Gao, and Z. Han, "MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices," in *Proc. Chin. Conf. Biometric Recognit.* Cham, Switzerland: Springer, 2018, pp. 428–438.

[39] Y.-H. Chen, W.-Y. Chen, Y.-T. Chen, B.-C. Tsai, Y. F. Wang, and M. Sun, "No more discrimination: Cross city adaptation of road scene segmenters," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2011–2020.

[40] K.-H. Chow, L. Liu, W. Wei, F. Ilhan, and Y. Wu, "STDLens: Model hijacking-resilient federated learning for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16343–16351.

[41] M. E. Chowdhury, M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, and A. I. N. A. Emadi, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132665–132676, 2020.

[42] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: Extending MNIST to handwritten letters," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2921–2926.

[43] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.

[44] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jul. 2005, pp. 886–893.

[45] L. N. Darlow, E. J. Crowley, A. Antoniou, and A. J. Storkey, "CINIC-10 is not ImageNet or CIFAR-10," 2018, *arXiv:1810.03505*.

[46] A. B. de Luca, G. Zhang, X. Chen, and Y. Yu, "Mitigating data heterogeneity in federated learning with data augmentation," 2022, *arXiv:2206.09979*.

[47] W. Deng, J. Hu, N. Zhang, B. Chen, and J. Guo, "Fine-grained face verification: FGLFW database, baselines, and human-DCMN partnership," *Pattern Recognit.*, vol. 66, pp. 63–73, Dec. 2017.

[48] E. Diao, J. Ding, and V. Tarokh, "HeteroFL: Computation and communication efficient federated learning for heterogeneous clients," 2020, *arXiv:2010.01264*.

[49] Y. Ding, X. Wu, Z. Li, Z. Wu, S. Tan, Q. Xu, W. Pan, and Q. Yang, "An efficient industrial federated learning framework for AIoT: A face recognition application," 2022, *arXiv:2206.13398*.

[50] J. Dong, L. Wang, Z. Fang, G. Sun, S. Xu, X. Wang, and Q. Zhu, "Federated class-incremental learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10154–10163.

[51] J. Dong, D. Zhang, Y. Cong, W. Cong, H. Ding, and D. Dai, "Federated incremental semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3934–3943.

[52] X. Dong, S. Q. Zhang, A. Li, and H. T. Kung, "SphereFed: Hyperspherical federated learning," in *Computer Vision—ECCV 2022*. Cham, Switzerland: Springer, 2022, pp. 165–184.

[53] Q. Dou et al., "Federated deep learning for detecting COVID-19 lung abnormalities in CT: A privacy-preserving multinational validation study," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–11, Mar. 2021.

[54] Z. Du, J. Sun, A. Li, P.-Y. Chen, J. Zhang, H. Li, and Y. Chen, "Rethinking normalization methods in federated learning," in *Proc. 3rd Int. Workshop Distrib. Mach. Learn.*, Dec. 2022, pp. 16–22.

[55] J.-H. Duan, W. Li, D. Zou, R. Li, and S. Lu, "Federated learning with data-agnostic distribution fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 8074–8083.

[56] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends® Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[57] G. Elmas, S. U. Dar, Y. Korkmaz, E. Ceyani, B. Susam, M. Ozbey, S. Avestimehr, and T. Çukur, "Federated learning of generative image priors for MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 42, no. 7, pp. 1996–2009, Jul. 2023.

[58] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[59] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. (2012). *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results*. [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html

[60] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," in *Proc. NeurIPS Conf.*, vol. 33, Dec. 2020, pp. 3557–3568.

[61] X. Fang and M. Ye, "Robust federated learning with noisy and heterogeneous clients," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10062–10071.

[62] L. Fantauzzo, E. Fanì, D. Caldarola, A. Tavera, F. Cermelli, M. Ciccone, and B. Caputo, "FedDrive: Generalizing federated learning to semantic segmentation in autonomous driving," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 11504–11511.

[63] C.-M. Feng, B. Li, X. Xu, Y. Liu, H. Fu, and W. Zuo, "Learning federated visual prompt in null space for MRI reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 8064–8073.

[64] C.-M. Feng, Y. Yan, S. Wang, Y. Xu, L. Shao, and H. Fu, "Specificity-preserving federated learning for MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 42, no. 7, pp. 2010–2021, Jul. 2023.

[65] T. Feng, D. Bose, T. Zhang, R. Hebbar, A. Ramakrishna, R. Gupta, M. Zhang, S. Avestimehr, and S. Narayanan, "FedMultimodal: A benchmark for multimodal federated learning," 2023, *arXiv:2306.09486*.

[66] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in *Proc. 24th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2011, pp. 1–6.

[67] L. Gao, H. Fu, L. Li, Y. Chen, M. Xu, and C.-Z. Xu, "FedDC: Federated learning with non-IID data via local drift decoupling and correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10102–10111.

[68] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[69] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An efficient framework for clustered federated learning," in *Advances in Neural Information Processing Systems*, vol. 33, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Ed. Red Hook, NY, USA: Curran Associates, 2020, pp. 19586–19597.

[70] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2066–2073.

[71] X. Gong, A. Sharma, S. Karanam, Z. Wu, T. Chen, D. Doermann, and A. Innanje, "Ensemble attention distillation for privacy-preserving federated learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 15056–15066.

[72] P. Guo, P. Wang, J. Zhou, S. Jiang, and V. M. Patel, "Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2423–2432.

[73] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 87–102.

[74] J. Hamer, M. Mohri, and A. T. Suresh, "FedBoost: A communication-efficient algorithm for federated learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 3973–3983.

[75] J. Han, X. Liang, H. Xu, K. Chen, L. Hong, J. Mao, C. Ye, W. Zhang, Z. Li, X. Liang, and C. Xu, "SODA10M: A large-scale 2D self/semi-supervised object detection dataset for autonomous driving," 2021, *arXiv:2106.11118*.

[76] F. Hanzely and P. Richtárik, "Federated learning of a mixture of global and local models," 2020, *arXiv:2002.05516*.

[77] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1–10.

[78] C. He, S. Li, J. So, X. Zeng, M. Zhang, H. Wang, X. Wang, P. Vepakomma, A. Singh, H. Qiu, X. Zhu, J. Wang, L. Shen, P. Zhao, Y. Kang, Y. Liu, R. Raskar, Q. Yang, M. Annavaram, and S. Avestimehr, "FedML: A research library and benchmark for federated machine learning," 2020, *arXiv:2007.13518*.

[79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[80] Y. He, Y. Chen, X. Yang, H. Yu, Y.-H. Huang, and Y. Gu, "Learning critically: Selective self-distillation in federated learning on non-IID data," *IEEE Trans. Big Data*, early access, Jul. 11, 2022, doi: 10.1109/TBDATA.2022.3189703.

[81] Y. He, Y. Chen, X. Yang, Y. Zhang, and B. Zeng, "Class-wise adaptive self distillation for heterogeneous federated learning," in *Proc. 36th AAAI Conf. Artif. Intell.*, vol. 22, 2022, pp. 1–6.

[82] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, Jr., R. Moore, K. Chang, and S. Munishkumaran, "Current status of the digital database for screening mammogram," in *Digital Mammography: Nijmegen, 1998.* Springer, 1998, pp. 457–460.

[83] D. Hendrycks and T. G. Dietterich, "Benchmarking neural network robustness to common corruptions and surface variations," 2018, *arXiv:1807.01697*.

[84] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[85] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[86] K. Hsieh, A. Phanishayee, O. Mutlu, and P. Gibbons, "The non-IID data quagmire of decentralized machine learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4387–4398.

[87] T.-M. H. Hsu, H. Qi, and M. Brown, "Measuring the effects of non-identical data distribution for federated visual classification," 2019, *arXiv:1909.06335*.

[88] T.-M. H. Hsu, H. Qi, and M. Brown, "Federated visual classification with real-world data distribution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 76–92.

[89] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognit.*, 2008, pp. 1–13.

[90] W. Huang, M. Ye, and B. Du, "Learn from others and be yourself in heterogeneous federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10133–10143.

[91] W. Huang, M. Ye, Z. Shi, H. Li, and B. Du, "Rethinking federated learning with domain shift: A prototype view," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16312–16322.

[92] F. Ilhan, G. Su, and L. Liu, "ScaleFL: Resource-adaptive federated learning with heterogeneous clients," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 24532–24541.

[93] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 590–597.

[94] B. Isik, F. Pase, D. Gunduz, T. Weissman, and M. Zorzi, "Sparse random networks for communication-efficient federated learning," 2022, *arXiv:2209.15328*.

[95] S. Jain and K. R. Jerripothula, "Federated learning for commercial image sources," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 6523–6532.

[96] D. Jallepalli, N. C. Ravikumar, P. V. Badarinath, S. Uchil, and M. A. Suresh, "Federated learning for object detection in autonomous vehicles," in *Proc. IEEE 7th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Aug. 2021, pp. 107–114.

[97] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-SEG: A segmented polyp dataset," in *MultiMedia Modeling*. Cham, Switzerland: Springer, 2020, pp. 451–462.

[98] E. Jiang, Y. Jacky Zhang, and O. Koyejo, "Federated auto-weighted domain adaptation," 2023, *arXiv:2302.05049*.

[99] M. Jiang, H. R. Roth, W. Li, D. Yang, C. Zhao, V. Nath, D. Xu, Q. Dou, and Z. Xu, "Fair federated medical image segmentation via client contribution estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16302–16311.

[100] S. Jiang, C. G. Eberhart, M. Lim, H.-Y. Heo, Y. Zhang, L. Blair, Z. Wen, M. Holdhoff, D. Lin, P. Huang, H. Qin, A. Quinones-Hinojosa, J. D. Weingart, P. B. Barker, M. G. Pomper, J. Laterra, P. C. M. van Zijl, J. O. Blakeley, and J. Zhou, "Identifying recurrent malignant glioma after treatment using amide proton transfer-weighted MR imaging: A validation study with image-guided stereotactic biopsy," *Clin. Cancer Res.*, vol. 25, no. 2, pp. 552–561, Jan. 2019.

[101] Y. Jiang, J. Konečný, K. Rush, and S. Kannan, "Improving federated learning personalization via model agnostic meta learning," 2019, *arXiv:1909.12488*.

[102] A. Jiménez-Sánchez, M. Tardy, M. A. G. Ballester, D. Mateus, and G. Piella, "Memory-aware curriculum federated learning for breast cancer classification," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2022, Art. no. 107318.

[103] P. Kairouz et al., "Advances and open problems in federated learning," *Found. Trends® Mach. Learn.*, vol. 14, nos. 1–2, pp. 1–210, 2019.

[104] S. P. Karimireddy, M. Jaggi, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "Mime: Mimicking centralized stochastic algorithms in federated learning," 2020, *arXiv:2008.03606*.

[105] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5132–5143.

[106] H. Kataoka, K. Okayasu, A. Matsumoto, E. Yamagata, R. Yamada, N. Inoue, A. Nakamura, and Y. Satoh, "Pre-training without natural images," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–12.

[107] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The MegaFace benchmark: 1 million faces for recognition at scale," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4873–4882.

[108] S. Kim, S. Bae, H. Song, and S.-Y. Yun, "Re-thinking federated active learning based on inter-class diversity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3944–3953.

[109] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus benchmark A," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1931–1939.

[110] F. Knoll et al., "FastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiol., Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190007.

[111] I. Krasin, "OpenImages: A public dataset for large-scale multi-label and multi-class image classification," *Dataset*, vol. 2, no. 3, p. 18, 2017.

[112] A. Krizhevsky, "Learning multiple layers of features from tiny images," Toronto, ON, Canada, 2009. [Online]. Available: http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf

[113] R. Kumar, A. A. Khan, J. Kumar, N. A. Golilarz, S. Zhang, Y. Ting, C. Zheng, and W. Wang, "Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging," *IEEE Sensors J.*, vol. 21, no. 14, pp. 16301–16314, Jul. 2021.

[114] F. Lai, Y. Dai, S. Singapuram, J. Liu, X. Zhu, H. Madhyastha, and M. Chowdhury, "FedScale: Benchmarking model and system performance of federated learning at scale," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 11814–11827.

[115] Y. Le and X. Yang, "Tiny ImageNet visual recognition challenge," *CS 231N*, vol. 7, no. 7, p. 3, 2015.

[116] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Mar. 1998.

[117] G. Lee, M. Jeong, Y. Shin, S. Bae, and S.-Y. Yun, "Preservation of the global knowledge by not-true distillation in federated learning," 2021, *arXiv:2106.03097*.

[118] G. Legate, L. Caccia, and E. Belilovsky, "Re-weighted softmax cross-entropy to control forgetting in federated learning," 2023, *arXiv:2304.05260*.

[119] G. Lemaître, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau, "Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review," *Comput. Biol. Med.*, vol. 60, pp. 8–31, Jan. 2015.

[120] B. Li, M. N. Schmidt, T. S. Alstrøm, and S. U. Stich, "On the effectiveness of partial variance reduction in federated learning with heterogeneous data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3964–3973.

[121] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5543–5551.

[122] J. Li, A. S. Rakin, X. Chen, Z. He, D. Fan, and C. Chakrabarti, "ResSFL: A resistance transfer framework for defending model inversion attack in split federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10184–10192.

[123] M. Li, Q. Li, and Y. Wang, "Class balanced adaptive pseudo labeling for federated semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16292–16301.

[124] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10708–10717.

[125] S. Li, Y. Cheng, Y. Liu, W. Wang, and T. Chen, "Abnormal client behavior detection in federated learning," 2019, *arXiv:1910.09933*.

[126] S. Li, Y. Cheng, W. Wang, Y. Liu, and T. Chen, "Learning to detect malicious clients for robust federated learning," 2020, *arXiv:2002.00211*.

[127] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[128] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, vol. 2, 2020, pp. 429–450.

[129] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso, and A. Feng, "Privacy-preserving federated brain tumour segmentation," in *Machine Learning in Medical Imaging*. Cham, Switzerland: Springer, 2019.

[130] X. Li, H. Ding, W. Zhang, H. Yuan, J. Pang, G. Cheng, K. Chen, Z. Liu, and C. Change Loy, "Transformer-based visual segmentation: A survey," 2023, *arXiv:2304.09854*.

[131] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: Abide results," *Med. Image Anal.*, vol. 65, Mar. 2020, Art. no. 101765.

[132] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," 2021, *arXiv:2102.07623*.

[133] X.-C. Li, Y.-C. Xu, S. Song, B. Li, Y. Li, Y. Shao, and D.-C. Zhan, "Federated learning with position-aware neurons," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10072–10081.

[134] Z. Li, J. Zhang, L. Liu, and J. Liu, "Auditing privacy defenses in federated learning via generative gradient leakage," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10122–10132.

[135] X. Liang, Y. Lin, H. Fu, L. Zhu, and X. Li, "RSCFed: Random sampling consensus federated semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10144–10153.

[136] X. Liang, S. Shen, J. Liu, Z. Pan, E. Chen, and Y. Cheng, "Variance reduced local SGD with lower communication complexity," 2019, *arXiv:1912.12844*.

[137] D. Liao, X. Gao, Y. Zhao, and C. Xu, "Adaptive channel sparsity for federated learning under system heterogeneity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 20432–20441.

[138] T. Lin, L. Kong, S. U. Stich, and M. Jaggi, "Ensemble distillation for robust model fusion in federated learning," in *Proc. NIPS*, vol. 33, 2020, pp. 2351–2363.

[139] G. Litjens, P. Bandi, B. E. Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, Q. F. Manson, N. Stathonikos, A. Baidoshvili, P. van Diest, C. Wauters, M. van Dijk, and J. van der Laak, "1399 H&E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset," *GigaScience*, vol. 7, no. 6, pp. 1–20, Jun. 2018.

[140] G. Litjens et al., "Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge," *Med. Image Anal.*, vol. 18, no. 2, pp. 359–373, Feb. 2014.

[141] C.-T. Liu, C.-Y. Wang, S.-Y. Chien, and S.-H. Lai, "FedFR: Joint optimization federated framework for generic and personalized face recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, 2022, pp. 1656–1664.

[142] Q. Liu, C. Chen, J. Qin, Q. Dou, and P.-A. Heng, "FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1013–1023.

[143] Y. Liu, A. Huang, Y. Luo, H. Huang, Y. Liu, Y. Chen, L. Feng, T. Chen, H. Yu, and Q. Yang, "FedVision: An online visual object detection platform powered by federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 13172–13179.

[144] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[145] N. Lu, Z. Wang, X. Li, G. Niu, Q. Dou, and M. Sugiyama, "Federated learning from only unlabeled data with class-conditional-sharing clients," 2022, *arXiv:2204.03304*.

[146] W. Lu, X. Hu, J. Wang, and X. Xie, "FedCLIP: Fast generalization and personalization for CLIP in federated learning," 2023, *arXiv:2302.13485*.

[147] J. Luo, X. Wu, Y. Luo, A. Huang, Y. Huang, Y. Liu, and Q. Yang, "Real-world image datasets for federated learning," 2019, *arXiv:1910.11089*.

[148] K. Luo, X. Li, Y. Lan, and M. Gao, "GradMA: A gradient-memory-based accelerated federated learning with alleviated catastrophic forgetting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3708–3717.

[149] Y. Luopan, R. Han, Q. Zhang, C. Harold Liu, and G. Wang, "FedKNOW: Federated continual learning with signature task knowledge integration at edge," 2022, *arXiv:2212.01738*.

[150] C. Ma, J. Li, M. Ding, H. H. Yang, F. Shu, T. Q. S. Quek, and H. V. Poor, "On safeguarding privacy and security in the framework of federated learning," *IEEE Netw.*, vol. 34, no. 4, pp. 242–248, Jul. 2020.

[151] J. Ma, S. Naas, S. Sigg, and X. Lyu, "Privacy-preserving federated learning based on multi-key homomorphic encryption," *Int. J. Intell. Syst.*, vol. 37, no. 9, pp. 5880–5901, Sep. 2022.

[152] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, Q. Zhu, G. Dong, J. He, Z. He, T. Cao, Y. Zhu, Z. Nie, and X. Yang, "Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation," *Med. Phys.*, vol. 48, no. 3, pp. 1197–1210, Mar. 2021.

[153] X. Ma, J. Zhang, S. Guo, and W. Xu, "Layer-wised model aggregation for personalized federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10082–10091.

[154] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, and P. Grother, "IARPA Janus benchmark—C: Face dataset and protocol," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 158–165.

[155] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, in Proceedings of Machine Learning Research, vol. 54, A. Singh and J. Zhu, Eds. PMLR, Apr. 2017, pp. 1273–1282.

[156] M. Mendieta, T. Yang, P. Wang, M. Lee, Z. Ding, and C. Chen, "Local learning matters: Rethinking data heterogeneity in federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8397–8406.

[157] Q. Meng, F. Zhou, H. Ren, T. Feng, G. Liu, and Y. Lin, "Improving federated learning face recognition via privacy-agnostic clusters," 2022, *arXiv:2201.12467*.

[158] B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[159] J. Miao, Z. Yang, L. Fan, and Y. Yang, "FedSeg: Class-heterogeneous federated learning for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 8042–8052.

[160] U. Michieli and M. Ozay, "Are all users treated fairly in federated learning systems?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2021, pp. 2318–2322.

[161] U. Michieli, M. Toldo, and M. Ozay, "Federated learning via attentive margin of semantic feature representations," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 1517–1535, Jan. 2023.

[162] A. Mora, I. Tenison, P. Bellavista, and I. Rish, "Knowledge distillation for federated learning: A practical guide," 2022, *arXiv:2211.04742*.

[163] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, no. 2, pp. 236–248, 2012.

[164] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "AgeDB: The first manually collected, in-the-wild age database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1997–2005.

[165] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Gener. Comput. Syst.*, vol. 115, pp. 619–640, Feb. 2021.

[166] X. Mu, Y. Shen, K. Cheng, X. Geng, J. Fu, T. Zhang, and Z. Zhang, "FedProc: Prototypical contrastive federated learning on non-IID data," 2021, *arXiv:2109.12273*.

[167] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *Proc. NIPS Deep Learn. Unsupervised Feature Learn. Workshop*, 2011, pp. 1–20.

[168] G. Neuhold, T. Ollmann, S. R. Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5000–5009.

[169] B.-L. Nguyen, T. C. Cao, and B. Le, "Meta-learning and personalization layer in federated learning," in *Proc. Asian Conf. Intell. Inf. Database Syst.* Cham, Switzerland: Springer, 2022, pp. 209–221.

[170] J. Nguyen, J. Wang, K. Malik, M. Sanjabi, and M. Rabbat, "Where to begin? On the impact of pre-training and initialization in federated learning," 2022, *arXiv:2206.15387*.

[171] Y. Niu and W. Deng, "Federated learning for face recognition with gradient correction," in *Proc. AAAI Conf. Artif. Intelligence*, vol. 36, 2022, pp. 1999–2007.

[172] J. I. Orlando et al., "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101570.

[173] S. Pati, U. Baid, B. Edwards, M. J. Sheller, P. Foley, G. A. Reina, S. Thakur, C. Sako, M. Bilello, C. Davatzikos, J. Martin, P. Shah, B. Menze, and S. Bakas, "The federated tumor segmentation (FeTS) tool: An open-source solution to further solid tumor research," *Phys. Med. Biol.*, vol. 67, no. 20, Oct. 2022, Art. no. 204002.

[174] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1406–1415.

[175] X. Peng, Z. Huang, Y. Zhu, and K. Saenko, "Federated adversarial domain adaptation," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–13.

[176] E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. K. Baum, S. Acharyya, E. F. Conant, L. L. Fajardo, L. Bassett, and C. D'Orsi, "Diagnostic performance of digital versus film mammography for breast-cancer screening," *New England J. Med.*, vol. 353, no. 17, pp. 1773–1783, 2005.

[177] D. Połap and M. Woźniak, "Meta-heuristic as manager in federated learning approaches for image processing purposes," *Appl. Soft Comput.*, vol. 113, Dec. 2021, Art. no. 107872.

[178] Z. Qin, L. Yang, Q. Wang, Y. Han, and Q. Hu, "Reliable and interpretable personalized federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 20422–20431.

[179] Z. Qu, X. Li, X. Han, R. Duan, C. Shen, and L. Chen, "How to prevent the poor performance clients for personalized federated learning?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 12167–12176.

[180] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečný, S. Kumar, and H. B. McMahan, "Adaptive federated optimization," 2020, *arXiv:2003.00295*.

[181] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[182] A. Reisizadeh, F. Farnia, R. Pedarsani, and A. Jadbabaie, "Robust federated learning: The case of affine distribution shifts," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21554–21565.

[183] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On data banks and privacy homomorphisms," *Found. Secure Comput.*, vol. 4, no. 11, pp. 169–180, 1978.

[184] G. Rizzoli, F. Barbato, and P. Zanuttigh, "Multimodal semantic segmentation in autonomous driving: A review of current approaches and future perspectives," *Technologies*, vol. 10, no. 4, p. 90, Jul. 2022.

[185] H. R. Roth et al., "Federated learning for breast density classification: A real-world implementation," in *Proc. MICCAI Workshop Domain Adaptation Represent. Transf.* Cham, Switzerland: Springer, 2020, pp. 181–191.

[186] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
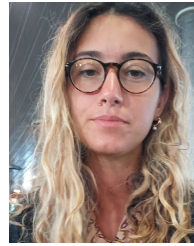
[187] F. Sattler, T. Korjakow, R. Rischke, and W. Samek, "FEDAUX: Leveraging unlabeled auxiliary data in federated learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 1, 2021, doi: 10.1109/TNNLS.2021.3129371.

[188] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3710–3722, Aug. 2021.

[189] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9.

[190] O. Shamir, N. Srebro, and T. Zhang, "Communication-efficient distributed optimization using an approximate Newton-type method," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1000–1008.

[191] E. Shang, Z. Yang, H. Liu, J. Du, and X. Wang, "FedFR: Evaluation and selection of loss functions for federated face recognition," in *Collaborative Computing: Networking, Applications and Worksharing*. Cham, Switzerland: Springer, 2023, pp. 95–114.

[192] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 92–104.

[193] D. Shenaj, E. Fanì, M. Toldo, D. Caldarola, A. Tavera, U. Michieli, M. Ciccone, P. Zanuttigh, and B. Caputo, "Learning across domains and devices: Style-driven source-free domain adaptation in clustered federated learning," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 444–454.

[194] D. Shenaj, M. Toldo, A. Rigon, and P. Zanuttigh, "Asynchronous federated continual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 1–12.

[195] Y. Shi, Y. Liu, K. Wei, L. Shen, X. Wang, and D. Tao, "Make landscape flatter in differentially private federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 24552–24562.

[196] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2015, pp. 909–910.

[197] J. Sivaswamy, S. Krishnadas, A. Chakravarty, G. Joshi, and A. S. Tabish, "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis," *JSM Biomed. Imag. Data Papers*, vol. 2, no. 1, p. 1004, Mar. 2015.

[198] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1631–1642.

[199] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: A multi-class classification competition," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2011, pp. 1453–1460.

[200] S. Su, B. Li, C. Zhang, M. Yang, and X. Xue, "Cross-domain federated object detection," 2022, *arXiv:2206.14996*.

[201] J. Sun, A. Li, L. Duan, S. Alam, X. Deng, X. Guo, H. Wang, M. Gorlatova, M. Zhang, H. Li, and Y. Chen, "FedSEA: A semi-asynchronous federated learning framework for extremely heterogeneous devices," in *Proc. 20th ACM Conf. Embedded Networked Sensor Syst.*, Nov. 2022, pp. 106–119.

[202] J. Sun, A. Li, B. Wang, H. Yang, H. Li, and Y. Chen, "Soteria: Provable defense against privacy leakage in federated learning from representation perspective," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9307–9315.

[203] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, and C. Zhang, "FedProto: Federated prototype learning across heterogeneous clients," in *Proc. AAAI Conf. Artif. Intell.*, vol. 1, 2022, p. 3.

[204] Y. Tan, G. Long, J. Ma, L. Liu, T. Zhou, and J. Jiang, "Federated learning from pre-trained models: A contrastive learning approach," 2022, *arXiv:2209.10083*.

[205] Y. N. Tan, V. P. Tinh, P. D. Lam, N. H. Nam, and T. A. Khoa, "A transfer learning approach to breast cancer classification in a federated learning framework," *IEEE Access*, vol. 11, pp. 27462–27476, 2023.

[206] M. Tang, X. Ning, Y. Wang, J. Sun, Y. Wang, H. Li, and Y. Chen, "FedCor: Correlation-based active client selection strategy for heterogeneous federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10092–10101.

[207] D. Teney, E. Abbasnejad, S. Lucey, and A. Van den Hengel, "Evading the simplicity bias: Training a diverse set of models discovers solutions with superior OOD generalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16740–16751.

[208] Y. Tian, Y. Wan, L. Lyu, D. Yao, H. Jin, and L. Sun, "FedBERT: When federated learning meets pre-training," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 4, pp. 1–26, Aug. 2022.

[209] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," *Mach. Vis. Appl.*, vol. 25, no. 3, pp. 633–647, Apr. 2014.

[210] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1–9, Aug. 2018.

[211] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie, "The iNaturalist species classification and detection dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8769–8778.

[212] F. Varno, M. Saghayi, L. R. Sevyeri, S. Gupta, S. Matwin, and M. Havaei, "AdaBest: Minimizing client drift in federated learning via adaptive bias estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 710–726.

[213] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5385–5394.

[214] P. Voigt and A. Von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*, 1st ed. Cham, Switzerland: Springer, 2017.

[215] H. Wang, Y. Li, W. Xu, R. Li, Y. Zhan, and Z. Zeng, "DaFKD: Domain-aware federated knowledge distillation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 20412–20421.

[216] H. Wang, K. Sreenivasan, S. Rajput, H. Vishwakarma, S. Agarwal, J.-Y. Sohn, K. Lee, and D. Papailiopoulos, "Attack of the tails: Yes, you really can backdoor federated learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 16070–16084.

[217] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5265–5274.

[218] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, and Y. Khazaeni, "Federated learning with matched averaging," 2020, *arXiv:2002.06440*.

[219] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 33, Dec. 2020, pp. 7611–7623.

[220] J. Wang, V. Tantia, N. Ballas, and M. Rabbat, "SlowMo: Improving communication-efficient distributed SGD with slow momentum," 2019, *arXiv:1910.00643*.

[221] J. Wang, G. Xie, Y. Huang, Y. Zheng, Y. Jin, and F. Zheng, "FedMed-ATL: Misaligned unpaired cross-modality NeuroImage synthesis via affine transform loss," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 1522–1531.

[222] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Nov. 2020.

[223] M. Wang and W. Deng, "Mitigating bias in face recognition using skewness-aware reinforcement learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9319–9328.

[224] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang, "Racial faces in the wild: Reducing racial bias by information maximization adaptation network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 692–702.

[225] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471.

[226] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2019, pp. 2512–2520.

[227] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3454–3469, 2020.

[228] T. Weyand, A. Araujo, B. Cao, and J. Sim, "Google landmarks dataset v2—A large-scale benchmark for instance-level recognition and retrieval," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2572–2581.

[229] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother, "IARPA Janus benchmark—B face dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 592–600.

[230] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.

[231] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3485–3492.

[232] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang, "Learning from massive noisy labeled data for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2691–2699.

[233] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 12077–12090.

[234] B. Xiong, X. Yang, F. Qi, and C. Xu, "A unified framework for multimodal federated learning," *Neurocomputing*, vol. 480, pp. 110–118, Apr. 2022.

[235] Y. Xiong, R. Wang, M. Cheng, F. Yu, and C.-J. Hsieh, "FedDM: Iterative distribution matching for communication-efficient federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16323–16332.

[236] A. Xu, W. Li, P. Guo, D. Yang, H. Roth, A. Hatamizadeh, C. Zhao, D. Xu, H. Huang, and Z. Xu, "Closing the generalization gap of cross-silo federated medical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 20834–20843.

[237] J. Xu, Z. Chen, T. Q. S. Quek, and K. F. Ernest Chong, "FedCorr: Multi-stage federated learning for label noise correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10184–10193.

[238] X. Xu, H. H. Deng, T. Chen, T. Kuang, J. C. Barber, D. Kim, J. Gateno, J. J. Xia, and P. Yan, "Federated cross learning for medical image segmentation," 2022, *arXiv:2204.02450*.

[239] Y.-Y. Xu, C.-S. Lin, and Y.-C.-F. Wang, "Bias-eliminating augmentation learning for debiased federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 20442–20452.

[240] B. Yan, J. Wang, J. Cheng, Y. Zhou, Y. Zhang, Y. Yang, Li Liu, H. Zhao, C. Wang, and B. Liu, "Experiments of federated learning for COVID-19 chest X-ray images," in *Proc. Int. Conf. Artif. Intell. Secur.* Cham, Switzerland: Springer, 2021, pp. 41–53.

[241] Z. Yan, J. Wicaksana, Z. Wang, X. Yang, and K.-T. Cheng, "Variation-aware federated learning with multi-source decentralized medical image data," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2615–2628, Jul. 2021.

[242] D. Yang, Z. Xu, W. Li, A. Myronenko, H. R. Roth, S. Harmon, S. Xu, B. Turkbey, E. Turkbey, X. Wang, W. Zhu, G. Carrafiello, F. Patella, M. Cariati, H. Obinata, H. Mori, K. Tamura, P. An, B. J. Wood, and D. Xu, "Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101992.

[243] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, "MedMNIST v2—A large-scale lightweight benchmark for 2D and 3D biomedical image classification," *Sci. Data*, vol. 10, no. 1, p. 41, Jan. 2023.

[244] L. Yang, J. Huang, W. Lin, and J. Cao, "Personalized federated learning on non-IID data via group-based meta-learning," *ACM Trans. Knowl. Discovery Data*, vol. 17, no. 4, pp. 1–20, Aug. 2023.

[245] C.-H. Yao, B. Gong, H. Qi, Y. Cui, Y. Zhu, and M.-H. Yang, "Federated multi-target domain adaptation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1081–1090.

[246] D. Yao, W. Pan, Y. Dai, Y. Wan, X. Ding, H. Jin, Z. Xu, and L. Sun, "Local–global knowledge distillation in heterogeneous federated learning with non-IID data," 2021, *arXiv:2107.00051*.

[247] X. Yao and L. Sun, "Continual local training for better initialization of federated models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 1736–1740.

[248] D. Ye, R. Yu, M. Pan, and Z. Han, "Federated learning in vehicular edge computing: A selective model aggregation approach," *IEEE Access*, vol. 8, pp. 23920–23935, 2020.

[249] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014, *arXiv:1411.7923*.

[250] H. Yin, P. Molchanov, J. M. Alvarez, Z. Li, A. Mallya, D. Hoiem, N. K. Jha, and J. Kautz, "Dreaming to distill: Data-free knowledge transfer via DeepInversion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8712–8721.

[251] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–11, Jul. 2021.

[252] J. Yoon, W. Jeong, G. Lee, E. Yang, and S. J. Hwang, "Federated continual learning with weighted inter-client transfer," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 12073–12086.

[253] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.

[254] F. Yu, W. Zhang, Z. Qin, Z. Xu, D. Wang, C. Liu, Z. Tian, and X. Chen, "Fed^2: Feature-aligned federated learning," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 2066–2074.

[255] Q. Yu, Y. Liu, Y. Wang, K. Xu, and J. Liu, "Multimodal federated learning via contrastive representation ensemble," 2023, *arXiv:2302.08888*.

[256] T. Yu, E. Bagdasaryan, and V. Shmatikov, "Salvaging federated learning by local adaptation," 2020, *arXiv:2002.04758*.

[257] M. Yurochkin, M. Agarwal, S. Ghosh, K. Greenewald, N. Hoang, and Y. Khazaeni, "Bayesian nonparametric federated learning of neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7252–7261.

[258] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowl.-Based Syst.*, vol. 216, Jan. 2021, Art. no. 106775.

[259] F. Zhang, K. Kuang, Z. You, T. Shen, J. Xiao, Y. Zhang, C. Wu, Y. Zhuang, and X. Li, "Federated unsupervised representation learning," 2020, *arXiv:2010.08982*.

[260] H. Zhang and L. Hanzo, "Federated learning assisted multi-UAV networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 14104–14109, Nov. 2020.

[261] L. Zhang, Y. Luo, Y. Bai, B. Du, and L.-Y. Duan, "Federated learning for non-IID data via unified feature learning and optimization objective alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4400–4408.

[262] L. Zhang, L. Shen, L. Ding, D. Tao, and L.-Y. Duan, "Fine-tuning global model via data-free knowledge distillation for non-IID federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10164–10173.

[263] L. Zhang, D. Wu, and X. Yuan, "FedZKT: Zero-shot knowledge transfer towards resource-constrained federated learning with heterogeneous on-device models," in *Proc. IEEE 42nd Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2022, pp. 928–938.

[264] R. Zhang, Q. Xu, J. Yao, Y. Zhang, Q. Tian, and Y. Wang, "Federated domain generalization with generalization adjustment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3954–3963.

[265] W. Zhang, T. Zhou, Q. Lu, X. Wang, C. Zhu, H. Sun, Z. Wang, S. K. Lo, and F.-Y. Wang, "Dynamic-fusion-based federated learning for COVID-19 detection," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15884–15891, Nov. 2021.

[266] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–12.

[267] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, and P. Xie, "COVID-CT-dataset: A CT scan dataset about COVID-19," 2020, *arXiv:2003.13865*.

[268] J. C. Zhao, A. R. Elkordy, A. Sharma, Y. H. Ezzeldin, S. Avestimehr, and S. Bagchi, "The resource problem of using linear layer leakage attack in federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3974–3983.

[269] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," 2018, *arXiv:1806.00582*.

[270] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ADE20K dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5122–5130.

[271] T. Zhou, J. Zhang, and D. H. K. Tsang, "FedFA: Federated learning with feature anchors to align features and classifiers for heterogeneous data," 2022, *arXiv:2211.09299*.

[272] J. Zhu, X. Ma, and M. B. Blaschko, "Confidence-aware personalized federated learning via variational expectation maximization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 24542–24551.

[273] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–13.

[274] M. Zhu, Z. Chen, and Y. Yuan, "FedDM: Federated weakly supervised segmentation via annotation calibration and gradient de-conflicting," *IEEE Trans. Med. Imag.*, vol. 42, no. 6, pp. 1632–1643, Jun. 2023.

[275] Z. Zhu, J. Hong, and J. Zhou, "Data-free knowledge distillation for heterogeneous federated learning," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 12878–12889.

[276] W. Zhuang, X. Gan, X. Zhang, Y. Wen, S. Zhang, and S. Yi, "Federated unsupervised domain adaptation for face recognition," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2022, pp. 1–6.

GIULIA RIZZOLI (Student Member, IEEE) received the M.Sc. degree (Hons.) in telecommunication engineering from the University of Padova, in 2021, where she is currently pursuing the Ph.D. degree. As part of the master's thesis, she conducted research with Sony, Stuttgart, Germany. Her research interests include multi-modal learning, domain adaptation, federated, and continual learning applied to computer vision tasks.

DONALD SHENAJ (Student Member, IEEE) received the B.Sc. degree (Hons.) in electronics engineering from the University of Bologna, in 2019, and the M.Sc. degree (Hons.) in telecommunications engineering from the University of Padova, in 2021, where he is currently pursuing the Ph.D. degree. He is also a Visiting Researcher with the Mila—Quebec AI Institute. His research interests include federated learning, continual learning, and domain adaptation.

PIETRO ZANUTTIGH (Member, IEEE) received the Ph.D. degree from the University of Padova, in 2007, where he is currently an Associate Professor with the Department of Information Engineering. He works in the computer vision and machine learning fields, with a special focus on domain adaptation and incremental learning in semantic segmentation, federated learning for computer vision applications, 3D acquisition with ToF sensors, sensor fusion, and hand gesture recognition.

• • •