Romance never dies

Wenke Huang, Mang Ye and Bo Du

You only live once. Do what you love.

Version I – February 2023

Show my great respect to my grandmother (Xiangfen Cai) and my parents.

Contents

1 Federated Learning

1.1 Theory

Some theoretical analysis about federated learning

Federated Optimization: Distributed Machine Learning (arXiv'16 [148])
Lower bounds and optimal algorithms for personalized FL (NeurIPS'20 [92])
Convergence rate of FedAvg on NonIID is $\mathcal{O}(\frac{1}{T})$ (arXiv'19 [178])
Linearly converge in NonIID with part clients $\mathcal{O}(\frac{1}{\sqrt{nKT}}+\frac{1}{T})$ (ICLR'21 [359])
Federated versions of adaptive optimizers (ICLR'21 [236])
Linear speedup for FedAvg in statistics and system heter (ICLR'21 [235])
Quantify and estimate influence over the model parameters (AAAI'21 [349])
Sharp lower bounds for homo and heter FedAvg (AISTAS'22 [75])
Federated Minimax Optimization: Improved Convergence (ICML'22 [249])

1.2 Efficiency

Communication and Computation efficiency are important in federated learning

Structured updates and sketched updates (arXiv'16 [149])
MOCHA (NeurIPS'17 [260]): Systems-aware via multi-task learning
Matcha (ICC'19 [305]): Speed up via matching decomposition sampling
AFL (arXiv'19 [76]): Select with probability conditioned rather uniformity
FedAsync has near-linear convergence to a global optimum (arXiv'19 [335])
LotteryFL (arXiv'20 [161]): Exploit lottery ticket hypothesis
Throughput-optimal topology design for cross-silo FL (NeurIPS'20 [206])
HeteroFL (ICLR'21 [54]): Local model with diff computation complexity
Attention-based and dynamic fraction for client selection (IJCAIW'21 [37])
Compress gradient with error-compensated compression (AAAI'21 [68])
FedPara (ICLR'22 [121]): Low-rank hadamard product to re-parameterize
FedAGM (arXiv'22 [139]): Global gradient guides client updates
FedNew (ICML'22 [58]): Not transmit Hessian from clients to server
FedDST (AAAI'22 [13]) Dynamic Sparse train sub-networks in FL

1.3 Label Skew

Focus on label distribution inconsistency

FedAvg (AISTATS'17 [208]): Average parameter to realize communication
FedProx (MLSys'20 [177]): Introduce proximal term (global parameter)
FedCurv (NeurIPS'19 [258]): Leverage hessian matrix (global parameter)
FedGKT (NeurIPS'20 [100]): Group knowledge transfer for large CNNs
DNNs with batch norm are vulnerable to skewed label (ICML'20 [111])
FLMoE (arXiv'20 [388]): Using a mixture of experts
SCAFFOLD (ICML'20 [136]): Correct the update direction (gradient)
FedNova (NeurIPS'20 [304]): Normalized average for objective inconsistency
FedDyn (ICLR'21 [2]): Based on Dynamic Regularization
FedMix (ICLR'21 [376]): Leverage Mixup without leaking privacy

FEDGEN (ICML'21 [437]): Data-free knowledge distillation via generator
MOON (CVPR'21 [172]): Align feature-level agreement with global model
FedUFO(ICCV'21 [400]): Unify feature learning and optimization objective
Fed-ZDAC (CVPRW'21 [93]): Fair fl with zero-shot data augmentation
CCVR (NeurIPS'21 [198]): Approximated GMM to calibrate classifier
FEDPA (ICLR'21 [4]) Federated learning via posterior averaging
Unlabeled auxiliary data for pretraining and distillation (TNNLS'21 [243])
FedProc (arXiv'21 [215]): Global prototypes with InfoNCE
FedOPT (ICLR'21 [237]): Adaptive optim for heterogeneity and efficiency
FedRS (KDD'21 [181]): Restricted Softmax to limit missing classes' weights
Fed^2 (KDD'21 [380]): A firm structure-feature alignment
FedDC (CVPR'22 [69]): An auxiliary local drift variable to track the gap
FedSAM (ICML'22 [234]): Local generality via sharpness ware minimization
FedLC (ICML'22 [397]): Calibrate logits based on probability of each class
FedMLB (ICML'22 [140]): Align to multiple auxiliary branches
Implicit Gradient Alignment in Federated Learning (AAAI'22 [47])
Initial learning phase plays a critical role in FL (AAAI'22 [351])
FedAlign (CVPR'22 [209]): Local learning matters in data heter FL
FedFTG (CVPR'22 [401]): Fine-tuning Global Model via Data-Free KD
Leverage conditional generative adversarial networks in FL (IJCAI'22 [327])
HarmoFL (AAAI'22 [125]): Normalize amplitude and weight-perturbation
FedNTD (NeurIPS'22 [154]): Preserve the global view for not-true classes

1.3.1 Personalized

Local model with personalization operation
FEDPER (arXiv'19 [8]): Base layers and personalized layers
Divide model into federated and private parameters (arXiv'19 [17])
FL personalization via model agnostic meta learning (arxiv'19 [126])
Mix local and global models with optimal mixing parameter (arXiv'20 [53])
share generic feature and keep a local private classifier (arXiv'20 [360])
MAFL: An initial shared model for current or new users (NeurIPS'20 [61])
pFedME (NeurIPS'20 [268]): Moreau envelopes with bi-level optimization
LG-FEDAVG (NeurIPSW'20 [184]): Global model + Local representation
FML (arXiv'20 [252]): Leverage shared model to conduct mutual learning
IFCA (NeurIPS'20 [74]): Alternately estimate the cluster identities
FEDRECON (NeurIPS'21 [259]): Model-agnostic for scale learning
FedAMP (AAAI'21 [120]): Similar clients to collaborate more
pFedHN (ICML'21 [246]): Central hypernetwork outputs personal model
Ditto (ICML'21 [175]): A lightweight personalization for global fl objective
FedAD (ICCV'21 [77]): Ensemble attention distill on unlabeled public data
FedFomo (ICLR'21 [403]): Personalized weighted combination of others
PGFL (arXiv'21 [210]): Prototypes and attention for aggregation
WAFFLe (Access'22 [94]): Weight anonymized factorization for fl
FedProto (AAAI'22 [270]): Global prototypes with L1 regularization
Personalize could benefits with small personal parameters (ICML'22 [231])

Alternating update outperforms the simultaneous update (ICML'22 [231])
kNN-Per (ICML'22 [205]): Interpolate global with local k-nearest neighbors
pFedLA (CVPR'22 [201]): Discern the importance of each layer
Channel Decoupling for Personalization Federated Learning (CVPR'22 [254])
FED-ROD (ICLR'22 [29]): Decouple generic and personalized with two tasks
SplitMix (ICLR'22 [106]): In-situ customization and adversarial robustness
FedBABU (ICLR'22 [219]): Only updates model body during FL
SFL (AAAI'22 [27]): Client-wise relations and Graph- based topology
FedSoft (AAAI'22 [242]: Soft clustered FL with proximal local updating
FedReg (ICLR'22 [341]): Regularize parameters with generated pseudo data
FedPCL (NeurIPS'22 [271]): Pretrained backbones and contrast prototypes
On the convergence of clustered federated learning (arXiv'22 [200])
FedFA (ICLR'23 [426]): Federated feature aug via Gaussian distribution

1.3.2 Model Heterogeneity

Model with different architecture
Cronus (arXiv'19 [23]): Public dataset (soft labels) is used with local data
KT-pFL (NeurIPS'21 [396]): Coefficient matrix for local predictions
FedMD (NeurIPS'19 [162]): Rely on related public data to distillation
RCFL (KDDW'20 [83]): Overlapping labels with weight aggregation
FedDF (NeurIPS'20 [186]): Ensemble distillation for model fusion
FEDMD-NFDP (IJCAI'21 [265]): FedMD in Noise-Free Differential Privacy
FedHeNN (ICML'22 [203]): CKA for architecture agnostic federated learning

1.4 Domain Skew

Focus on feature distribution inconsistency
A domain expert model and a private gating mechanism (NeurIPS'19 [228])
FADA (ICLR'20 [225]): Focus on domain adaption in federated learning
FedBN (ICLR'21 [179]): Except Batch Normalization module
FedDG (CVPR'21 [190]): Federated domain generalization in medical image
FADE (KDD'21 [107]): Yield biased and unfair models for minority groups
ADCOL (arXiv'21 [171]): Introduce discriminator module
FedSM (CVPR'22 [340]): Model selector to decide the closest model
FedCG (CVPRW'21 [19]): Cluster-driven graph FL over multiple domains
CGPFL (IJCAI'22 [274]): Handle context-level heterogeneity in FL
Generalization FL: out-of-sample gap and participation gap (ICLR'22 [382])
FPL (CVPR'23 [118]): Rethink FL under domain shift: a prototype view

1.4.1 Model Heterogeneity

Model with different architecture
FCCL (CVPR'22 [116]): Align Cross-Correlation Matrix with public data
FSMAFL (ACMMM'22 [117]): Few-shot model agnostic federated learning
Data heterogeneity limits knowledge distillation (ICLR'22 [3])

1.5 Aggregation

Model aggregation in federated learning
SPAHM (NeurIPS'19 [386]): Bayesian non-parameter to match clients' weights
FedMA (ICLR'20 [303]): Matching and average feature for global model
Robust FL with attack-adaptive aggregation (IJCAIW'21 [294])
FEDBE (ICLR'21 [28]): Aggregate via Bayesian model Ensemble
FEDDUALAVG (ICML'21 [383]) Conduct averaging in the dual space
Learnable aggregation weight via Dirichlet distribution (arXiv'21 [332])
Layer-wise model aggregation for scalable FL (CVPR'22 [159])
Select a small diverse clients, carrying representative gradient (ICLR'21 [10])
FedCor (CVPR'22 [273]): Correlation-Based Active Client Selection
MAB-RFL (IJCAI'22 [295]): Robust aggregate with adaptive client selection

1.6 Robustness

Impact of pervasive training data flaw
CoMT (AAAI'20 [91]): Craft instances and subset-select the training set
RHFL (CVPR'22 [63]): Robust fl with noisy and heterogeneous clients
FedCorr (CVPR'22 [345]) Multi-Stage FL for Label Noise Correction
MAB-RFL (IJCAI'22 [295]): Robust aggregate with adaptive client selection
FedRN (CIKM'22 [143]): Retrieve k-reliable neighbors and identify examples

1.7 Security

Secure discussion in federated learning
Norm clipping and "weak" differential privacy (arXiv'19 [266])
FL is susceptible to edge-case backdoors (NeurIPS'20 [302])
Reconstruct images from parameter gradients (NeurIPS'20 [73])
FPCA (NeurIPS'20 [79]): Combine FL and differential privacy
Private federated learning with laplacian smoothing (NeurIPSW'20 [185])
Soteria (CVPR'21 [264]) Perturb representation to decrease reconstruction
Clip FL: Convergence and Client-Level Differential Privacy (ICML'22 [412])
Obtain Private Data in FL with Modified Models (ICLR'22 [66])
Privacy Leakage of Adversarial Training Models (CVPR'22 [395])

1.8 Long-Tailed

Focus on Long-Tailed distribution
Local imbalance and global imbalance in FL (AAAI'21 [306])
CReFF (IJCAI'22 [248]): Classifier re-training with federated features
An Agnostic Approach to FL with Class Imbalance (ICLR'22 [255])

1.9 Unsupervised learning

Federated learning meets unsupervised learning
FURL (arXiv'20 [392]): Federated unsupervised representation learning.
FedSSL (ICLR'22 [438]): Federated learning meets Self-Supervised learning.
Orchestra (ICML'22 [196]): Cluster-based SSL exploits federated hierarchy

1.10 Semi-Supervised Learning

Semi-supervised learning faces federated learning
Semi-supervised Vertical FL with MultiView Training (IJCAIW'20 [135])
FedMatch (ICLR'21 [123]): Inter-client consistency and disjoint parameters
Private Semi-Supervised Federated Learning (IJCAI'22 [62])

1.11 Novel direction

I fail to classify into existing major research fields
Lazily aggregated quantized gradient for FL (TPAMI'20 [263])
Federated deep learning via neural architecture search (CVPRW'20 [99])
Federated Bayesian optimization via Thompson sampling (NeurIPS'20 [46])
Transformers are more robust to heterogeneous FL (CVPR'22 [233])
Federated hyperparameter tuning (NeurIPS'21 [137])
Communication-efficient random for multi-kernel online fl (TPAMI'21 [108])
Heterogeneous Ensemble KD for Large Models in FL (IJCAI'22 [41])
MaKEr (IJCAI'22 [32]): Meta-learning knowledge extrapolation for KG

1.12 Multi Modality

FL faces multi modality problem
FL for vision-and-language grounding problems (AAAI'20 [187])
Cross-modal federated human activity recognition (AAAI'22 [362])

1.13 Continual Learning

FL faces continual learning scenario
FL continual learning via weighted inter-client transfer (ICML'21 [375])
CFeD (IJCAI'22 [202]): Continual FL based on knowledge distillation

1.14 Application

Applications in federated learning
Intellectual property rights: A critical history (06 [207])
Federated Learning for Mobile Keyboard Prediction (arXiv'18 [97])
Federated Prediction Model for Stroke Prevention (arxiv'20 [131])
Federated Person Re-identification via Benchmark Analysis (MM'20 [439])
Federated Learning with Humor Recognition (IJCAI'20 [85])
Fedvision: An Online Visual Object Detection Platform (AAAI'20 [193])

FedDG: FL Domain Generalization on Medical Image (CVPR'21 [190])
FedReID (AAAI'21 [323]): Multi-Domain Labels for Person ReID
FedMAT (IJCAI'22 [251]): Cross-individual human activity recognition
FedScale (ICML'22 [153]): Scalable and reproducible benchmark for reality

1.15 Survey

Surey about federated learning
Federated learning with Non-IID Data (arXiv'18 [420])
Federated machine learning: Concept and applications (TIST'19 [361])
3 Approaches for Personalization with Applications to FL (arXiv'20 [204])
Survey of personalization techniques for federated learning (WordS4'20 [152])
Privacy and Robustness in FL: Attacks and Defenses (arXiv'20 [199])
FL: Challenges, Methods, and Future Directions (SPM'20 [176])
Personalized fl for intelligent IoT applications (OJ-COMS'20 [326])
Towards Personalized Federated Learning (arXiv'21 [269])
Federated ML: Survey, classification, criteria and directions (CST'21 [293])
Federated Learning: Issues in Medical Application (FDSE'21 [374])
A Survey on Federated Learning Systems (TKDE'21 [173])
A survey on security and privacy of federated learning (FGCS'21 [214])
Advances and open problems in federated learning (FTML'21 [132])
Federated learning on non-IID data: A survey (NC'21 [427])
FL on non-iid data silos: An experimental study (ICDE'22 [169])
Federated Learning for Smart Healthcare: A Survey (CSUR'22 [217])

2 Self-Supervised Learning

We show the report linear evaluation accuracy on ImageNet for ResNet50

2.1 Theory

Theoretical analysis of self-supervised learning especially contrastive learning
Similar pairs are assumed to be the same latent class (ICML'19 [245])
Good views retain task-relevant rather irrelevant info (NeurIPS'20 [285])
Alignment of positive; uniformity of feats on hypersphere (ICML'20 [309])
Multi embed, each of which is invariant to all but one aug (ICLR'21 [334])
CL is hardness-aware; τ controls the strength of penalties (CVPR'21 [298])
Reconstruction SSL with appro conditional independence (NeurIPS'21 [155])
No single self-supervised method dominates overall (CVPR'21 [59])
Positive align; Centers diverge; Augmentation concentrate (arXiv'21 [119])
CL shares low/mid but overfit semantic info of data (ICLR'21 [418])
Aug removes features associated with spurious dense noise (ICML'21 [319])
Aug creates chaos for intra-class and CL climbs ladder (ICLR'22 [314])
Dimensional collapse: vectors spans a lower-dim subspace (ICLR'22 [129])
Different SSL methods have a similar gradient structure (CVPR'22 [276])

Two hardness-aware independently with dual temperature (CVPR'22 [391])
Local geometry decides the learning behavior of SSL (arXiv'22 [440])
A probabilistic model selects optimal negative samples scale (IJCAI'22 [321])
Incorporate function class bias into transfer bounds (ICML'22 [244])
AggNCE (NeurIPS'22 [44]): Aligns the mean feature of multiple views
CL probably over-fits to shared information between views (CVPR'22 [300])
Treat each aug differently and learn hierarchical invariance (CVPR'22 [398])
MLP Improves transferability of unsupervised pretraining (CVPR'22 [313])
The intermediate layer update is governed by covariance (arXiv'20 [286])
The nonlinear learning dynamics of non-contrastive SSL (ICML'21 [280])
Dimensional collapse: vectors spans a lower-dim subspace (ICLR'22 [129])
Siamese keeps a lower variance in target than source (CVPR'22 [310])
Expansion magnifies local to global consistency in same class (ICLR'21 [315])
Spectral decomposition on augmentation graph (NeurIPS'21 [95])
SSL are more robust to imbalance than supervised learning (ICLR'22 [188])
CL disentangles domain and class rather learn domain-invar (ICML'22 [250])
Pos has more connection with same class in cross-domain (NeurIPS'22 [96])

2.2 Baseline

Some famous methods
InstDisc (CVPR'18 [329]): Instance Discriminate and memory bank
DeepCluster (ECCV'18 [20]): Cluster features and assign pseudo-labels
CPC (arXiv'18 [220]): Conduct predictive learning on sequential information
InvaSpread (CVPR'19 [370, 369]): End-to-end learning with siamese net
MOCO (CVPR'20 [102]): Query should be similar to key than others - 60.6
SimCLR (ICML'20 [34]): Data aug and nonlinear transformation - 69.3
PIRL (CVPR'20 [212]): Semantic representations should be invariant - 63.6
CMC (ECCV'20 [282]): Mutual infor maximization for different views - 66.2
BYOL (NeurIPS'20 [80]): Online and target networks interact - 74.3
SwAV (NeurIPS'20 [21]): Clusters data and enforce consistency - 71.8
SimSiam (CVPR'21 [35]): Predictor+stop-gradient is important - 71.3
W-MSE (ICML'21 [60]): Align only positive and Whitening to avoid collapse
DINO (ICCV'21 [22]): SSL provides new properties to Vision Transformer
TWIST (arXiv'21 [297]): Prediction consistency + sharpness + diversity
E-SSL (ICLR'22 [48]): Highlight the invariance and equivariance in SSL
MEC (NeurIPS'22 [191]): SSL need admit the maximum entropy

2.3 Dimension Decorrelation

Redundancy-reduction principle
Barlow Twins (ICML'21 [387]): Cross-correlation reduces redundancy - 73.2
VICReg (ICLR'22 [11]): Invariance + Variance + Covariance - 73.2
TiCo (arXiv'22 [428]): Min transform invariance and covariance contrast
ARB (CVPR'22 [407]): Group + Shuffle Dimension Alignment
ZeroCL (ICLR'22 [409]): Instance and Dimension Alignment

MetaMask (NeurIPS'22 [164]): Dimension redundancy and confounder - 73.9
VICRegL (NeurIPS'22 [12]): VICReg with local visual features

2.4 Positiveness

Discover postiveness in SSL
MixCo (NeurIPSW'20 [142]): Semi-positives from pos and neg mixup
MSF (CVPR'21 [151]): Leverage kmeans to enlarge positive pairs - 72.4
NNCLR (ICCV'21 [57]): Positive from nearest neighbor - 74.5
i-Mix (ICLR'21 [157]): Virtual labels for a batch and mixture
ContrastiveCrop (CVPR'22 [226]): Semantic-aware and center-suppressed
Un-Mix (AAAI'22 [256]): Soft distance concept on label space in CL
SDMP (CVPR'22 [238]): Mixed images as additional positive pairs
RINCE (AAAI'22 [105]): Exploit information about a similarity ranking
SNCLR (ICLR'23 [7]): Highly-correlated instances as soft neighbors - 75.3

2.5 Negativeness

Discover negativeness in SSL
DCL (NeurIPS'20 [42]): Correct the sampling bias of negative examples
MoChi (NeurIPS'20 [133]): Mix hardest negative with itself and query
ReSSL (NeurIPS'21 [421]): Instance relational consistency with different aug
NonSemNeg (NeurIPS'21 [71]): Negative samples only preserve non-semantics
HCL (ICLR'21 [239]): Hard negative are mapped nearby but should be far.
RELIC (ICLR'21 [213]): Enforce invariant prediction of proxy targets - 70.3
CO2 (ICLR'21 [316]): Similarities between query and negatives - 68.0
ISD (ICCV'21 [278]): Use a soft similarity for negative images - 69.8
DecoupCL (arXiv'21 [371]): Decouple positive and negative samples - 69.5

2.6 Positiveness and Negativeness

Leverage positive and negative simultaneously
FeatTrans (ICCV'21 [431]): Harder positives and diversified negatives
MetAug (ICML'22 [165]): Margin-injected meta feature augmentation

2.7 Cluster and Prototypes

Cluster and prototypes for SSL
DnC (ICCV'21 [281]): Train individual experts to recover local consistency
PCL (ICLR'21 [166]): Introduce prototypes to encode semantic structures
GCC (CVPR'21 [422]): Clustering-friendly features and compact cluster
HCSC (CVPR'22 [86]): Capture hierarchical semantic structures

2.8 Scene Images

Complicated sene images for SSL

ORL (NeurIPS'22 [338]): Pretrained SSL discovers object-level semantic
UniVIP (CVPR'22 [182]): Scene-scene, scene-instance, and instance-instance

2.9 Detection and Segmentation

Prior SSL focuses on classification, but ignores detection and segmentation

InsLoc (CVPR'21 [356]): Foreground is copy-pasted onto background.
PixPro (CVPR'21 [339]: A pixel-to-propagation consistency task
DetCo (CVPR'21 [336]): Multi-level supervision + Global and local CL.
PLRC (CVPR'22 [9]): Point-level region contrast for localization and holism

2.10 Distillation

SSL faces knowledge distillation

Compress (NeurIPS'20 [1]): Mimic relative similarity between the samples
SimDis (arXiv'21 [82]): A simple distillation baselines
SEED (ICLR'21 [64]): Mimic similarity distribution inferred by teacher
ProtoCPC (ICLR'22 [156]): Prototypical probabilistic discrepancy for CL
BINGO (ICLR'22 [344]): Bags related instances by matching similarities
DisCO (ECCV'22 [70]): A consistency regularization to align embedding

2.11 Supervised Learning

Introduce CL into supervised learning

SCL (NeurIPS'20 [138]: Supervised contrastive learning
HiMulConE (CVPR'22 [408]): Leverage all labels and preserve hierarchy
LOOK (ICLR'22 [65]): Consider intra-class difference via kmeans
THANOS (ICML'22 [31]): Balanced spread + class permutation invariance

2.11.1 Long Tailed

Handle long tailed problem

Labels bring supervision and bias in class-imbalanced (NeurIPS'20 [365])
KCL (ICLR'21 [134]): k-positive contrastive learning
SDCLR (ICML'21 [127]): A self-competitor online by pruning target
Hybrid-SC (CVPR'21 [308]): Hybrid network includes SCL and CE
PaCo (ICCV'21 [45]): Parametric class-wise learnable centers to re-balance
TSC (CVPR'22 [174]): Force all classes to maintain uniformity
BCL (CVPR'22 [430]): Class-averaging and class-complement
SSL is more robust to dataset imbalance (ICLR'22 [188])

3 Graph Learning

3.1 Preliminary

Base research about graph learning

Spectral networks and locally connected networks on graphs (ICLR'14 [15])
DeepWalk (SIGKDD'14 [227]): Structure by truncated random walks
node2vec (SIGKDD'16 [81]): Learn continuous representations for nodes
Fast localized convolutional filters on graphs (NeurIPS'16 [49])
GCN (ICLR'17 [146]): Localized first-order approximation of spectral gcn
GraphSAGE (NeurIPS'17 [89]): Generate embeddings from neighborhood
GAT (ICLR'18 [290]): Specify different weights to different neighborhoods
GIN (ICLR'19 [160]): GNN under the neighborhood aggregation framework
DeepGCNs (ICCV'19 [163]): Residual/dense connect and dilated conv
EdgeConv (TOG'19 [312]): Capture local geometric and recover topology
PNA (NeurIPS'20 [43]): Multiple aggregators and degree-scalers
OrthoGConv (AAAI'22 [84]) Maintain orthogonality of feature transform

3.2 Analysis

Some analysis about graph learning

GCN model is a special form of Laplacian smoothing (AAAI'18 [170])
SGC (ICML'19 [322]): Remove nonlinearity and collapsing weight matrices
DGN (NeurIPS'20 [424]): Normalize group of similar nodes
MADGAP (AAAI'20 [24]): Low information-to-noise ratio leads over-smoothing
CSSL (AAAI'21 [389]): Contrastive learning pretrain alleviates overfitting
Training GNNs for long-range face over-squashing (ICLR'21 [5])
Graph-MLP (arXiv'21 [113]): Contrast neighbor via adjacency information
AIR (SIGKDD'22 [411]): Divide GNNs into transformation and propagation
A dynamic graph attention variant for GAT (ICLR'22 [14])
DeCorr (SIGKDD'22 [128]): Feature overcorrelation with over-smoothing
OrthoReg (arXiv'22 [6]): The orthogonality of node correlation matrix
Deep GCNs suffer from poor gradient flow during training (NeurIPS'22 [122])
MLPInit (ICLR'23 [90]): Accerlate GNN via converged PeerMLP initialize
PMLP (ICLR'23 [354]): MLP in training and GNN in testing

3.3 Augmentation

Graph learning with data augmentation

DropEdge (ICLR'20 [241] TPAMI'22 [115]): Randomly removing edges
GraphMix (AAAI'21 [292]): Train a FCN jointly with the GNN
GAUG (AAAI'21 [419]): Edge predictors encode class-homophilic structure
KDGA (NeurIPS'22 [324]): Leverage KD to reduce negative augmentation
FLAG (CPVR'22 [150]): Aug node features with gradient-based adversarial

3.4 Self-Supervised Learning

Graph learning with self-supervised learning
DGI (ICLR'19 [291]): Local mutual max on graph's patch representations
InfoGraph (ICLR'20 [262]): Maximize mutual info from overall and local
GRACE (ICML'20 [435]): Remove edge, mask node: structure, attribute
GCC (SIGKDD'20 [232]): Graph contrastive coding for GNN pre-training
GraphCL (NeurIPS'20 [378]): Four types of graph augmentations for GCL
MVGRL (ICML'20 [98]): Node and graph (first-order and graph diffusion)
InfoGCL ([342]): Information-aware graph contrastive learning
GCA (WWW'21 [436]): Identify important edges and feature dimensions
AD-GCL (NeurIPS'21 [267]): Adversarial graph augmentation for GCL
CCA-SSG (NeurIPS'21 [394]): Canonical correlation analysis with GCL
GraphLoG (ICML'21 [347]): Hierarchical prototypes for global-semantic
JOAO (ICML'21 [377]): Automatically select data augmentation for GraphCL
BGRL (ICLRW'21 [279]): Graph contrastive learning with BYOL
MGSSL (NeurIPS'21 [414]): Motif captures subgraphs (functional group)
AFGRL (AAAI'22 [158]): Positive by local structure and global semantics
SAIL (AAAI'22 [381]): Intra- and inter-graph knowledge self-distillation
HGRL (CIKM'22 [30]): Preserve original feat and capture general neighbor
GraphMAE (SIGKDD'22 [110]): Recover feature with scaled cosine error
BYOV (WSDM'22 [379]): InfoMin and InfoBN guides prior learning
GSCL (arXiv'22 [218]): Rank node representation in the neighborhood
ProGCL (ICML'22 [331]): Most negatives with larger similarities are false
SimGRACE (WWW'22 [330]): GNN model with its perturbed version
BAGE (TPAMI'22 [405]):

3.5 Federated Graph Learning

Graph learning with federated scenario
FedGraphNN (ICLRW'21 [101]): Federated Learning Benchmark for GNN
GCFL (NeurIPS'21 [337]): Graph clustered FL based on gradient
FedSage (NeurIPS'21 [399]): Subgraph FL with missing neighbor generation

3.6 Confidence Calibration

Calibration of GNNs
Improve GNN calibration to increase trustworthiness (ICMLW'19 [277])
CaGCN (NeurIPS'21 [311]): A topology-aware post-hoc GNNs calibration
GATS (NeurIPS'22 [112]): Nodewise temperature scaling with attention
DR-GST (WWW'22 [189]): High-confidence nodes bring distribution shift

3.7 Distillation

Knowledge distillation in graph learning
LSP (CVPR'20 [364]): Local structure preserving for GCN Distillation

TinyGNN (SIGKDD'20 [350]): Peer-aware module with neighbor distillation
RDD (SIGMOD'20 [410]): Define node and edge reliability in a graph
G-CRD (arXiv'21 [130]): Preserve global topology via contrastive learning
GFKD (IJCAI'21 [51]): GNN Distill without observable graph
GLNN (ICLR'22 [406]): Eliminate graph dependency via distilled MLPs
LLP (arXiv'22 [87]): Distill relational knowledge for MLP
BGNN (arXiv'22 [88]): Complementary knowledge from different GNNs
MSKD (AAAI'22 [390]): Learn from multiple GNNs with attention weight
GKD (NeurIPS'22 [355]): Neural Heat Kernel to encapsulate geometric

3.7.1 Self-Knowledge Distillation

Self-knowledge distillation in graph learning
IGSD (arXiv'20 [393]): Iterative self-distillation to contrast graph pairs
GNN-SD (IJCAI'21 [36]): Shallow layers as supervision for deeper layers
GSDN (arXiv'22 [325]): Graph self-distillation on neighborhood

3.8 Multimodal Learning

Multimodal learning with Graph learning
MMGCN (ACMMM'19 [318]): Multi-modal GCN for Recommendation

3.9 HyperGraph

HyperGraph learning
HyperGCL (NeurIPS'22[317]): Fabricate and generate contrastive views for

3.10 Application

Some applications on graph learning
PinSage (KDD'18 [373]): GCNN for web-scale recommender systems
GAMENet (AAAI'19 [247]): Medical recommend via drug-drug interact
SemGCN (CVPR'19 [417]) Local and global semantic relations for Pose Regression
Graph convolutional neural network in social networks (NaNA'19 [284])

3.11 Survey

Some survey about Graph Learning
Survey of graph: Problems, techniques, and applications (TKDE'18 [18])
A Comprehensive Survey on Graph Neural Networks (TNNLS'20 [328])
Survey on Graph Structure Learning (arXiv'21 [434])
Graph self-supervised learning: A survey (TKDE'22 [194])
Data augmentation for deep graph learning: A survey (arXiv'22 [55])

4 Knowledge Distillation

4.1 Logits

Logits-level knowledge distillation

KD (arXiv'15 [104]): Transfer knowledge from big to small model
ESKD (ICCV'19 [40]): More accurate teacher is not a good teacher
TAKD (AAAI'20 [211]): An assistant to bridge the gap
PAD (ECCV'20 [413]): Adaptive sample weighting: easy is highlighted
Tf-KD (CVPR'20 [384]): KD is a learned label smoothing regularization
GLD (ICCV'21 [145]): global and local logits + relational information
MT-BERT (ACL'21 [320]): Multi-teacher for language model compression.
SFTN (NeurIPS'21 [221]): Learn the teacher which is friendly to students
DKD (CVPR'22 [416]): Target distillation and non-target distillation
DIST (NeurIPS'22 [114]): Correlation-based loss for intra and inter-class
ATS (NeurIPS'22 [180]): Guidance, Smooth and class discriminability
NKD (arXiv'22 [366]): Rethinking knowledge distillation via cross-entropy

4.2 Feature

Feature-level knowledge distillation

FitNet (ICLR'15 [240]): Map the hidden layer from student to teacher
AT (ICLR'17 [147]): Optimize FitNet and use the attention map of features
Overhaul (ICCV'19 [103]): Feature transform with margin ReLU
SSKD (ECCV'20 [343]): Combine self-supervision with KD
FNKD (ECCV'20 [346]): Feature normalized replace the unified temperature
HSAKD (AAAI'21 [352]): Capture hierarchical self-supervised knowledge
SemCKD (AAAI'21 [26]): Assign each tea layers for stu via attention
DKMF (TPAMI'21 [299]): Mimick both feature magnitude and direction
ReviewKD (CVPR'21 [33]): Cross-stage connection
SCKD (ICCV'21 [433]): Relative gradient direction control distillation
CID (NeurIPS'21 [50]): Capture sample and class and remove bias
SimKD (CVPR'22 [25]): Align feature and reuse teacher classifier
MGD (ECCV'22 [367]): Mask student feature and generate teacher feature
KF (ECCV'22 [363]): Decompose pretrained network into factor networks

4.3 Relation

Relation-level knowledge distillation

GiftKD (CVPR'17 [372]): FSP computes inner product between layers
RKD (CVPR'19 [223]): Pairwise and ternary relations of examples
IRG (CVPR'19 [192]): Instance relationships and feature transformation
CCKD (ICCV'19 [224]): Transfer correlation between instances
SP (ICCV'19 [287]): Preserve the pairwise similarities
CRD (ICLR'20 [283]): A contrastive objective to transfer knowledge
Intra-category and inter-category structure distillation (ECCV'20 [38])

HKD (CVPR'21 [425]): Distill holistic knowledge based on graph
CRCD (CVPR'21 [429]): Inter-sample relation and representation learning
REFILLED (TPAMI'22 [368]): Comparison match and adaptive local kd

4.4 Self-Distillation

Self Distillation Paper
BAN (ICML'18 [67]): An ensemble of multiple students generations
DDSD (AAAI'19 [348]): Align feature and logits on data distortion
Snapshot (CVPR'19 [358]): Signal from prior iteration rather generation
MultiExit (ICCV'19 [230]): Encourages early exits to mimic later exits
BYOT (ICCV'19 [402]): Deeper knowledge is squeezed into shallow ones
SAD (ICCV'19 [109]): Attention maps from deeper as targets for lower layers
CS-KD (CVPR'20 [385]): Align output of different sample of same class
Self-Distillation as Instance-Specific Label Smoothing (NeurIPS'20 [415])
PS-KD (ICCV'21 [141]): Previous epoch model as teacher to conduct KD
BAKE (arXiv'21 [72]): Inter-sample affinities weight to provide soft label
KR (AAAI'21 [56]: Alleviate optimizing conflict and maintain class relations
FRSKD (CVPR'21 [124]): Utilize an auxiliary self-teacher network
DLB (CVPR'22 [253]): Constrain half of batch coinciding with previous
Zipf (ECCV'22 [183]): Use prediction to get soft label under Zipf distribution
MixSKD (ECCV'22 [353]): MixUp alignment on feature and logits
Tf-FD (ECCV'22 [167]): Reusing channel and layer-wise features
FER (AAAI'22 [333]): Restrict behavior on the correctly-classified samples

4.5 Theory

Some explanation about knowledge distillation.
Compress the function learned by a complex into a smaller (KDD'06 [16])
LUPI (JMLR'15 [289]): Similarity control and knowledge transfer
RNN (ICASSP'16 [275]): Simple model is effectively to guide complex model
Unifying distillation and privileged information (ICLR'16 [195])
Convolution models need to be deep and convolutional (ICLR'17 [288])
The teacher needs to be more tolerant (lower accuracy) (AAAI'19 [357])
Data geometry, optimization bias and strong monotonicity (ICML'19 [229])
In noise-free data, teacher label smoothing hurts student (NeurIPS'19 [216])
Rescale student gradient and inject prior knowledge (arXiv'20 [272])
KD brings various visual concepts and stable optimization (CVPR'20 [39])
In noisy data, teacher label smoothing benefits student (ICML'20 [197])
label smooth decrease variance and lower mean prediction (ICLR'21 [257])
MSE does not elongate representation compared with KL (IJCAI'21 [144])
Good student doesn't imply good distillation fidelity (NeurIPS'21 [261])
Rethink soft label: sample-wise bias-variance tradeoff (ICLR'21 [423])
More and different knowledge points + stable optimization (TPAMI'22 [404])
Link KD to information bottleneck principle (NeurIPS'22 [296])
Pruned teacher is more effective than the original teacher (ECCV'22 [222])

SHAKE (NeurIPS'22 [168]): Extra shadow head as a proxy teacher
PEBKD (ECCV'22 [52]): Student adaptively find blind knowledge region
Teacher knowledge of some classes is undistillable (NeurIPS'22 [432])
Good DA reduces covariance of teacher-student cross-entropy (NeurIPS'22 [301])

4.6 Survey

Some survey about knowledge distillation.
Knowledge Distillation: A Survey (IJCV'21 [78])
Knowledge distillation: A review and new outlooks (PAMI'21 [307])

References

[1] Soroush Abbasi Koohpayegani, Ajinkya Tejankar, and Hamed Pirsiavash. Compress: Self-supervised learning by compressing representations. In *NeurIPS*, pages 12980–12992, 2020. 11

[2] Durmus Alp Emre Acar, Yue Zhao, Ramon Matas, Matthew Mattina, Paul Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. In *ICLR*, 2021. 3

[3] Andrei Afonin and Sai Praneeth Karimireddy. Towards model-agnostic federated learning using knowledge distillation. In *ICLR*, 2022. 5

[4] Maruan Al-Shedivat, Jennifer Gillenwater, Eric Xing, and Afshin Rostamizadeh. Federated learning via posterior averaging: A new perspective and practical algorithms. In *ICLR*, 2021. 4

[5] Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *ICLR*, 2021. 12

[6] Anonymous. Orthoreg: Improving graph-regularized MLPs via orthogonality regularization. *arXiv*, 2022. 12

[7] Anonymous. Soft neighbors are positive supporters in contrastive visual representation learning. In *ICLR*, 2023. 10

[8] Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019. 4

[9] Yutong Bai, Xinlei Chen, Alexander Kirillov, Alan Yuille, and Alexander C Berg. Point-level region contrast for object detection pre-training. In *CVPR*, pages 16061–16070, 2022. 11

[10] Ravikumar Balakrishnan, Tian Li, Tianyi Zhou, Nageen Himayat, Virginia Smith, and Jeff Bilmes. Diverse client selection for federated learning via submodular maximization. In *ICLR*, 2021. 6

[11] Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *ICLR*, 2022. 9

[12] Adrien Bardes, Jean Ponce, and Yann LeCun. Vicregl: Self-supervised learning of local visual features. In *NeurIPS*, 2022. 10

[13] Sameer Bibikar, Haris Vikalo, Zhangyang Wang, and Xiaohan Chen. Federated dynamic sparse training: Computing less, communicating less, yet learning better. In *AAAI*, pages 6080–6088, 2022. 3

[14] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *ICLR*, 2022. 12

[15] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *ICLR*, 2014. 12

[16] Cristian Buciluǎ, Rich Caruana, and Alexandru Niculescu-Mizil. Model compression. In *ACM SIGKDD*, pages 535–541, 2006. 16

[17] Duc Bui, Kshitiz Malik, Jack Goetz, Honglei Liu, Seungwhan Moon, Anuj Kumar, and Kang G Shin. Federated user representation learning. *arXiv preprint arXiv:1909.12535*, 2019. 4

[18] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE TKDE*, 30(9):1616–1637, 2018. 14

[19] Debora Caldarola, Massimiliano Mancini, Fabio Galasso, Marco Ciccone, Emanuele Rodolà, and Barbara Caputo. Cluster-driven graph federated learning over multiple domains. In *CVPRW*, pages 2749–2758, 2021. 5

[20] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *ECCV*, pages 132–149, 2018. 9

[21] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *NeurIPS*, pages 9912–9924, 2020. 9

[22] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, pages 9650–9660, 2021. 9

[23] Hongyan Chang, Virat Shejwalkar, Reza Shokri, and Amir Houmansadr. Cronus: Robust and heterogeneous collaborative learning with black-box knowledge transfer. *arXiv preprint arXiv:1912.11279*, 2019. 5

[24] Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *AAAI*, pages 3438–3445, 2020. 12

[25] Defang Chen, Jian-Ping Mei, Hailin Zhang, Can Wang, Yan Feng, and Chun Chen. Knowledge distillation with the reused teacher classifier. In *CVPR*, pages 11933–11942, 2022. 15

[26] Defang Chen, Jian-Ping Mei, Yuan Zhang, Can Wang, Zhe Wang, Yan Feng, and Chun Chen. Cross-layer distillation with semantic calibration. In *AAAI*, pages 7028–7036, 2021. 15

[27] Fengwen Chen, Guodong Long, Zonghan Wu, Tianyi Zhou, and Jing Jiang. Personalized federated learning with graph. In *AAAI*, 2022. 5

[28] Hong-You Chen and Wei-Lun Chao. Fedbe: Making bayesian model ensemble applicable to federated learning. In *ICLR*, 2021. 6

[29] Hong-You Chen and Wei-Lun Chao. On bridging generic and personalized federated learning for image classification. In *ICLR*, 2022. 5

[30] Jingfan Chen, Guanghui Zhu, Yifan Qi, Chunfeng Yuan, and Yihua Huang. Towards self-supervised learning on graphs with heterophily. In *CIKM*, pages 201–211, 2022. 13

[31] Mayee Chen, Daniel Y Fu, Avanika Narayan, Michael Zhang, Zhao Song, Kayvon Fatahalian, and Christopher Ré. Perfectly balanced: Improving transfer and robustness of supervised contrastive learning. In *ICML*, pages 3090–3122, 2022. 11

[32] Mingyang Chen, Wen Zhang, Zhen Yao, Xiangnan Chen, Mengxiao Ding, Fei Huang, and Huajun Chen. Meta-learning based knowledge extrapolation for knowledge graphs in the federated setting. In *IJCAI*, pages 1966–1972, 2022. 7

[33] Pengguang Chen, Shu Liu, Hengshuang Zhao, and Jiaya Jia. Distilling knowledge via knowledge review. In *CVPR*, pages 5008–5017, 2021. 15

[34] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607, 2020. 9

[35] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *CVPR*, pages 15750–15758, 2021. 9

[36] Yuzhao Chen, Yatao Bian, Xi Xiao, Yu Rong, Tingyang Xu, and Junzhou Huang. On self-distilling graph neural network. In *IJCAI*, 2021. 14

[37] Zihan Chen, Kai Fong Ernest Chong, and Tony QS Quek. Dynamic attention-based communication-efficient federated learning. In *IJCAI Workshop*, 2021. 3

[38] Zailiang Chen, Xianxian Zheng, Hailan Shen, Ziyang Zeng, Yukun Zhou, and Rongchang Zhao. Improving knowledge distillation via category structure. In *ECCV*, pages 205–219, 2020. 15

[39] Xu Cheng, Zhefan Rao, Yilan Chen, and Quanshi Zhang. Explaining knowledge distillation by quantifying the knowledge. In *CVPR*, pages 12925–12935, 2020. 16

[40] Jang Hyun Cho and Bharath Hariharan. On the efficacy of knowledge distillation. In *CVPR*, pages 4794–4802, 2019. 15

[41] Yae Jee Cho, Andre Manoel, Gauri Joshi, Robert Sim, and Dimitrios Dimitriadis. Heterogeneous ensemble knowledge transfer for training large models in federated learning. In *IJCAI*, 2022. 7

[42] Ching-Yao Chuang, Joshua Robinson, Yen-Chen Lin, Antonio Torralba, and Stefanie Jegelka. Debiased contrastive learning. In *NeurIPS*, pages 8765–8775, 2020. 10

[43] Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal neighbourhood aggregation for graph nets. In *NeurIPS*, pages 13260–13271, 2020. 12

[44] Jingyi Cui, Weiran Huang, Yifei Wang, and Yisen Wang. Aggnce: Asymptotically identifiable contrastive learning. In *NeurIPS*, 2022. 9

[45] Jiequan Cui, Zhisheng Zhong, Shu Liu, Bei Yu, and Jiaya Jia. Parametric contrastive learning. In *ICCV*, pages 715–724, 2021. 11

[46] Zhongxiang Dai, Bryan Kian Hsiang Low, and Patrick Jaillet. Federated bayesian optimization via thompson sampling. In *NeurIPS*, pages 9687–9699, 2020. 7

[47] Yatin Dandi, Luis Barba, and Martin Jaggi. Implicit gradient alignment in distributed and federated learning. In *AAAI*, pages 6454–6462, 2022. 4

[48] Rumen Dangovski, Li Jing, Charlotte Loh, Seungwook Han, Akash Srivastava, Brian Cheung, Pulkit Agrawal, and Marin Soljacic. Equivariant self-supervised learning: Encouraging equivariance in representations. In *ICLR*, 2022. 9

[49] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *NeurIPS*, 2016. 12

[50] Xiang Deng and Zhongfei Zhang. Comprehensive knowledge distillation with causal intervention. In *NeurIPS*, volume 34, pages 22158–22170, 2021. 15

[51] Xiang Deng and Zhongfei Zhang. Graph-free knowledge distillation for graph neural networks. In *IJCAI*, 2021. 14

[52] Xiang Deng, Jian Zheng, and Zhongfei Zhang. Personalized education: Blind knowledge distillation. In *ECCV*, 2022. 17

[53] Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. Adaptive personalized federated learning. *arXiv preprint arXiv:2003.13461*, 2020. 4

[54] Enmao Diao, Jie Ding, and Vahid Tarokh. Heterofl: Computation and communication efficient federated learning for heterogeneous clients. In *ICLR*, 2021. 3

[55] Kaize Ding, Zhe Xu, Hanghang Tong, and Huan Liu. Data augmentation for deep graph learning: A survey. *arXiv preprint arXiv:2202.08235*, 2022. 14

[56] Qianggang Ding, Sifan Wu, Tao Dai, Hao Sun, Jiadong Guo, Zhang-Hua Fu, and Shutao Xia. Knowledge refinery: Learning from decoupled label. In *AAAI*, pages 7228–7235, 2021. 16

[57] Debidatta Dwibedi, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet, and Andrew Zisserman. With a little help from my friends: Nearest-neighbor contrastive learning of visual representations. In *ICCV*, pages 9588–9597, 2021. 10

[58] Anis Elgabli, Chaouki Ben Issaid, Amrit Singh Bedi, Ketan Rajawat, Mehdi Bennis, and Vaneet Aggarwal. Fednew: A communication-efficient and privacy-preserving newton-type method for federated learning. In *ICML*, pages 5861–5877, 2022. 3

[59] Linus Ericsson, Henry Gouk, and Timothy M Hospedales. How well do self-supervised models transfer? In *CVPR*, pages 5414–5423, 2021. 8

[60] Aleksandr Ermolov, Aliaksandr Siarohin, Enver Sangineto, and Nicu Sebe. Whitening for self-supervised representation learning. In *ICML*, pages 3015–3024, 2021. 9

[61] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In *NeurIPS*, pages 3557–3568, 2020. 4

[62] Chenyou Fan, Junjie Hu, and Jianwei Huang. Private semi-supervised federated learning. In *IJCAI*, pages 2009–2015, 2022. 7

[63] Xiuwen Fang and Mang Ye. Robust federated learning with noisy and heterogeneous clients. In *CVPR*, 2022. 6

[64] Zhiyuan Fang, Jianfeng Wang, Lijuan Wang, Lei Zhang, Yezhou Yang, and Zicheng Liu. Seed: Self-supervised distillation for visual representation. In *ICLR*, 2021. 11

[65] Yutong Feng, Jianwen Jiang, Mingqian Tang, Rong Jin, and Yue Gao. Rethinking supervised pre-training for better downstream transferring. In *ICLR*, 2022. 11

[66] Liam Fowl, Jonas Geiping, Wojtek Czaja, Micah Goldblum, and Tom Goldstein. Robbing the fed: Directly obtaining private data in federated learning with modified models. In *ICLR*, 2022. 6

[67] Tommaso Furlanello, Zachary Lipton, Michael Tschannen, Laurent Itti, and Anima Anandkumar. Born again neural networks. In *ICML*, pages 1607–1616, 2018. 16

[68] Hongchang Gao, An Xu, and Heng Huang. On the convergence of communication-efficient local sgd for federated learning. In *AAAI*, pages 7510–7518, 2021. 3

[69] Liang Gao, Huazhu Fu, Li Li, Yingwen Chen, Ming Xu, and Cheng-Zhong Xu. Feddc: Federated learning with non-iid data via local drift decoupling and correction. In *CVPR*, 2022. 4

[70] Yuting Gao, Jia-Xin Zhuang, Ke Li, Hao Cheng, Xiaowei Guo, Feiyue Huang, Rongrong Ji, and Xing Sun. Disco: Remedy self-supervised learning on lightweight models with distilled contrastive learning. In *ECCV*, 2022. 11

[71] Songwei Ge, Shlok Mishra, Chun-Liang Li, Haohan Wang, and David Jacobs. Robust contrastive learning using negative samples with diminished semantics. In *NeurIPS*, 2021. 10

[72] Yixiao Ge, Ching Lam Choi, Xiao Zhang, Peipei Zhao, Feng Zhu, Rui Zhao, and Hongsheng Li. Self-distillation with batch knowledge ensembling improves imagenet classification. *arXiv preprint arXiv:2104.13298*, 2021. 16

[73] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients-how easy is it to break privacy in federated learning? In *NeurIPS*, pages 16937–16947, 2020. 6

[74] Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. In *NeurIPS*, pages 19586–19597, 2020. 4

[75] Margalit Glasgow, Honglin Yuan, and Tengyu Ma. Sharp bounds for federated averaging (local sgd) and continuous perspective. In *AISTATS*, 2022. 3

[76] Jack Goetz, Kshitiz Malik, Duc Bui, Seungwhan Moon, Honglei Liu, and Anuj Kumar. Active federated learning. *arXiv preprint arXiv:1909.12641*, 2019. 3

[77] Xuan Gong, Abhishek Sharma, Srikrishna Karanam, Ziyan Wu, Terrence Chen, David Doermann, and Arun Innanje. Ensemble attention distillation for privacy-preserving federated learning. In *ICCV*, pages 15076–15086, 2021. 4

[78] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *IJCV*, pages 1789–1819, 2021. 17

[79] Andreas Grammenos, Rodrigo Mendoza Smith, Jon Crowcroft, and Cecilia Mascolo. Federated principal component analysis. In *NeurIPS*, volume 33, pages 6453–6464, 2020. 6

[80] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent - a new approach to self-supervised learning. In *NeurIPS*, pages 21271–21284, 2020. 9

[81] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *ACM SIGKDD*, pages 855–864, 2016. 12

[82] Jindong Gu, Wei Liu, and Yonglong Tian. Simple distillation baselines for improving small self-supervised models. *arXiv preprint arXiv:2106.11304*, 2021. 11

[83] Gautham Krishna Gudur and Satheesh Kumar Perepu. Resource-constrained federated learning with heterogeneous labels and models for human activity recognition. In *ACM SIGKDD Workshop*, pages 57–69, 2020. 5

[84] Kai Guo, Kaixiong Zhou, Xia Hu, Yu Li, Yi Chang, and Xin Wang. Orthogonal graph neural networks. In *AAAI*, pages 3996–4004, 2022. 12

[85] Xu Guo, Pengwei Xing, Siwei Feng, Boyang Li, and Chunyan Miao. Federated learning with diversified preference for humor recognition. In *IJCAI Workshop*, 2020. 7

[86] Yuanfan Guo, Minghao Xu, Jiawen Li, Bingbing Ni, Xuanyu Zhu, Zhenbang Sun, and Yi Xu. Hcsc: Hierarchical contrastive selective coding. In *CVPR*, pages 9706–9715, 2022. 10

[87] Zhichun Guo, William Shiao, Shichang Zhang, Yozen Liu, Nitesh Chawla, Neil Shah, and Tong Zhao. Linkless link prediction via relational distillation. *arXiv preprint arXiv:2210.05801*, 2022. 14

[88] Zhichun Guo, Chunhui Zhang, Yujie Fan, Yijun Tian, Chuxu Zhang, and Nitesh Chawla. Boosting graph neural networks via adaptive knowledge distillation. *arXiv preprint arXiv:2210.05920*, 2022. 14

[89] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NeurIPS*, 2017. 12

[90] Xiaotian Han, Tong Zhao, Yozen Liu, Xia Hu, and Neil Shah. MLPInit: Embarrassingly simple GNN training acceleration with MLP initialization. In *ICLR*, 2023. 12

[91] Yufei Han and Xiangliang Zhang. Robust federated learning via collaborative machine teaching. In *AAAI*, pages 4075–4082, 2020. 6

[92] Filip Hanzely, Slavomír Hanzely, Samuel Horváth, and Peter Richtárik. Lower bounds and optimal algorithms for personalized federated learning. In *NeurIPS*, pages 2304–2315, 2020. 3

[93] Weituo Hao, Mostafa El-Khamy, Jungwon Lee, Jianyi Zhang, Kevin J Liang, Changyou Chen, and Lawrence Carin Duke. Towards fair federated learning with zero-shot data augmentation. In *CVPRW*, pages 3310–3319, 2021. 4

[94] Weituo Hao, Nikhil Mehta, Kevin J Liang, Pengyu Cheng, Mostafa El-Khamy, and Lawrence Carin. Waffle: Weight anonymized factorization for federated learning. *Access*, pages 49207–49218, 2022. 4

[95] Jeff Z HaoChen, Colin Wei, Adrien Gaidon, and Tengyu Ma. Provable guarantees for self-supervised deep learning with spectral contrastive loss. In *NeurIPS*, pages 5000–5011, 2021. 9

[96] Jeff Z HaoChen, Colin Wei, Ananya Kumar, and Tengyu Ma. Beyond separability: Analyzing the linear transferability of contrastive representations to related subpopulations. In *NeurIPS*, 2022. 9

[97] Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*, 2018. 7

[98] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In *ICML*, pages 4116–4126, 2020. 13

[99] Chaoyang He, Murali Annavaram, and Salman Avestimehr. Fednas: Federated deep learning via neural architecture search. In *CVPRW*, 2020. 7

[100] Chaoyang He, Murali Annavaram, and Salman Avestimehr. Group knowledge transfer: Federated learning of large cnns at the edge. In *NeurIPS*, pages 14068–14080, 2020. 3

[101] Chaoyang He, Keshav Balasubramanian, Emir Ceyani, Carl Yang, Han Xie, Lichao Sun, Lifang He, Liangwei Yang, Philip S Yu, Yu Rong, et al. Fedgraphnn: A federated learning benchmark system for graph neural networks. In *ICLR Workshop*, 2021. 13

[102] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020. 9

[103] Byeongho Heo, Jeesoo Kim, Sangdoo Yun, Hyojin Park, Nojun Kwak, and Jin Young Choi. A comprehensive overhaul of feature distillation. In *ICCV*, pages 1921–1930, 2019. 15

[104] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 15

[105] David T Hoffmann, Nadine Behrmann, Juergen Gall, Thomas Brox, and Mehdi Noroozi. Ranking info noise contrastive estimation: Boosting contrastive learning via ranked positives. In *AAAI*, 2022. 10

[106] Junyuan Hong, Haotao Wang, Zhangyang Wang, and Jiayu Zhou. Efficient split-mix federated learning for on-demand and in-situ customization. In *ICLR*, 2022. 5

[107] Junyuan Hong, Zhuangdi Zhu, Shuyang Yu, Zhangyang Wang, Hiroko H Dodge, and Jiayu Zhou. Federated adversarial debiasing for fair and transferable representations. In *ACM SIGKDD*, pages 617–627, 2021. 5

[108] Songnam Hong and Jeongmin Chae. Communication-efficient randomized algorithm for multi-kernel online federated learning. *IEEE PAMI*, 44(12):9872–9886, 2021. 7

[109] Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy. Learning lightweight lane detection cnns by self attention distillation. In *ICCV*, pages 1013–1021, 2019. 16

[110] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *ACM SIGKDD*, pages 594–604, 2022. 13

[111] Kevin Hsieh, Amar Phanishayee, Onur Mutlu, and Phillip Gibbons. The non-iid data quagmire of decentralized machine learning. In *ICML*, pages 4387–4398, 2020. 3

[112] Hans Hao-Hsun Hsu, Yuesong Shen, Christian Tomani, and Daniel Cremers. What makes graph neural networks miscalibrated? In *NeurIPS*, 2022. 13

[113] Yang Hu, Haoxuan You, Zhecan Wang, Zhicheng Wang, Erjin Zhou, and Yue Gao. Graph-mlp: node classification without message passing in graph. *arXiv preprint arXiv:2106.04051*, 2021. 12

[114] Tao Huang, Shan You, Fei Wang, Chen Qian, and Chang Xu. Knowledge distillation from a stronger teacher. In *NeurIPS*, 2022. 15

[115] Wenbing Huang, Yu Rong, Tingyang Xu, Fuchun Sun, and Junzhou Huang. Tackling over-smoothing for general graph convolutional networks. *IEEE PAMI*, 2020. 12

[116] Wenke Huang, Mang Ye, and Bo Du. Learn from others and be yourself in heterogeneous federated learning. In *CVPR*, 2022. 5

[117] Wenke Huang, Mang Ye, Bo Du, and Xiang Gao. Few-shot model agnostic federated learning. In *ACM MM*, pages 7309–7316, 2022. 5

[118] Wenke Huang, Mang Ye, Zekun Shi, He Li, and Bo Du. Rethinking federated learning with domain shift: A prototype view. In *CVPR*, 2023. 5

[119] Weiran Huang, Mingyang Yi, and Xuyang Zhao. Towards the generalization of contrastive self-supervised learning. *arXiv preprint arXiv:2111.00743*, 2021. 8

[120] Yutao Huang, Lingyang Chu, Zirui Zhou, Lanjun Wang, Jiangchuan Liu, Jian Pei, and Yong Zhang. Personalized cross-silo federated learning on non-iid data. In *AAAI*, pages 7865–7873, 2021. 4

[121] Nam Hyeon-Woo, Moon Ye-Bin, and Tae-Hyun Oh. Fedpara: Low-rank hadamard product for communication-efficient federated learning. In *ICLR*, 2022. 3

[122] Ajay Jaiswal, Peihao Wang, Tianlong Chen, Justin F Rousseau, Ying Ding, and Zhangyang Wang. Old can be gold: Better gradient flow can make vanilla-gcns great again. In *NeurIPS*, 2022. 12

[123] Wonyong Jeong, Jaehong Yoon, Eunho Yang, and Sung Ju Hwang. Federated semi-supervised learning with inter-client consistency & disjoint learning. In *ICLR*, 2021. 7

[124] Mingi Ji, Seungjae Shin, Seunghyun Hwang, Gibeom Park, and Il-Chul Moon. Refine myself by teaching myself: Feature refinement via self-knowledge distillation. In *CVPR*, pages 10664–10673, 2021. 16

[125] Meirui Jiang, Zirui Wang, and Qi Dou. Harmofl: Harmonizing local and global drifts in federated learning on heterogeneous medical images. In *AAAI*, pages 1087–1095, 2022. 4

[126] Yihan Jiang, Jakub Konečnỳ, Keith Rush, and Sreeram Kannan. Improving federated learning personalization via model agnostic meta learning. *arXiv preprint arXiv:1909.12488*, 2019. 4

[127] Ziyu Jiang, Tianlong Chen, Bobak J Mortazavi, and Zhangyang Wang. Self-damaging contrastive learning. In *ICML*, pages 4927–4939, 2021. 11

[128] Wei Jin, Xiaorui Liu, Yao Ma, Charu Aggarwal, and Jiliang Tang. Feature overcorrelation in deep graph neural networks: A new perspective. In *ACM SIGKDD*, pages 709–719, 2022. 12

[129] Li Jing, Pascal Vincent, Yann LeCun, and Yuandong Tian. Understanding dimensional collapse in contrastive self-supervised learning. In *ICLR*, 2022. 8, 9

[130] Chaitanya K Joshi, Fayao Liu, Xu Xun, Jie Lin, and Chuan-Sheng Foo. On representation knowledge distillation for graph neural networks. *arXiv preprint arXiv:2111.04964*, 2021. 14

[131] Ce Ju, Ruihui Zhao, Jichao Sun, Xiguang Wei, Bo Zhao, Yang Liu, Hongshan Li, Tianjian Chen, Xinwei Zhang, Dashan Gao, et al. Privacy-preserving technology to help millions of people: Federated prediction model for stroke prevention. *arXiv preprint arXiv:2006.10517*, 2020. 7

[132] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Found. Trends Mach. Learn.*, pages 1–210, 2021. 8

[133] Yannis Kalantidis, Mert Bulent Sariyildiz, Noe Pion, Philippe Weinzaepfel, and Diane Larlus. Hard negative mixing for contrastive learning. In *NeurIPS*, pages 21798–21809, 2020. 10

[134] Bingyi Kang, Yu Li, Sa Xie, Zehuan Yuan, and Jiashi Feng. Exploring balanced feature spaces for representation learning. In *ICLR*, 2021. 11

[135] Yan Kang, Yang Liu, and Tianjian Chen. Fedmvt: Semi-supervised vertical federated learning with multiview training. In *IJCAI Workshop*, 2020. 7

[136] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank J Reddi, Sebastian U Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for on-device federated learning. In *ICML*, 2020. 3

[137] Mikhail Khodak, Renbo Tu, Tian Li, Liam Li, Maria-Florina F Balcan, Virginia Smith, and Ameet Talwalkar. Federated hyperparameter tuning: Challenges, baselines, and connections to weight-sharing. In *NeurIPS*, pages 19184–19197, 2021. 7

[138] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *NeurIPS*, 2020. 11

[139] Geeho Kim, Jinkyu Kim, and Bohyung Han. Communication-efficient federated learning with acceleration of global momentum. *arXiv preprint arXiv:2201.03172*, 2022. 3

[140] Jinkyu Kim, Geeho Kim, and Bohyung Han. Multi-level branched regularization for federated learning. In *ICML*, pages 11058–11073, 2022. 4

[141] Kyungyul Kim, ByeongMoon Ji, Doyoung Yoon, and Sangheum Hwang. Self-knowledge distillation with progressive refinement of targets. In *ICCV*, pages 6567–6576, 2021. 16

[142] Sungnyun Kim, Gihun Lee, Sangmin Bae, and Se-Young Yun. Mixco: Mix-up contrastive learning for visual representation. In *NeurIPS Workshop*, 2020. 10

[143] SangMook Kim, Wonyoung Shin, Soohyuk Jang, Hwanjun Song, and Se-Young Yun. Fedrn: Exploiting k-reliable neighbors towards robust federated learning. In *CIKM*, pages 972–981, 2022. 6

[144] Taehyeon Kim, Jaehoon Oh, NakYil Kim, Sangwook Cho, and Se-Young Yun. Comparing kullback-leibler divergence and mean squared error loss in knowledge distillation. In *IJCAI*, 2021. 16

[145] Youmin Kim, Jinbae Park, YounHo Jang, Muhammad Ali, Tae-Hyun Oh, and Sung-Ho Bae. Distilling global and local with densely connected relations. In *ICCV*, pages 6290–6300, 2021. 15

[146] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017. 12

[147] Nikos Komodakis and Sergey Zagoruyko. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. In *ICLR*, 2017. 15

[148] Jakub Konečnỳ, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*, 2016. 3

[149] Jakub Konečnỳ, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016. 3

[150] Kezhi Kong, Guohao Li, Mucong Ding, Zuxuan Wu, Chen Zhu, Bernard Ghanem, Gavin Taylor, and Tom Goldstein. Robust optimization as data augmentation for large-scale graphs. In *CVPR*, pages 60–69, 2022. 12

[151] Soroush Abbasi Koohpayegani, Ajinkya Tejankar, and Hamed Pirsiavash. Mean shift for self-supervised learning. In *CVPR*, pages 10326–10335, 2021. 10

[152] Viraj Kulkarni, Milind Kulkarni, and Aniruddha Pant. Survey of personalization techniques for federated learning. In *WorldS4*, pages 794–797, 2020. 8

[153] Fan Lai, Yinwei Dai, Sanjay Singapuram, Jiachen Liu, Xiangfeng Zhu, Harsha Madhyastha, and Mosharaf Chowdhury. Fedscale: Benchmarking model and system performance of federated learning at scale. In *ICML*, pages 11814–11827, 2022. 8

[154] Gihun Lee, Minchan Jeong, Yongjin Shin, Sangmin Bae, and Se-Young Yun. Preservation of the global knowledge by not-true distillation in federated learning. In *NeurIPS*, 2022. 4

[155] Jason D Lee, Qi Lei, Nikunj Saunshi, and Jiacheng Zhuo. Predicting what you already know helps: Provable self-supervised learning. In *NeurIPS*, pages 309–323, 2021. 8

[156] Kyungmin Lee. Prototypical contrastive predictive coding. In *ICLR*, 2021. 11

[157] Kibok Lee, Yian Zhu, Kihyuk Sohn, Chun-Liang Li, Jinwoo Shin, and Honglak Lee. i-mix: A domain-agnostic strategy for contrastive representation learning. In *ICLR*, 2021. 10

[158] Namkyeong Lee, Junseok Lee, and Chanyoung Park. Augmentation-free self-supervised learning on graphs. In *AAAI*, pages 7372–7380, 2022. 13

[159] Sunwoo Lee, Tuo Zhang, Chaoyang He, and Salman Avestimehr. Layer-wise adaptive model aggregation for scalable federated learning. In *CVPR*, 2022. 6

[160] Keyulu Xu Weihua Hu Jure Leskovec and Stefanie Jegelka. How powerful are graph neural networks. In *ICLR*, 2019. 12

[161] Ang Li, Jingwei Sun, Binghui Wang, Lin Duan, Sicheng Li, Yiran Chen, and Hai Li. Lotteryfl: Personalized and communication-efficient federated learning with lottery ticket hypothesis on non-iid datasets. *arXiv preprint arXiv:2008.03371*, 2020. 3

[162] Daliang Li and Junpu Wang. Fedmd: Heterogenous federated learning via model distillation. In *NeurIPS Workshop*, 2019. 5

[163] Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. Deepgcns: Can gcns go as deep as cnns? In *CVPR*, pages 9267–9276, 2019. 12

[164] Jiangmeng Li, Wenwen Qiang, Yanan Zhang, Wenyi Mo, Changwen Zheng, Bing Su, and Hui Xiong. Metamask: Revisiting dimensional confounder for self-supervised learning. In *NeurIPS*, 2022. 10

[165] Jiangmeng Li, Wenwen Qiang, Changwen Zheng, Bing Su, and Hui Xiong. Metaug: Contrastive learning via meta feature augmentation. In *ICML*, 2022. 10

[166] Junnan Li, Pan Zhou, Caiming Xiong, and Steven CH Hoi. Prototypical contrastive learning of unsupervised representations. In *ICLR*, 2021. 10

[167] Lujun Li. Self-regulated feature learning via teacher-free feature distillation. In *ECCV*, pages 347–363, 2022. 16

[168] Lujun Li and Zhe Jin. Shadow knowledge distillation: Bridging offline and online knowledge transfer. In *NeurIPS*, 2022. 17

[169] Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-iid data silos: An experimental study. In *ICDE*, pages 965–978, 2022. 8

[170] Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *AAAI*, 2018. 12

[171] Qinbin Li, Bingsheng He, and Dawn Song. Adversarial collaborative learning on non-iid features. *arXiv*, 2021. 5

[172] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *CVPR*, pages 10713–10722, 2021. 4

[173] Qinbin Li, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Yuan Li, Xu Liu, and Bingsheng He. A survey on federated learning systems: vision, hype and reality for data privacy and protection. *IEEE TKDE*, 2021. 8

[174] Tianhong Li, Peng Cao, Yuan Yuan, Lijie Fan, Yuzhe Yang, Rogerio S Feris, Piotr Indyk, and Dina Katabi. Targeted supervised contrastive learning for long-tailed recognition. In *CVPR*, pages 6918–6928, 2022. 11

[175] Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In *ICML*, pages 6357–6368, 2021. 4

[176] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE SPM*, pages 50–60, 2020. 8

[177] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *arXiv preprint arXiv:1812.06127*, 2020. 3

[178] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang. On the convergence of fedavg on non-iid data. *arXiv preprint arXiv:1907.02189*, 2019. 3

[179] Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fed{bn}: Federated learning on non-{iid} features via local batch normalization. In *ICLR*, 2021. 5

[180] Xin-Chun Li, Wen shu Fan, Shaoming Song, Yinchuan Li, bingshuai Li, yunfeng shao, and De-Chuan Zhan. Asymmetric temperature scaling makes larger networks teach well again. In *NeurIPS*, 2022. 15

[181] Xin-Chun Li and De-Chuan Zhan. Fedrs: Federated learning with restricted softmax for label distribution non-iid data. In *ACM SIGKDD*, pages 995–1005, 2021. 4

[182] Zhaowen Li, Yousong Zhu, Fan Yang, Wei Li, Chaoyang Zhao, Yingying Chen, Zhiyang Chen, Jiahao Xie, Liwei Wu, Rui Zhao, et al. Univip: A unified framework for self-supervised visual pre-training. In *CVPR*, pages 14627–14636, 2022. 11

[183] Jiajun Liang, Linze Li, Zhaodong Bing, Borui Zhao, Yao Tang, Bo Lin, and Haoqiang Fan. Efficient one pass self-distillation with zipf's label smoothing. In *ECCV*, 2022. 16

[184] Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B Allen, Randy P Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency. Think locally, act globally: Federated learning with local and global representations. In *NeurIPS Workshop*, 2020. 4

[185] Zhicong Liang, Bao Wang, Quanquan Gu, Stanley Osher, and Yuan Yao. Exploring private federated learning with laplacian smoothing. In *NeurIPS Workshop*, 2020. 6

[186] Tao Lin, Lingjing Kong, Sebastian U Stich, and Martin Jaggi. Ensemble distillation for robust model fusion in federated learning. In *NeurIPS*, pages 2351–2363, 2020. 5

[187] Fenglin Liu, Xian Wu, Shen Ge, Wei Fan, and Yuexian Zou. Federated learning for vision-and-language grounding problems. In *AAAI*, pages 11572–11579, 2020. 7

[188] Hong Liu, Jeff Z HaoChen, Adrien Gaidon, and Tengyu Ma. Self-supervised learning is more robust to dataset imbalance. In *ICLR*, 2022. 9, 11

[189] Hongrui Liu, Binbin Hu, Xiao Wang, Chuan Shi, Zhiqiang Zhang, and Jun Zhou. Confidence may cheat: Self-training on graph neural networks under distribution shift. In *WWW*, pages 1248–1258, 2022. 13

[190] Quande Liu, Cheng Chen, Jing Qin, Qi Dou, and Pheng-Ann Heng. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *CVPR*, pages 1013–1023, 2021. 5, 8

[191] Xin Liu, Zhongdao Wang, Yali Li, and Shengjin Wang. Self-supervised learning via maximum entropy coding. In *NeurIPS*, 2022. 9

[192] Yufan Liu, Jiajiong Cao, Bing Li, Chunfeng Yuan, Weiming Hu, Yangxi Li, and Yunqiang Duan. Knowledge distillation via instance relationship graph. In *CVPR*, pages 7096–7104, 2019. 15

[193] Yang Liu, Anbu Huang, Yun Luo, He Huang, Youzhi Liu, Yuanyuan Chen, Lican Feng, Tianjian Chen, Han Yu, and Qiang Yang. Fedvision: An online visual object detection platform powered by federated learning. In *AAAI*, pages 13172–13179, 2020. 7

[194] Yixin Liu, Ming Jin, Shirui Pan, Chuan Zhou, Yu Zheng, Feng Xia, and Philip Yu. Graph self-supervised learning: A survey. *IEEE TKDE*, 2022. 14

[195] David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. Unifying distillation and privileged information. In *ICLR*, 2016. 16

[196] Ekdeep Singh Lubana, Chi Ian Tang, Fahim Kawsar, Robert P Dick, and Akhil Mathur. Orchestra: Unsupervised federated learning via globally consistent clustering. In *ICML*, 2022. 7

[197] Michal Lukasik, Srinadh Bhojanapalli, Aditya Menon, and Sanjiv Kumar. Does label smoothing mitigate label noise? In *ICML*, pages 6448–6458, 2020. 16

[198] Mi Luo, Fei Chen, Dapeng Hu, Yifan Zhang, Jian Liang, and Jiashi Feng. No fear of heterogeneity: Classifier calibration for federated learning with non-iid data. In *NeurIPS*, 2021. 4

[199] Lingjuan Lyu, Han Yu, Xingjun Ma, Lichao Sun, Jun Zhao, Qiang Yang, and Philip S Yu. Privacy and robustness in federated learning: Attacks and defenses. *arXiv preprint arXiv:2012.06337*, 2020. 8

[200] Jie Ma, Guodong Long, Tianyi Zhou, Jing Jiang, and Chengqi Zhang. On the convergence of clustered federated learning. *arXiv preprint arXiv:2202.06187*, 2022. 5

[201] Xiaosong Ma, Jie Zhang, Song Guo, and Wenchao Xu. Layer-wised model aggregation for personalized federated learning. In *CVPR*, pages 10092–10101, 2022. 5

[202] Yuhang Ma, Zhongle Xie, Jue Wang, Ke Chen, and Lidan Shou. Continual federated learning based on knowledge distillation. In *IJCAI*, pages 2182–2188, 2022. 7

[203] Disha Makhija, Xing Han, Nhat Ho, and Joydeep Ghosh. Architecture agnostic federated learning for neural networks. In *ICML*, 2022. 5

[204] Yishay Mansour, Mehryar Mohri, Jae Ro, and Ananda Theertha Suresh. Three approaches for personalization with applications to federated learning. *arXiv preprint arXiv:2002.10619*, 2020. 8

[205] Othmane Marfoq, Giovanni Neglia, Richard Vidal, and Laetitia Kameni. Personalized federated learning through local memorization. In *ICML*, pages 15070–15092, 2022. 5

[206] Othmane Marfoq, Chuan Xu, Giovanni Neglia, and Richard Vidal. Throughput-optimal topology design for cross-silo federated learning. In *NeurIPS*, pages 19478–19487, 2020. 3

[207] Christopher May and Susan K Sell. *Intellectual property rights: A critical history*. Lynne Rienner Publishers Boulder, 2006. 7

[208] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, pages 1273–1282, 2017. 3

[209] Matias Mendieta, Taojiannan Yang, Pu Wang, Minwoo Lee, Zhengming Ding, and Chen Chen. Local learning matters: Rethinking data heterogeneity in federated learning. In *CVPR*, pages 8397–8406, 2022. 4

[210] Umberto Michieli and Mete Ozay. Prototype guided federated learning of visual feature representations. *arXiv preprint arXiv:2105.08982*, 2021. 4

[211] Seyed Iman Mirzadeh, Mehrdad Farajtabar, Ang Li, Nir Levine, Akihiro Matsukawa, and Hassan Ghasemzadeh. Improved knowledge distillation via teacher assistant. In *AAAI*, pages 5191–5198, 2020. 15

[212] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *CVPR*, pages 6707–6717, 2020. 9

[213] Jovana Mitrovic, Brian McWilliams, Jacob Walker, Lars Buesing, and Charles Blundell. Representation learning via invariant causal mechanisms. In *ICLR*, 2021. 10

[214] Viraaji Mothukuri, Reza M Parizi, Seyedamin Pouriyeh, Yan Huang, Ali Dehghantanha, and Gautam Srivastava. A survey on security and privacy of federated learning. *Future Generation Computer Systems*, pages 619–640, 2021. 8

[215] Xutong Mu, Yulong Shen, Ke Cheng, Xueli Geng, Jiaxuan Fu, Tao Zhang, and Zhiwei Zhang. Fedproc: Prototypical contrastive federated learning on non-iid data. *arXiv preprint arXiv:2109.12273*, 2021. 4

[216] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. When does label smoothing help? In *NeurIPS*, 2019. 16

[217] Dinh C Nguyen, Quoc-Viet Pham, Pubudu N Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia Dobre, and Won-Joo Hwang. Federated learning for smart healthcare: A survey. *CSUR*, pages 1–37, 2022. 8

[218] Zhiyuan Ning, Pengfei Wang, Pengyang Wang, Ziyue Qiao, Wei Fan, Denghui Zhang, Yi Du, and Yuanchun Zhou. Graph soft-contrastive learning via neighborhood ranking. *arXiv preprint arXiv:2209.13964*, 2022. 13

[219] Jaehoon Oh, Sangmook Kim, and Se-Young Yun. Fedbabu: Towards enhanced representation for federated image classification. In *ICLR*, 2022. 5

[220] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 9

[221] Dae Young Park, Moon-Hyun Cha, Daesin Kim, Bohyung Han, et al. Learning student-friendly teacher networks for knowledge distillation. In *NeurIPS*, pages 13292–13303, 2021. 15

[222] Jinhyuk Park and Albert No. Prune your model before distill it. In *ECCV*, pages 120–136, 2022. 16

[223] Wonpyo Park, Dongju Kim, Yan Lu, and Minsu Cho. Relational knowledge distillation. In *CVPR*, pages 3967–3976, 2019. 15

[224] Baoyun Peng, Xiao Jin, Jiaheng Liu, Dongsheng Li, Yichao Wu, Yu Liu, Shunfeng Zhou, and Zhaoning Zhang. Correlation congruence for knowledge distillation. In *CVPR*, pages 5007–5016, 2019. 15

[225] Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. Federated adversarial domain adaptation. In *ICLR*, 2020. 5

[226] Xiangyu Peng, Kai Wang, Zheng Zhu, Mang Wang, and Yang You. Crafting better contrastive views for siamese representation learning. In *CVPR*, pages 16031–16040, 2022. 10

[227] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *ACM SIGKDD*, pages 701–710, 2014. 12

[228] Daniel Peterson, Pallika Kanani, and Virendra J Marathe. Private federated learning with domain adaptation. In *NeurIPS*, 2019. 5

[229] Mary Phuong and Christoph Lampert. Towards understanding knowledge distillation. In *ICML*, pages 5142–5151, 2019. 16

[230] Mary Phuong and Christoph H Lampert. Distillation-based training for multi-exit architectures. In *ICCV*, pages 1355–1364, 2019. 16
[231] Krishna Pillutla, Kshitiz Malik, Abdel-Rahman Mohamed, Mike Rabbat, Maziar Sanjabi, and Lin Xiao. Federated learning with partial model personalization. In *ICML*, pages 17716–17758, 2022. 4, 5
[232] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. Gcc: Graph contrastive coding for graph neural network pre-training. In *ACM SIGKDD*, pages 1150–1160, 2020. 13
[233] Liangqiong Qu, Yuyin Zhou, Paul Pu Liang, Yingda Xia, Feifei Wang, Ehsan Adeli, Li Fei-Fei, and Daniel Rubin. Rethinking architecture design for tackling data heterogeneity in federated learning. In *CVPR*, pages 10061–10071, 2022. 7
[234] Zhe Qu, Xingyu Li, Rui Duan, Yao Liu, Bo Tang, and Zhuo Lu. Generalized federated learning via sharpness aware minimization. In *ICML*, 2022. 4
[235] Zhaonan Qu, Kaixiang Lin, Jayant Kalagnanam, Zhaojian Li, Jiayu Zhou, and Zhengyuan Zhou. Federated learning's blessing: Fedavg has linear speedup. In *ICLR*, 2021. 3
[236] Sashank Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečnỳ, Sanjiv Kumar, and H Brendan McMahan. Adaptive federated optimization. In *ICLR*, 2021. 3
[237] Sashank Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečnỳ, Sanjiv Kumar, and H Brendan McMahan. Adaptive federated optimization. In *ICLR*, 2021. 4
[238] Sucheng Ren, Huiyu Wang, Zhengqi Gao, Shengfeng He, Alan Yuille, Yuyin Zhou, and Cihang Xie. A simple data mixing prior for improving self-supervised learning. In *CVPR*, pages 14595–14604, 2022. 10
[239] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. Contrastive learning with hard negative samples. In *ICLR*, 2021. 10
[240] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. Fitnets: Hints for thin deep nets. In *ICLR*, 2015. 15
[241] Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. Dropedge: Towards deep graph convolutional networks on node classification. In *ICLR*, 2020. 12
[242] Yichen Ruan and Carlee Joe-Wong. Fedsoft: Soft clustered federated learning with proximal local updating. In *AAAI*, pages 8124–8131, 2022. 5
[243] Felix Sattler, Tim Korjakow, Roman Rischke, and Wojciech Samek. Fedaux: Leveraging unlabeled auxiliary data in federated learning. *IEEE TNNLS*, 2021. 4
[244] Nikunj Saunshi, Jordan Ash, Surbhi Goel, Dipendra Misra, Cyril Zhang, Sanjeev Arora, Sham Kakade, and Akshay Krishnamurthy. Understanding contrastive learning requires incorporating inductive biases. In *ICML*, 2022. 9
[245] Nikunj Saunshi, Orestis Plevrakis, Sanjeev Arora, Mikhail Khodak, and Hrishikesh Khandeparkar. A theoretical analysis of contrastive unsupervised representation learning. In *ICML*, pages 5628–5637, 2019. 8
[246] Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. Personalized federated learning using hypernetworks. In *ICML*, pages 9489–9502, 2021. 4
[247] Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. Gamenet: Graph augmented memory networks for recommending medication combination. In *AAAI*, pages 1126–1133, 2019. 14

[248] Xinyi Shang, Yang Lu, Gang Huang, and Hanzi Wang. Federated learning on heterogeneous and long-tailed data via classifier re-training with federated features. In *IJCAI*, 2022. 6

[249] Pranay Sharma, Rohan Panda, Gauri Joshi, and Pramod Varshney. Federated minimax optimization: Improved convergence analyses and algorithms. In *ICML*, pages 19683–19730, 2022. 3

[250] Kendrick Shen, Robbie Jones, Ananya Kumar, Sang Michael Xie, Jeff Z HaoChen, Tengyu Ma, and Percy Liang. Connect, not collapse: Explaining contrastive learning for unsupervised domain adaptation. In *ICML*, 2022. 9

[251] Qiang Shen, Haotian Feng, Rui Song, Stefano Teso, Fausto Giunchiglia, and Hao Xu. Federated multi-task attention for cross-individual human activity recognition. In *IJCAI*, 2022. 8

[252] Tao Shen, Jie Zhang, Xinkang Jia, Fengda Zhang, Gang Huang, Pan Zhou, Kun Kuang, Fei Wu, and Chao Wu. Federated mutual learning. *arXiv preprint arXiv:2006.16765*, 2020. 4

[253] Yiqing Shen, Liwu Xu, Yuzhe Yang, Yaqian Li, and Yandong Guo. Self-distillation from the last mini-batch for consistency regularization. In *CVPR*, pages 11943–11952, 2022. 16

[254] Yiqing Shen, Yuyin Zhou, and Lequan Yu. Cd2-pfed: Cyclic distillation-guided channel decoupling for model personalization in federated learning. In *CVPR*, pages 10041–10050, 2022. 5

[255] Zebang Shen, Juan Cervino, Hamed Hassani, and Alejandro Ribeiro. An agnostic approach to federated learning with class imbalance. In *ICLR*, 2022. 6

[256] Zhiqiang Shen, Zechun Liu, Zhuang Liu, Marios Savvides, Trevor Darrell, and Eric Xing. Un-mix: Rethinking image mixtures for unsupervised visual representation learning. In *AAAI*, pages 2216–2224, 2022. 10

[257] Zhiqiang Shen, Zechun Liu, Dejia Xu, Zitian Chen, Kwang-Ting Cheng, and Marios Savvides. Is label smoothing truly incompatible with knowledge distillation: An empirical study. In *ICLR*, 2021. 16

[258] Neta Shoham, Tomer Avidor, Aviv Keren, Nadav Israel, Daniel Benditkis, Liron Mor-Yosef, and Itai Zeitak. Overcoming forgetting in federated learning on non-iid data. In *NeurIPS Workshop*, 2019. 3

[259] Karan Singhal, Hakim Sidahmed, Zachary Garrett, Shanshan Wu, John Rush, and Sushant Prakash. Federated reconstruction: Partially local federated learning. In *NeurIPS*, volume 34, pages 11220–11232, 2021. 4

[260] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated multi-task learning. In *NeurIPS*, 2017. 3

[261] Samuel Stanton, Pavel Izmailov, Polina Kirichenko, Alexander A Alemi, and Andrew G Wilson. Does knowledge distillation really work? In *NeurIPS*, 2021. 16

[262] Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. In *ICLR*, 2020. 13

[263] Jun Sun, Tianyi Chen, Georgios B Giannakis, Qinmin Yang, and Zaiyue Yang. Lazily aggregated quantized gradient innovation for communication-efficient federated learning. *IEEE PAMI*, 44(4):2031–2044, 2020. 7

[264] Jingwei Sun, Ang Li, Binghui Wang, Huanrui Yang, Hai Li, and Yiran Chen. Provable defense against privacy leakage in federated learning from representation perspective. In *CVPR*, 2021. 6

[265] Lichao Sun and Lingjuan Lyu. Federated model distillation with noise-free differential privacy. In *IJCAI*, 2021. 5

[266] Ziteng Sun, Peter Kairouz, Ananda Theertha Suresh, and H Brendan McMahan. Can you really backdoor federated learning? *arXiv preprint arXiv:1911.07963*, 2019. 6

[267] Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. Adversarial graph augmentation to improve graph contrastive learning. *NeurIPS*, 19:15920–15933, 2021. 13

[268] Canh T. Dinh, Nguyen Tran, and Josh Nguyen. Personalized federated learning with moreau envelopes. In *NeurIPS*, pages 21394–21405, 2020. 4

[269] Alysa Ziying Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *arXiv preprint arXiv:2103.00710*, 2021. 8

[270] Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated prototype learning across heterogeneous clients. In *AAAI*, 2022. 4

[271] Yue Tan, Guodong Long, Jie Ma, Lu Liu, Tianyi Zhou, and Jing Jiang. Federated learning from pre-trained models: A contrastive learning approach. In *NeurIPS*, 2022. 5

[272] Jiaxi Tang, Rakesh Shivanna, Zhe Zhao, Dong Lin, Anima Singh, Ed H Chi, and Sagar Jain. Understanding and improving knowledge distillation. *arXiv preprint arXiv:2002.03532*, 2020. 16

[273] Minxue Tang, Xuefei Ning, Yitu Wang, Jingwei Sun, Yu Wang, Hai Li, and Yiran Chen. Fedcor: Correlation-based active client selection strategy for heterogeneous federated learning. In *CVPR*, pages 10102–10111, 2022. 6

[274] Xueyang Tang, Song Guo, and Jingcai Guo. Personalized federated learning with contextualized generalization. In *IJCAI*, pages 2241–2247, 2022. 5

[275] Zhiyuan Tang, Dong Wang, and Zhiyong Zhang. Recurrent neural network training with dark knowledge transfer. In *ICASSP*, pages 5900–5904, 2016. 16

[276] Chenxin Tao, Honghui Wang, Xizhou Zhu, Jiahua Dong, Shiji Song, Gao Huang, and Jifeng Dai. Exploring the equivalence of siamese self-supervised learning via a unified gradient framework. In *CVPR*, pages 14431–14440, 2022. 8

[277] Leonardo Teixeira, Brian Jalaian, and Bruno Ribeiro. Are graph neural networks miscalibrated? In *ICML Workshop*, 2019. 13

[278] Ajinkya Tejankar, Soroush Abbasi Koohpayegani, Vipin Pillai, Paolo Favaro, and Hamed Pirsiavash. Isd: Self-supervised learning by iterative similarity distillation. In *ICCV*, 2021. 10

[279] Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Rémi Munos, Petar Veličković, and Michal Valko. Bootstrapped representation learning on graphs. In *ICLR Workshop*, 2021. 13

[280] Yuandong Tian, Xinlei Chen, and Surya Ganguli. Understanding self-supervised learning dynamics without contrastive pairs. In *ICML*, 2021. 9

[281] Yonglong Tian, Olivier J Henaff, and Aäron van den Oord. Divide and contrast: Self-supervised learning from uncurated data. In *ICCV*, pages 10063–10074, 2021. 10

[282] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *ECCV*, pages 776–794. Springer, 2020. 9

[283] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive representation distillation. In *ICLR*, 2020. 15

[284] Yupan Tian, Yang Niu, Jun Yan, and Feng Tian. Inferring private attributes based on graph convolutional neural network in social networks. In *NaNA*, pages 186–190, 2019. 14

[285] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. What makes for good views for contrastive learning? In *NeurIPS*, volume 33, pages 6827–6839, 2020. 8

[286] Yuandong Tian, Lantao Yu, Xinlei Chen, and Surya Ganguli. Understanding self-supervised learning with dual deep networks. *arXiv preprint arXiv:2010.00578*, 2020. 9

[287] Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *ICCV*, pages 1365–1374, 2019. 15

[288] Gregor Urban, Krzysztof J Geras, Samira Ebrahimi Kahou, Ozlem Aslan, Shengjie Wang, Rich Caruana, Abdelrahman Mohamed, Matthai Philipose, and Matt Richardson. Do deep convolutional nets really need to be deep and convolutional? In *ICLR*, 2017. 16

[289] Vladimir Vapnik, Rauf Izmailov, et al. Learning using privileged information: similarity control and knowledge transfer. *JMLR*, pages 2023–2049, 2015. 16

[290] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018. 12

[291] Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. In *ICLR*, 2019. 13

[292] Vikas Verma, Meng Qu, Kenji Kawaguchi, Alex Lamb, Yoshua Bengio, Juho Kannala, and Jian Tang. Graphmix: Improved training of gnns for semi-supervised learning. In *AAAI*, pages 10024–10032, 2021. 12

[293] Omar Abdel Wahab, Azzam Mourad, Hadi Otrok, and Tarik Taleb. Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems. *IEEE CST*, pages 1342–1397, 2021. 8

[294] Ching Pui Wan and Qifeng Chen. Robust federated learning with attack-adaptive aggregation. In *IJCAI Workshop*, 2021. 6

[295] Wei Wan, Shengshan Hu, Jianrong Lu, Leo Yu Zhang, Hai Jin, and Yuanyuan He. Shielding federated learning: Robust aggregation with adaptive client selection. In *IJCAI*, 2022. 6

[296] Chaofei Wang, Qisen Yang, Rui Huang, Shiji Song, and Gao Huang. Efficient knowledge distillation from model checkpoints. In *NeurIPS*, 2022. 16

[297] Feng Wang, Tao Kong, Rufeng Zhang, Huaping Liu, and Hang Li. Self-supervised learning by estimating twin class distributions. *arXiv preprint arXiv:2110.07402*, 2021. 9

[298] Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss. In *CVPR*, pages 2495–2504, 2021. 8

[299] Guo-Hua Wang, Yifan Ge, and Jianxin Wu. Distilling knowledge by mimicking features. *IEEE PAMI*, 2021. 15

[300] Haoqing Wang, Xun Guo, Zhi-Hong Deng, and Yan Lu. Rethinking minimal sufficient representation in contrastive learning. In *CVPR*, pages 16041–16050, 2022. 9

[301] Huan Wang, Suhas Lohit, Mike Jones, and Yun Fu. What makes a ”good” data augmentation in knowledge distillation – a statistical perspective. In *NeurIPS*, 2022. 17

[302] Hongyi Wang, Kartik Sreenivasan, Shashank Rajput, Harit Vishwakarma, Saurabh Agarwal, Jy-yong Sohn, Kangwook Lee, and Dimitris Papailiopoulos. Attack of the tails: Yes, you really can backdoor federated learning. In *NeurIPS*, pages 16070–16084, 2020. 6

[303] Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. In *ICLR*, 2020. 6

[304] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. In *NeurIPS*, pages 7611–7623, 2020. 3

[305] Jianyu Wang, Anit Kumar Sahu, Zhouyi Yang, Gauri Joshi, and Soummya Kar. Matcha: Speeding up decentralized sgd via matching decomposition sampling. In *ICC*, pages 299–300, 2019. 3

[306] Lixu Wang, Shichao Xu, Xiao Wang, and Qi Zhu. Addressing class imbalance in federated learning. In *AAAI*, pages 10165–10173, 2021. 6

[307] Lin Wang and Kuk-Jin Yoon. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE PAMI*, 2021. 17

[308] Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. Contrastive learning based hybrid networks for long-tailed image classification. In *CVPR*, pages 943–952, 2021. 11

[309] Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *ICML*, pages 9929–9939, 2020. 8

[310] Xiao Wang, Haoqi Fan, Yuandong Tian, Daisuke Kihara, and Xinlei Chen. On the importance of asymmetry for siamese representation learning. In *CVPR*, pages 16570–16579, 2022. 9

[311] Xiao Wang, Hongrui Liu, Chuan Shi, and Cheng Yang. Be confident! towards trustworthy graph neural networks via confidence calibration. In *NeurIPS*, pages 23768–23779, 2021. 13

[312] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *TOG*, 38(5):1–12, 2019. 12

[313] Yizhou Wang, Shixiang Tang, Feng Zhu, Lei Bai, Rui Zhao, Donglian Qi, and Wanli Ouyang. Revisiting the transferability of supervised pretraining: an mlp perspective. In *CVPR*, pages 9183–9193, 2022. 9

[314] Yifei Wang, Qi Zhang, Yisen Wang, Jiansheng Yang, and Zhouchen Lin. Chaos is a ladder: A new understanding of contrastive learning. In *ICLR*, 2022. 8

[315] Colin Wei, Kendrick Shen, Yining Chen, and Tengyu Ma. Theoretical analysis of self-training with deep networks on unlabeled data. In *ICLR*, 2020. 9

[316] Chen Wei, Huiyu Wang, Wei Shen, and Alan Yuille. Co2: Consistent contrast for unsupervised visual representation learning. In *ICLR*, 2021. 10

[317] Tianxin Wei, Yuning You, Tianlong Chen, Yang Shen, Jingrui He, and Zhangyang Wang. Augmentations in hypergraph contrastive learning: Fabricated and generative. In *NeurIPS*, 2022. 14

[318] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. Mmgcn: Multi-modal graph convolution network for personalized recommendation of micro-video. In *ACM MM*, pages 1437–1445, 2019. 14

[319] Zixin Wen and Yuanzhi Li. Toward understanding the feature learning process of self-supervised contrastive learning. In *ICML*, pages 11112–11122, 2021. 8

[320] Chuhan Wu, Fangzhao Wu, and Yongfeng Huang. One teacher is enough? pre-trained language model distillation from multiple teachers. In *ACL-IJCNLP*, pages 4408–4413, 2021. 15

[321] Chuhan Wu, Fangzhao Wu, and Yongfeng Huang. Rethinking infonce: How many negative samples do you need? In *IJCAI*, 2022. 9

[322] Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *ICML*, pages 6861–6871, 2019. 12

[323] Guile Wu and Shaogang Gong. Decentralised learning from independent multi-domain labels for person re-identification. In *AAAI*, pages 2898–2906, 2021. 8

[324] Lirong Wu, Haitao Lin, Yufei Huang, , and Stan Z. Li. Knowledge distillation improves graph structure augmentation for graph neural networks. In *NeurIPS*, 2022. 12

[325] Lirong Wu, Jun Xia, Haitao Lin, Zhangyang Gao, Zicheng Liu, Guojiang Zhao, and Stan Z Li. Teaching yourself: Graph self-distillation on neighborhood for node classification. *arXiv preprint arXiv:2210.02097*, 2022. 14

[326] Qiong Wu, Kaiwen He, and Xu Chen. Personalized federated learning for intelligent iot applications: A cloud-edge based framework. *OJ-COMS*, pages 35–44, 2020. 8

[327] Yuezhou Wu, Yan Kang, Jiahuan Luo, Yuanqin He, and Qiang Yang. Fedcg: Leverage conditional gan for protecting privacy and maintaining competitive performance in federated learning. In *IJCAI*, pages 2334–2340, 2022. 4

[328] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE TNNLS*, pages 4–24, 2020. 14

[329] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *CVPR*, pages 3733–3742, 2018. 9

[330] Jun Xia, Lirong Wu, Jintao Chen, Bozhen Hu, and Stan Z Li. Simgrace: A simple framework for graph contrastive learning without data augmentation. In *WWW*, pages 1070–1079, 2022. 13

[331] Jun Xia, Lirong Wu, Ge Wang, Jintao Chen, and Stan Z Li. Progcl: Rethinking hard negative mining in graph contrastive learning. In *ICML*, pages 24332–24346, 2022. 13

[332] Yingda Xia, Dong Yang, Wenqi Li, Andriy Myronenko, Daguang Xu, Hirofumi Obinata, Hitoshi Mori, Peng An, Stephanie Harmon, Evrim Turkbey, et al. Auto-fedavg: learnable federated averaging for multi-institutional medical image segmentation. *arXiv preprint arXiv:2104.10195*, 2021. 6

[333] Deng Xiang, Xiao Yun, Long Bo, and Zhongfei Zhang. Reducing flipping errors in deep neural networks. In *AAAI*, 2022. 16

[334] Tete Xiao, Xiaolong Wang, Alexei A Efros, and Trevor Darrell. What should not be contrastive in contrastive learning. In *ICLR*, 2021. 8

[335] Cong Xie, Sanmi Koyejo, and Indranil Gupta. Asynchronous federated optimization. *arXiv preprint arXiv:1903.03934*, 2019. 3

[336] Enze Xie, Jian Ding, Wenhai Wang, Xiaohang Zhan, Hang Xu, Peize Sun, Zhenguo Li, and Ping Luo. Detco: Unsupervised contrastive learning for object detection. In *CVPR*, pages 8392–8401, 2021. 11

[337] Han Xie, Jing Ma, Li Xiong, and Carl Yang. Federated graph classification over non-iid graphs. In *NeurIPS*, pages 18839–18852, 2021. 13

[338] Jiahao Xie, Xiaohang Zhan, Ziwei Liu, Yew Soon Ong, and Chen Change Loy. Unsupervised object-level representation learning from scene images. In *NeurIPS*, pages 28864–28876, 2021. 11

[339] Zhenda Xie, Yutong Lin, Zheng Zhang, Yue Cao, Stephen Lin, and Han Hu. Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning. In *CVPR*, pages 16684–16693, 2021. 11

[340] An Xu, Wenqi Li, Pengfei Guo, Dong Yang, Holger R Roth, Ali Hatamizadeh, Can Zhao, Daguang Xu, Heng Huang, and Ziyue Xu. Closing the generalization gap of cross-silo federated medical image segmentation. In *CVPR*, pages 20866–20875, 2022. 5

[341] Chencheng Xu, Zhiwei Hong, Minlie Huang, and Tao Jiang. Acceleration of federated learning with alleviated forgetting in local training. *arXiv preprint arXiv:2203.02645*, 2022. 5

[342] Dongkuan Xu, Wei Cheng, Dongsheng Luo, Haifeng Chen, and Xiang Zhang. Infogcl: Information-aware graph contrastive learning. In *NeurIPS*, pages 30414–30425, 2021. 13

[343] Guodong Xu, Ziwei Liu, Xiaoxiao Li, and Chen Change Loy. Knowledge distillation meets self-supervision. In *ECCV*, pages 588–604, 2020. 15

[344] Haohang Xu, Jiemin Fang, Xiaopeng Zhang, Lingxi Xie, Xinggang Wang, Wenrui Dai, Hongkai Xiong, and Qi Tian. Bag of instances aggregation boosts self-supervised distillation. In *ICLR*, 2022. 11

[345] Jingyi Xu, Zihan Chen, Tony QS Quek, and Kai Fong Ernest Chong. Fedcorr: Multi-stage federated learning for label noise correction. In *CVPR*, pages 10184–10193, 2022. 6

[346] Kunran Xu, Lai Rui, Yishi Li, and Lin Gu. Feature normalized knowledge distillation for image classification. In *ECCV*, pages 664–680, 2020. 15

[347] Minghao Xu, Hang Wang, Bingbing Ni, Hongyu Guo, and Jian Tang. Self-supervised graph-level representation learning with local and global structure. In *ICML*, 2021. 13

[348] Ting-Bing Xu and Cheng-Lin Liu. Data-distortion guided self-distillation for deep neural networks. In *AAAI*, pages 5565–5572, 2019. 16

[349] Yihao Xue, Chaoyue Niu, Zhenzhe Zheng, Shaojie Tang, Chengfei Lyu, Fan Wu, and Guihai Chen. Toward understanding the influence of individual clients in federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10560–10567, 2021. 3

[350] Bencheng Yan, Chaokun Wang, Gaoyang Guo, and Yunkai Lou. TinyGNN: Learning Efficient Graph Neural Networks. In *ACM SIGKDD*, pages 1848–1856, 2020. 14

[351] Gang Yan, Hao Wang, and Jian Li. Seizing critical learning periods in federated learning. In *AAAI*, 2022. 4

[352] Chuanguang Yang, Zhulin An, Linhang Cai, and Yongjun Xu. Hierarchical self-supervised augmented knowledge distillation. In *AAAI*, 2021. 15

[353] Chuanguang Yang, Zhulin An, Helong Zhou, Linhang Cai, Xiang Zhi, Jiwen Wu, Yongjun Xu, and Qian Zhang. Mixskd: Self-knowledge distillation from mixup for image recognition. In *ECCV*, 2022. 16

[354] Chenxiao Yang, Qitian Wu, Jiahua Wang, and Junchi Yan. Graph neural networks are inherently good generalizers: Insights by bridging gnns and mlps. In *ICLR*, 2023. 12

[355] Chenxiao Yang, Qitian Wu, and Junchi Yan. Geometric knowledge distillation: Topology compression for graph neural networks. In *NeurIPS*, 2022. 14

[356] Ceyuan Yang, Zhirong Wu, Bolei Zhou, and Stephen Lin. Instance localization for self-supervised detection pretraining. In *CVPR*, pages 3987–3996, 2021. 11

[357] Chenglin Yang, Lingxi Xie, Siyuan Qiao, and Alan Yuille. Knowledge distillation in generations: More tolerant teachers educate better students. In *AAAI*, 2018. 16

[358] Chenglin Yang, Lingxi Xie, Chi Su, and Alan L Yuille. Snapshot distillation: Teacher-student optimization in one generation. In *CVPR*, pages 2859–2868, 2019. 16

[359] Haibo Yang, Minghong Fang, and Jia Liu. Achieving linear speedup with partial worker participation in non-iid federated learning. In *ICLR*, 2021. 3

[360] Lixuan Yang, Cedric Beliard, and Dario Rossi. Heterogeneous data-aware federated learning. *arXiv preprint arXiv:2011.06393*, 2020. 4

[361] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM TIST*, pages 1–19, 2019. 8

[362] Xiaoshan Yang, Baochen Xiong, Yi Huang, and Changsheng Xu. Cross-modal federated human activity recognition via modality-agnostic and modality-specific representation learning. In *AAAI*, 2022. 7

[363] Xingyi Yang, Jingwen Ye, and Xinchao Wang. Factorizing knowledge in neural networks. In *ECCV*, 2022. 15

[364] Yiding Yang, Jiayan Qiu, Mingli Song, Dacheng Tao, and Xinchao Wang. Distilling knowledge from graph convolutional networks. In *CVPR*, pages 7074–7083, 2020. 13

[365] Yuzhe Yang and Zhi Xu. Rethinking the value of labels for improving class-imbalanced learning. In *NeurIPS*, 2020. 11

[366] Zhendong Yang, Zhe Li, Yuan Gong, Tianke Zhang, Shanshan Lao, Chun Yuan, and Yu Li. Rethinking knowledge distillation via cross-entropy. *arXiv preprint arXiv:2208.10139*, 2022. 15

[367] Zhendong Yang, Zhe Li, Mingqi Shao, Dachuan Shi, Zehuan Yuan, and Chun Yuan. Masked generative distillation. In *ECCV*, 2022. 15

[368] Han-Jia Ye, Su Lu, and De-Chuan Zhan. Generalized knowledge distillation via relationship matching. *IEEE PAMI*, 2022. 16

[369] Mang Ye, Jianbing Shen, Xu Zhang, Pong C Yuen, and Shih-Fu Chang. Augmentation invariant and instance spreading feature for softmax embedding. *IEEE PAMI*, 2020. 9

[370] Mang Ye, Xu Zhang, Pong C Yuen, and Shih-Fu Chang. Unsupervised embedding learning via invariant and spreading instance feature. In *CVPR*, pages 6210–6219, 2019. 9

[371] Chun-Hsiao Yeh, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu, Yubei Chen, and Yann LeCun. Decoupled contrastive learning. *arXiv preprint arXiv:2110.06848*, 2021. 10

[372] Junho Yim, Donggyu Joo, Jihoon Bae, and Junmo Kim. A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. In *CVPR*, pages 4133–4141, 2017. 15

[373] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *ACM SIGKDD*, pages 974–983, 2018. 14

[374] Joo Hun Yoo, Hyejun Jeong, Jaehyeok Lee, and Tai-Myoung Chung. Federated learning: Issues in medical application. In *FDSE*, pages 3–22, 2021. 8

[375] Jaehong Yoon, Wonyong Jeong, Giwoong Lee, Eunho Yang, and Sung Ju Hwang. Federated continual learning with weighted inter-client transfer. In *ICML*, pages 12073–12086, 2021. 7

[376] Tehrim Yoon, Sumin Shin, Sung Ju Hwang, and Eunho Yang. Fedmix: Approximation of mixup under mean augmented federated learning. In *ICLR*, 2021. 3

[377] Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. Graph contrastive learning automated. In *ICML*, 2021. 13

[378] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. In *NeurIPS*, pages 5812–5823, 2020. 13

[379] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. Bringing your own view: Graph contrastive learning without prefabricated data augmentations. In *WSDM*, pages 1300–1309, 2022. 13

[380] Fuxun Yu, Weishan Zhang, Zhuwei Qin, Zirui Xu, Di Wang, Chenchen Liu, Zhi Tian, and Xiang Chen. Fed2: Feature-aligned federated learning. In *ACM SIGKDD*, pages 2066–2074, 2021. 4

[381] Lu Yu, Shichao Pei, Lizhong Ding, Jun Zhou, Longfei Li, Chuxu Zhang, and Xiangliang Zhang. Sail: Self-augmented graph contrastive learning. In *AAAI*, pages 8927–8935, 2022. 13

[382] Honglin Yuan, Warren Morningstar, Lin Ning, and Karan Singhal. What do we mean by generalization in federated learning? In *ICLR*, 2022. 5

[383] Honglin Yuan, Manzil Zaheer, and Sashank Reddi. Federated composite optimization. In *ICML*, pages 12253–12266, 2021. 6

[384] Li Yuan, Francis EH Tay, Guilin Li, Tao Wang, and Jiashi Feng. Revisiting knowledge distillation via label smoothing regularization. In *CVPR*, pages 3903–3911, 2020. 15

[385] Sukmin Yun, Jongjin Park, Kimin Lee, and Jinwoo Shin. Regularizing class-wise predictions via self-knowledge distillation. In *CVPR*, June 2020. 16

[386] Mikhail Yurochkin, Mayank Agarwal, Soumya Ghosh, Kristjan Greenewald, and Nghia Hoang. Statistical model aggregation via parameter matching. In *NeurIPS*, 2019. 6

[387] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *ICML*, 2021. 9

[388] Edvin Listo Zec, John Martinsson, Olof Mogren, Leon René Sütfeld, and Daniel Gillblad. Specialized federated learning using mixture of experts. *arXiv preprint arXiv:2010.02056*, 2020. 3

[389] Jiaqi Zeng and Pengtao Xie. Contrastive self-supervised learning for graph classification. In *AAAI*, pages 10824–10832, 2021. 12

[390] Chunhai Zhang, Jie Liu, Kai Dang, and Wenzheng Zhang. Multi-scale distillation from multiple graph neural networks. In *AAAI*, 2022. 14

[391] Chaoning Zhang, Kang Zhang, Trung X Pham, Axi Niu, Zhinan Qiao, Chang D Yoo, and In So Kweon. Dual temperature helps contrastive learning without many negative samples: Towards understanding and simplifying moco. In *CVPR*, pages 14441–14450, 2022. 9

[392] Fengda Zhang, Kun Kuang, Zhaoyang You, Tao Shen, Jun Xiao, Yin Zhang, Chao Wu, Yueting Zhuang, and Xiaolin Li. Federated unsupervised representation learning. *arXiv preprint arXiv:2010.08982*, 2020. 7

[393] Hanlin Zhang, Shuai Lin, Weiyang Liu, Pan Zhou, Jian Tang, Xiaodan Liang, and Eric P Xing. Iterative graph self-distillation. *arXiv preprint arXiv:2010.12609*, 2020. 14

[394] Hengrui Zhang, Qitian Wu, Junchi Yan, David Wipf, and Philip S Yu. From canonical correlation analysis to self-supervised graph neural networks. In *NeurIPS*, volume 34, pages 76–89, 2021. 13

[395] Jingyang Zhang, Yiran Chen, and Hai Li. Privacy leakage of adversarial training models in federated learning systems. In *CVPR*, pages 108–114, 2022. 6

[396] Jie Zhang, Song Guo, Xiaosong Ma, Haozhao Wang, Wenchao Xu, and Feijie Wu. Parameterized knowledge transfer for personalized federated learning. In *NeurIPS*, pages 10092–10104, 2021. 5

[397] Jie Zhang, Zhiqi Li, Bo Li, Jianghe Xu, Shuang Wu, Shouhong Ding, and Chao Wu. Federated learning with label distribution skew via logits calibration. In *ICML*, pages 26311–26329, 2022. 4

[398] Junbo Zhang and Kaisheng Ma. Rethinking the augmentation module in contrastive learning: Learning hierarchical augmentation invariance with expanded views. In *CVPR*, pages 16650–16659, 2022. 9

[399] Ke Zhang, Carl Yang, Xiaoxiao Li, Lichao Sun, and Siu Ming Yiu. Subgraph federated learning with missing neighbor generation. In *NeurIPS*, pages 6671–6682, 2021. 13

[400] Lin Zhang, Yong Luo, Yan Bai, Bo Du, and Ling-Yu Duan. Federated learning for non-iid data via unified feature learning and optimization objective alignment. In *ICCV*, pages 4420–4428, 2021. 4

[401] Lin Zhang, Li Shen, Liang Ding, Dacheng Tao, and Ling-Yu Duan. Fine-tuning global model via data-free knowledge distillation for non-iid federated learning. In *CVPR*, pages 10174–10183, 2022. 4

[402] Linfeng Zhang, Jiebo Song, Anni Gao, Jingwei Chen, Chenglong Bao, and Kaisheng Ma. Be your own teacher: Improve the performance of convolutional neural networks via self distillation. In *ICCV*, pages 3713–3722, 2019. 16

[403] Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and Jose M Alvarez. Personalized federated learning with first order model optimization. In *ICLR*, 2021. 4

[404] Quanshi Zhang, Xu Cheng, Yilan Chen, and Zhefan Rao. Quantifying the knowledge in a dnn to explain knowledge distillation for classification. *IEEE PAMI*, 2022. 16

[405] Rui Zhang, Yunxing Zhang, Chengjun Lu, and Xuelong Li. Unsupervised graph embedding via adaptive graph learning. *IEEE PAMI*, 2022. 13

[406] Shichang Zhang, Yozen Liu, Yizhou Sun, and Neil Shah. Graph-less neural networks: Teaching old mlps new tricks via distillation. In *ICLR*, 2022. 14

[407] Shaofeng Zhang, Lyn Qiu, Feng Zhu, Junchi Yan, Hengrui Zhang, Rui Zhao, Hongyang Li, and Xiaokang Yang. Align representations with base: A new approach to self-supervised learning. In *CVPR*, pages 16600–16609, 2022. 9

[408] Shu Zhang, Ran Xu, Caiming Xiong, and Chetan Ramaiah. Use all the labels: A hierarchical multi-label contrastive learning framework. In *CVPR*, pages 16660–16669, 2022. 11

[409] Shaofeng Zhang, Feng Zhu, Junchi Yan, Rui Zhao, and Xiaokang Yang. Zero-CL: Instance and feature decorrelation for negative-free symmetric contrastive learning. In *ICLR*, 2022. 9

[410] Wentao Zhang, Xupeng Miao, Yingxia Shao, Jiawei Jiang, Lei Chen, Olivier Ruas, and Bin Cui. Reliable Data Distillation on Graph Convolutional Network. In *SIGMOD*, pages 1399–1414, 2020. 14

[411] Wentao Zhang, Zeang Sheng, Ziqi Yin, Yuezihan Jiang, Yikuan Xia, Jun Gao, Zhi Yang, and Bin Cui. Model degradation hinders deep graph neural networks. In *ACM SIGKDD*, pages 2493–2503, 2022. 12

[412] Xinwei Zhang, Xiangyi Chen, Mingyi Hong, Steven Wu, and Jinfeng Yi. Understanding clipping for federated learning: Convergence and client-level differential privacy. In *ICML*, pages 26048–26067, 2022. 6

[413] Youcai Zhang, Zhonghao Lan, Yuchen Dai, Fangao Zeng, Yan Bai, Jie Chang, and Yichen Wei. Prime-aware adaptive distillation. In *ECCV*, pages 658–674, 2020. 15

[414] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. In *NeurIPS*, pages 15870–15882, 2021. 13

[415] Zhilu Zhang and Mert Sabuncu. Self-distillation as instance-specific label smoothing. In *NeurIPS*, pages 2184–2195, 2020. 16

[416] Borui Zhao, Quan Cui, Renjie Song, Yiyu Qiu, and Jiajun Liang. Decoupled knowledge distillation. In *CVPR*, 2022. 15

[417] Long Zhao, Xi Peng, Yu Tian, Mubbasir Kapadia, and Dimitris N Metaxas. Semantic graph convolutional networks for 3d human pose regression. In *CVPR*, pages 3425–3435, 2019. 14

[418] Nanxuan Zhao, Zhirong Wu, Rynson WH Lau, and Stephen Lin. What makes instance discrimination good for transfer learning? In *ICLR*, 2021. 8

[419] Tong Zhao, Yozen Liu, Leonardo Neves, Oliver Woodford, Meng Jiang, and Neil Shah. Data augmentation for graph neural networks. In *AAAI*, pages 11015–11023, 2021. 12

[420] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*, 2018. 8

[421] Mingkai Zheng, Shan You, Fei Wang, Chen Qian, Changshui Zhang, Xiaogang Wang, and Chang Xu. Ressl: Relational self-supervised learning with weak augmentation. In *NeurIPS*, pages 2543–2555, 2021. 10

[422] Huasong Zhong, Jianlong Wu, Chong Chen, Jianqiang Huang, Minghua Deng, Liqiang Nie, Zhouchen Lin, and Xian-Sheng Hua. Graph contrastive clustering. In *CVPR*, pages 9224–9233, 2021. 10

[423] Helong Zhou, Liangchen Song, Jiajie Chen, Ye Zhou, Guoli Wang, Junsong Yuan, and Qian Zhang. Rethinking soft labels for knowledge distillation: A bias-variance tradeoff perspective. In *ICLR*, 2021. 16

[424] Kaixiong Zhou, Xiao Huang, Yuening Li, Daochen Zha, Rui Chen, and Xia Hu. Towards deeper graph neural networks with differentiable group normalization. In *NeurIPS*, pages 4917–4928, 2020. 12

[425] Sheng Zhou, Yucheng Wang, Defang Chen, Jiawei Chen, Xin Wang, Can Wang, and Jiajun Bu. Distilling holistic knowledge with graph neural networks. In *ICCV*, pages 10387–10396, 2021. 16

[426] Tianfei Zhou and Ender Konukoglu. FedFA: Federated feature augmentation. In *ICLR*, 2023. 5

[427] Hangyu Zhu, Jinjin Xu, Shiqing Liu, and Yaochu Jin. Federated learning on non-IID data: A survey. *NC*, pages 371–390, 2021. 8

[428] Jiachen Zhu, Rafael M Moraes, Serkan Karakulak, Vlad Sobol, Alfredo Canziani, and Yann LeCun. Tico: Transformation invariance and covariance contrast for self-supervised visual representation learning. *arXiv preprint arXiv:2206.10698*, 2022. 9

[429] Jinguo Zhu, Shixiang Tang, Dapeng Chen, Shijie Yu, Yakun Liu, Mingzhe Rong, Aijun Yang, and Xiaohua Wang. Complementary relation contrastive distillation. In *CVPR*, pages 9260–9269, 2021. 16

[430] Jianggang Zhu, Zheng Wang, Jingjing Chen, Yi-Ping Phoebe Chen, and Yu-Gang Jiang. Balanced contrastive learning for long-tailed visual recognition. In *CVPR*, pages 6908–6917, 2022. 11

[431] Rui Zhu, Bingchen Zhao, Jingen Liu, Zhenglong Sun, and Chang Wen Chen. Improving contrastive learning by visualizing feature transformation. In *ICCV*, pages 10306–10315, 2021. 10

[432] Yichen Zhu, Ning Liu, Zhiyuan Xu, Xin Liu, Weibin Meng, Louis Wang, Zhicai Ou, and Jian Tang. Teach less, learn more: On the undistillable classes in knowledge distillation. In *NeurIPS*, 2022. 17

[433] Yichen Zhu and Yi Wang. Student customized knowledge distillation: Bridging the gap between student and teacher. In *ICCV*, pages 5057–5066, 2021. 15

[434] Yanqiao Zhu, Weizhi Xu, Jinghao Zhang, Yuanqi Du, Jieyu Zhang, Qiang Liu, Carl Yang, and Shu Wu. A survey on graph structure learning: Progress and opportunities. *arXiv e-prints*, pages arXiv–2103, 2021. 14

[435] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. In *ICML*, 2020. 13

[436] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph contrastive learning with adaptive augmentation. In *WWW*, pages 2069–2080, 2021. 13

[437] Zhuangdi Zhu, Junyuan Hong, and Jiayu Zhou. Data-free knowledge distillation for heterogeneous federated learning. In *ICML*, pages 12878–12889, 2021. 4

[438] Weiming Zhuang, Yonggang Wen, and Shuai Zhang. Divergence-aware federated self-supervised learning. In *ICLR*, 2022. 7

[439] Weiming Zhuang, Yonggang Wen, Xuesen Zhang, Xin Gan, Daiying Yin, Dongzhan Zhou, Shuai Zhang, and Shuai Yi. Performance optimization of federated person re-identification via benchmark analysis. In *ACM MM*, pages 955–963, 2020. 7

[440] Liu Ziyin, Ekdeep Singh Lubana, Masahito Ueda, and Hidenori Tanaka. What shapes the loss landscape of self-supervised learning? *arXiv preprint arXiv:2210.00638*, 2022. 9