Qualcomm

Qualcomm AI Research†

A Mutual Information Perspective on Federated Contrastive Learning

Christos Louizos ♤, Matthias Reisser, Denis Korzhenkov

♤ Eng., Senior Staff, Qualcomm Technologies Netherlands B.V.

{clouizos,mreisser,dkorzhenkov}@qti.qualcomm.com

† Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Self-supervised learning in the federated context

Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?

Self-supervised learning in the federated context

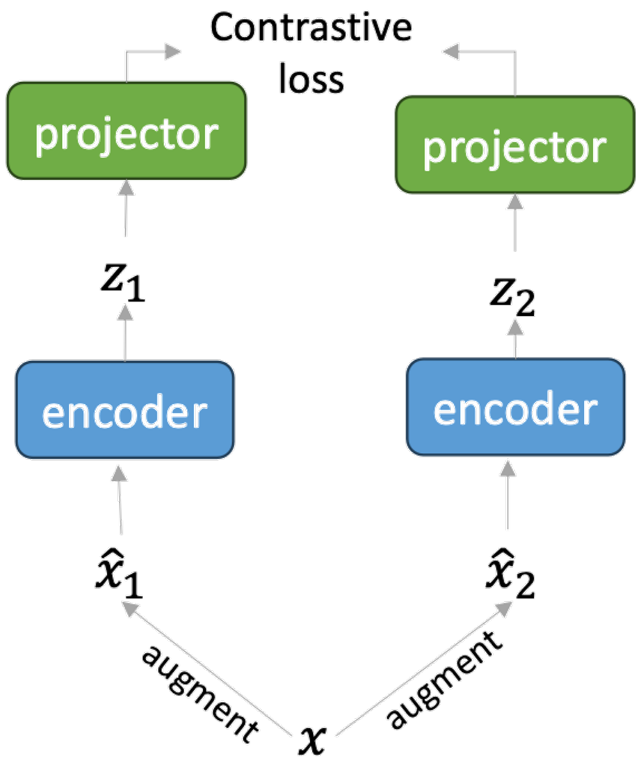
- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset

Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand

Self-supervised learning in the federated context

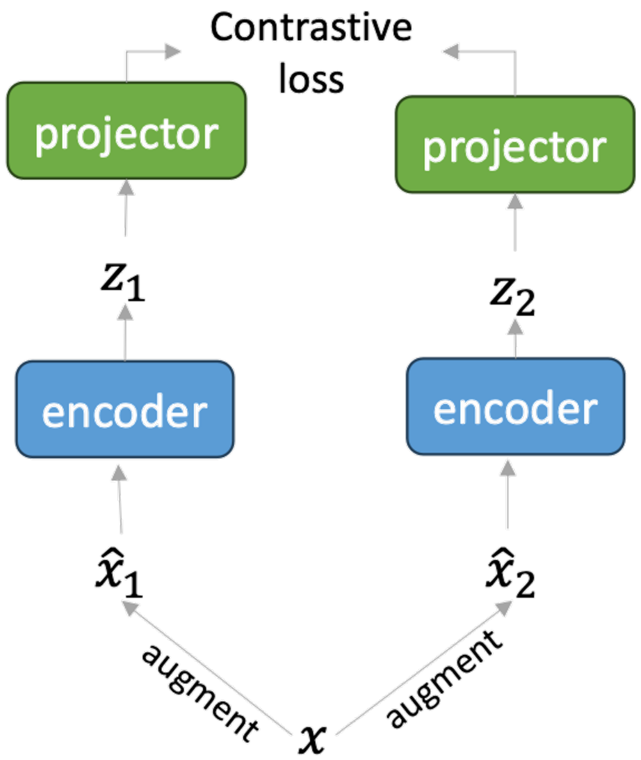
- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?



SimCLR

Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?
- This work:
  <u>We extend SimCLR to FL through a mutual-information perspective</u>

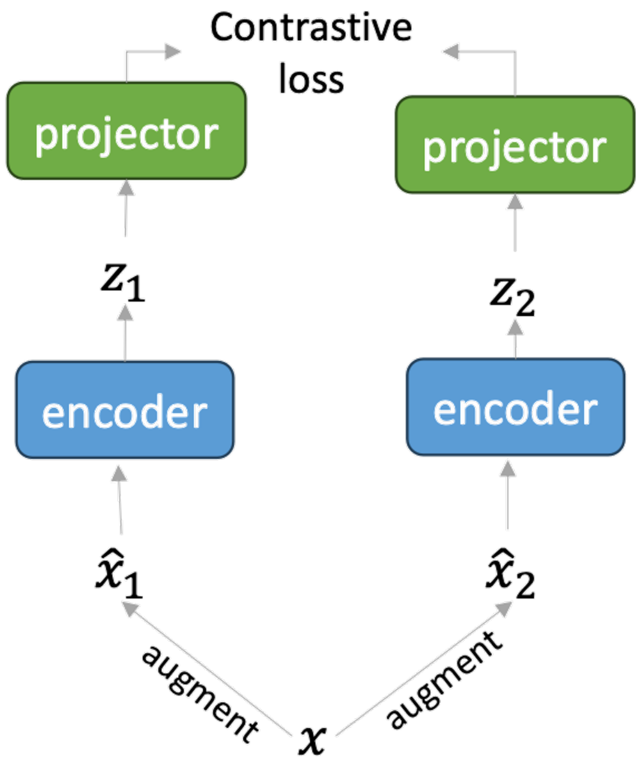


SimCLR

Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?
- This work:
  We extend SimCLR to FL through a mutual-information perspective

Bonus: extension also covers the semi-supervised setting



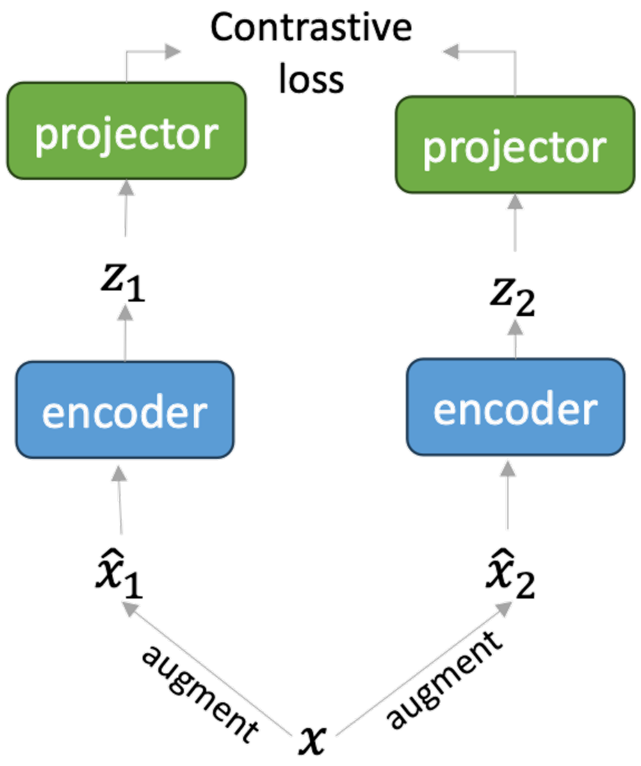
SimCLR

Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?
- This work:
  <u>We extend SimCLR to FL through a mutual-information perspective</u>

Bonus: extension also covers the semi-supervised setting

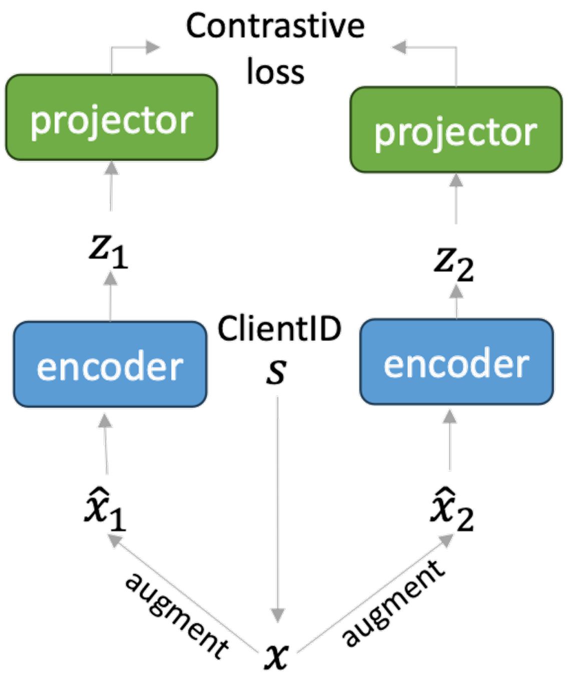
Second Bonus: insights translate to other self-supervised methods



SimCLR

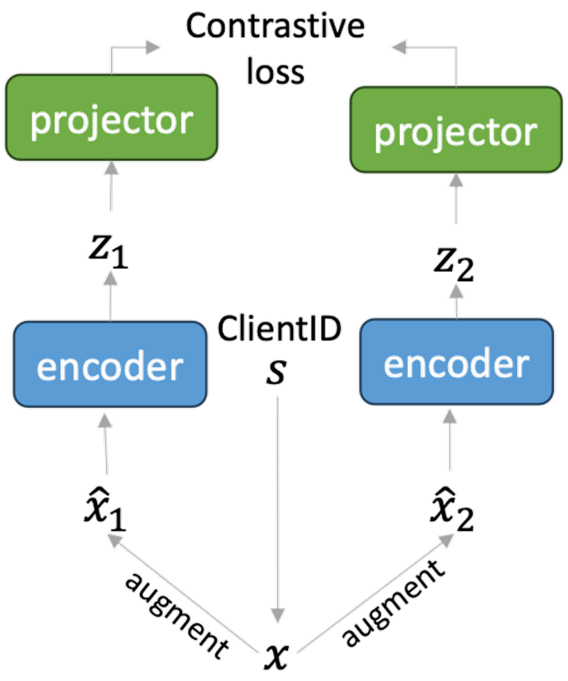
The SimCLR variants considered

The SimCLR variants considered

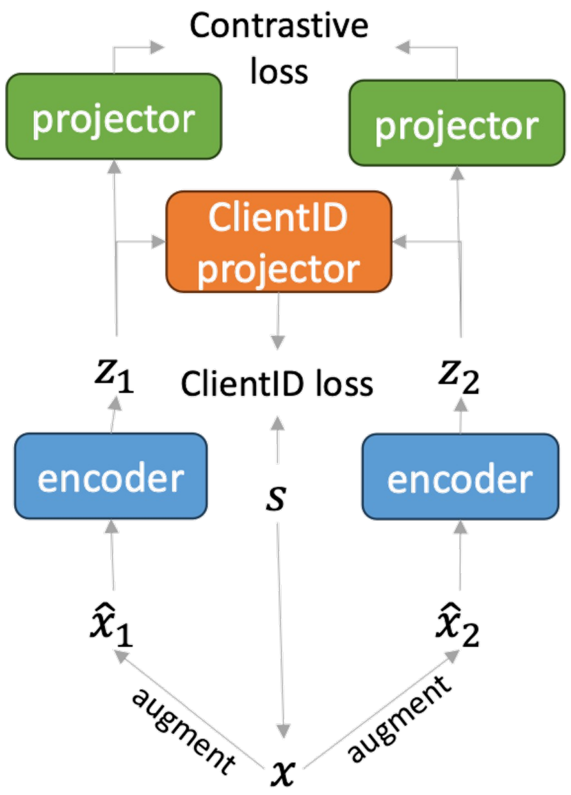


Local SimCLR: each client applies SimCLR on their own data, thus the federation implicitly optimizes a lower bound to the local $I(\boldsymbol{z}_1;\ \boldsymbol{z}_2|s)$

The SimCLR variants considered

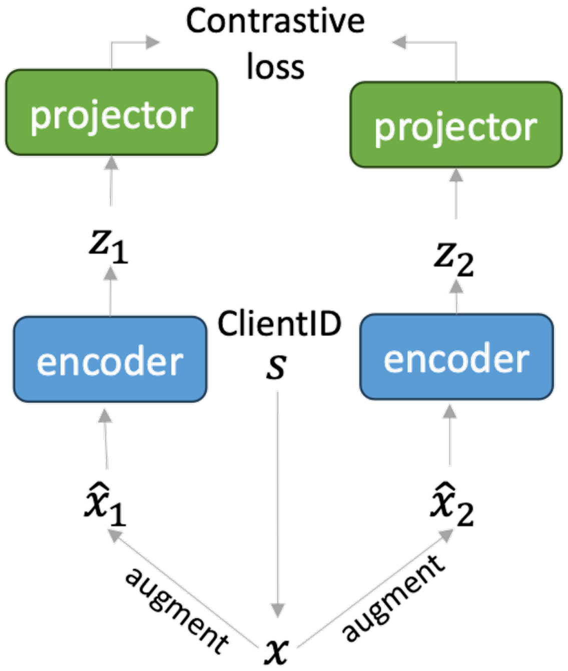


Local SimCLR: each client applies SimCLR on their own data, thus the federation implicitly optimizes a lower bound to the local $I(\boldsymbol{z}_1;\ \boldsymbol{z}_2|s)$

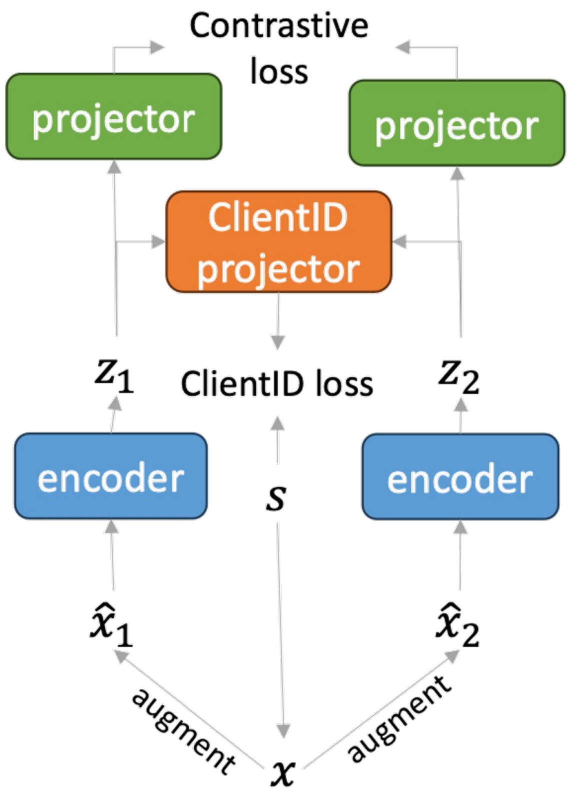


Federated SimCLR: each client also optimizes a client classifier, thus the federation implicitly optimizes a lower bound to the global $I(\boldsymbol{z}_1;\ \boldsymbol{z}_2)$

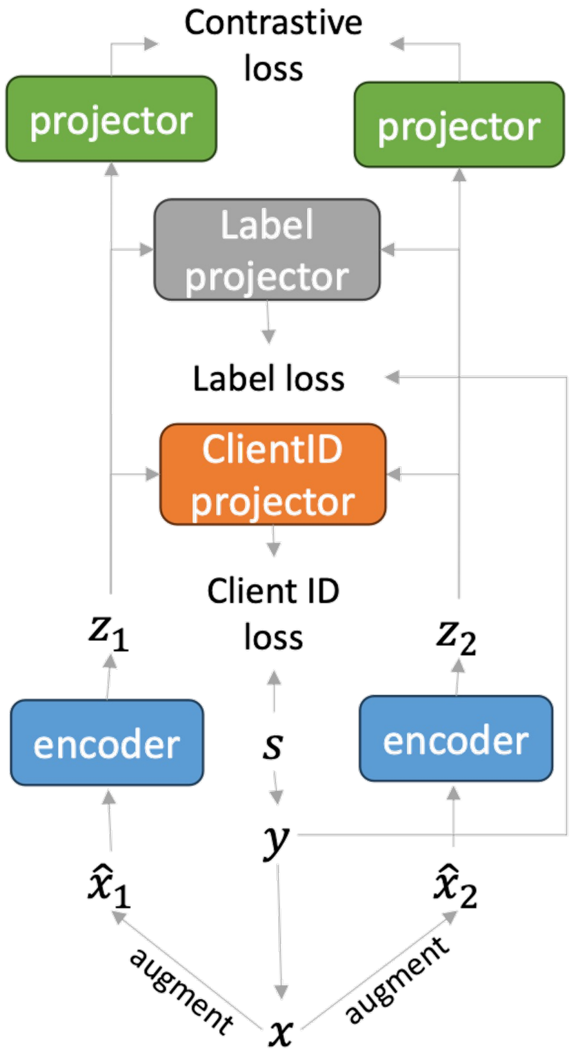
The SimCLR variants considered



Local SimCLR: each client applies SimCLR on their own data, thus the federation implicitly optimizes a lower bound to the local $I(z_1;\ z_2|s)$



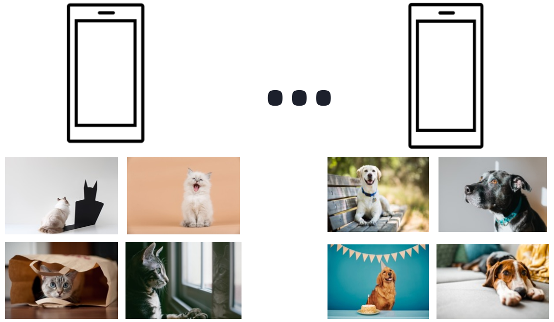
Federated SimCLR: each client also optimizes a client classifier, thus the federation implicitly optimizes a lower bound to the global $I(z_1;\ z_2)$



Supervised federated SimCLR: with an additonal label classifier and label dependent contrastive learning, we obtain a label-informed variant that also optimizes a lower bound to the global $I(z_1;\ z_2)$

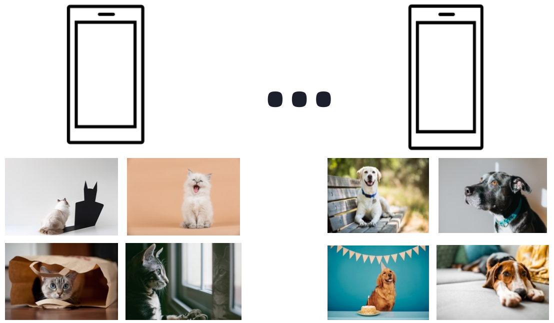
Federated Learning: the non-i.i.d. generative process

Federated Learning: the non-i.i.d. generative process

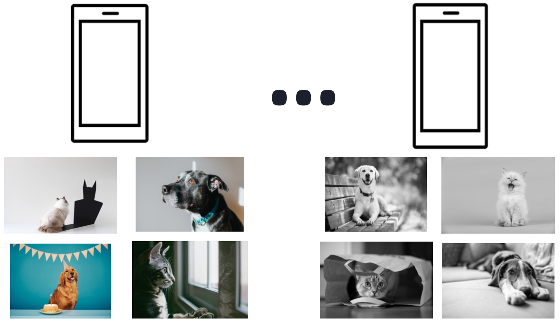


- Label-skew: the most common non-iid setting assumed in FL
  - Each client has a different class marginal, $p(y|s)$
  - The conditional feature distribution, $p(x|y)$, is the same for all clients

Federated Learning: the non-i.i.d. generative process

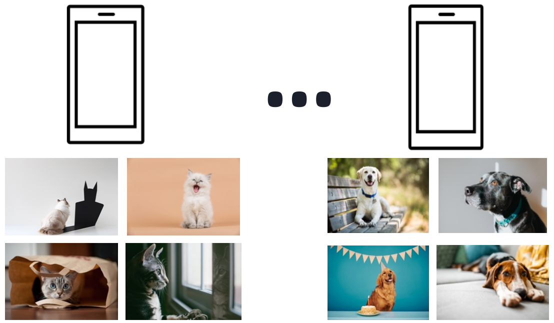


- Label-skew: the most common non-iid setting assumed in FL
  - Each client has a different class marginal, $p(y|s)$
  - The conditional feature distribution, $p(x|y)$, is the same for all clients

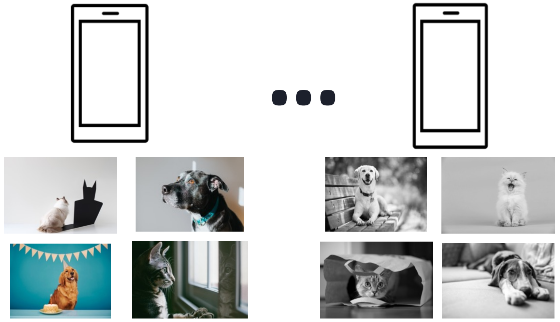


- Covariate shift: feature noise, independent of the label
  - Each client has the same local distribution over classes, $p(y)$
  - The conditional distribution over features, $p(x|y, s)$, varies due to, e.g., sensor noise

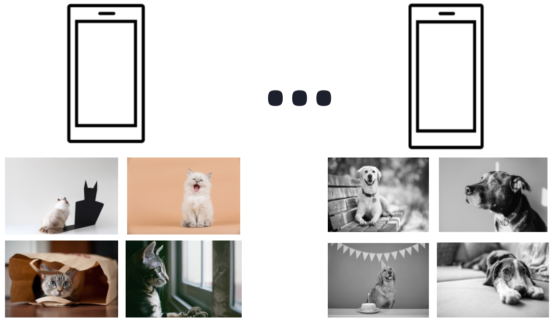
Federated Learning: the non-i.i.d. generative process



- Label-skew: the most common non-iid setting assumed in FL
  - Each client has a different class marginal, $p(y|s)$
  - The conditional feature distribution, $p(x|y)$, is the same for all clients

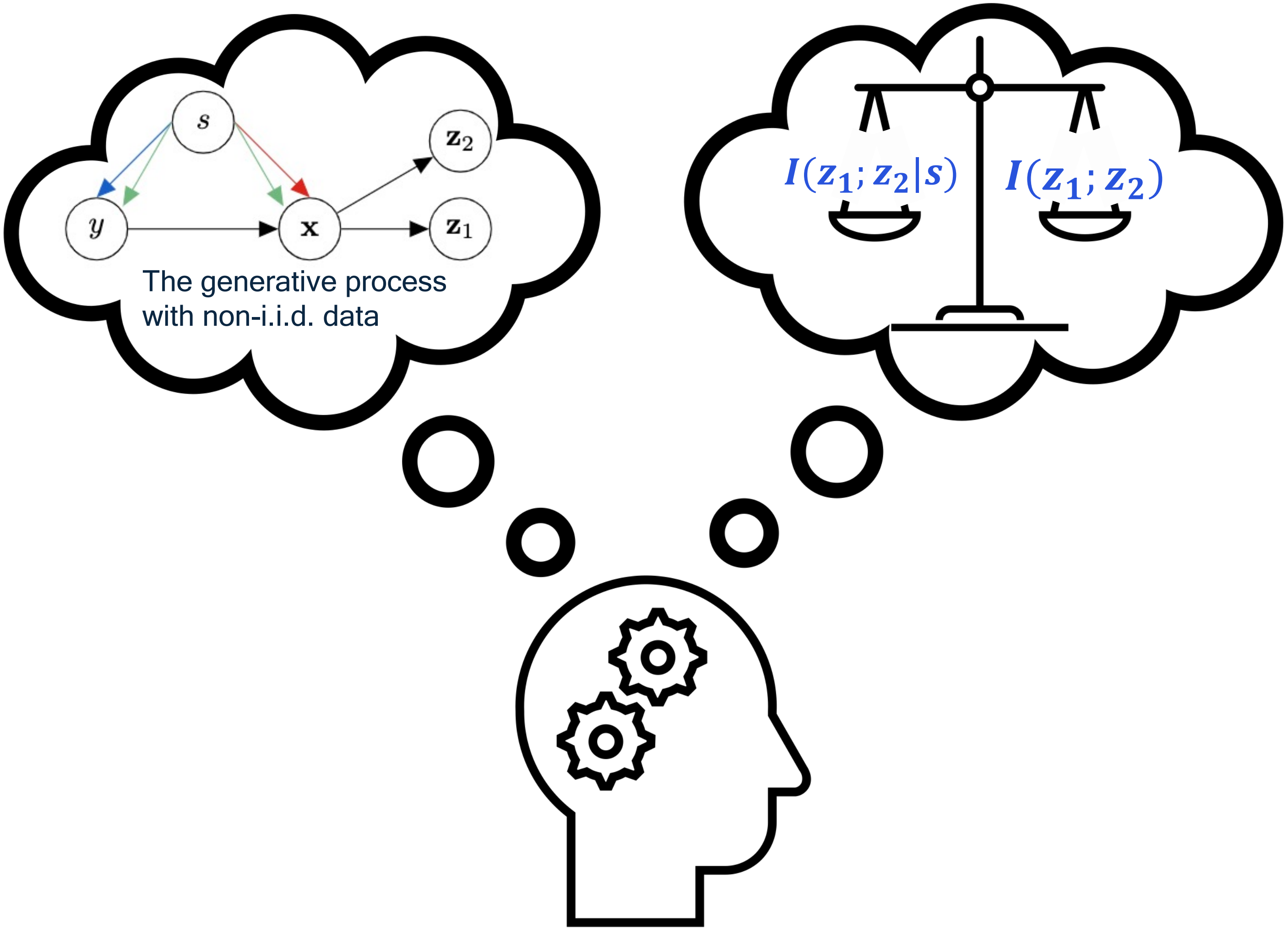


- Covariate shift: feature noise, independent of the label
  - Each client has the same local distribution over classes, $p(y)$
  - The conditional distribution over features, $p(x|y,s)$, varies due to, e.g., sensor noise



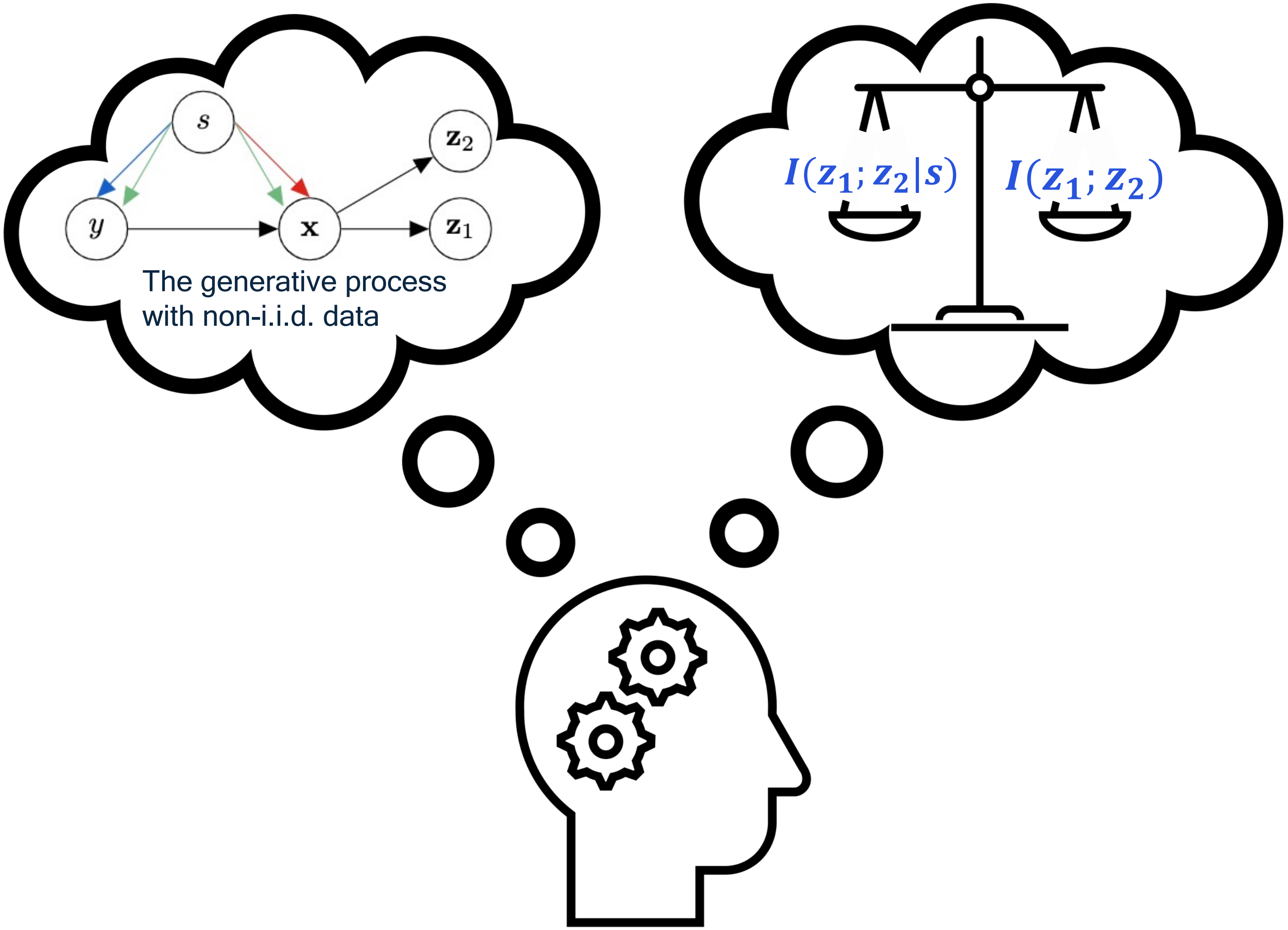
- Joint shift: a mixture of the two
  - Both the label marginal and feature distribution vary per client, $p(y|s)p(x|y,s)$

The effects of non-i.i.d.-ness on SimCLR



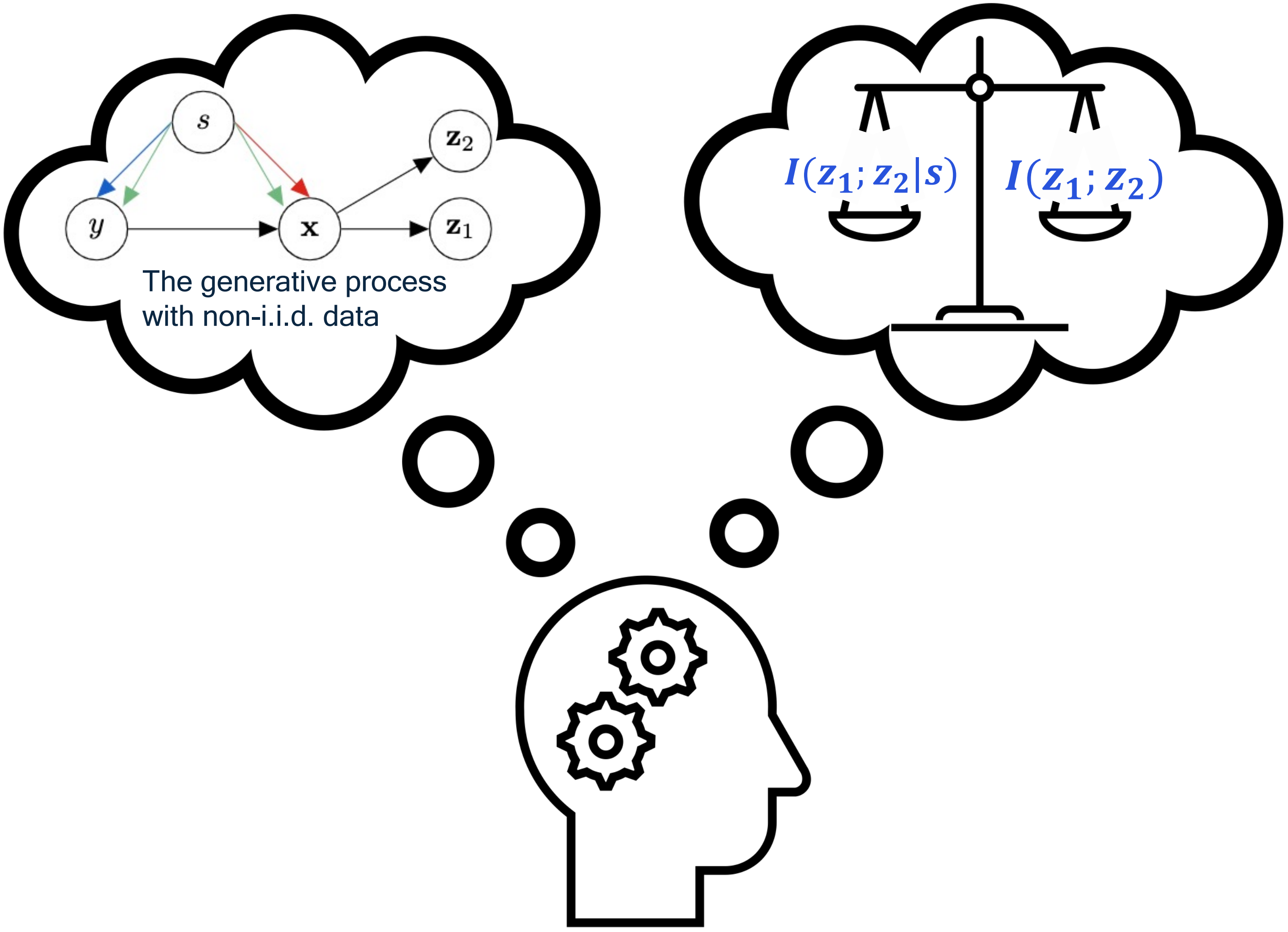
- What are the effects of data non-i.i.d.-ness?
  - Use local $I(z_1; z_2|s)$ or global $I(z_1; z_2)$?
  - The client classification task is what separates them

The effects of non-i.i.d.-ness on SimCLR



- What are the effects of data non-i.i.d.-ness?
  - Use local $I(z_1; z_2|s)$ or global $I(z_1; z_2)$?
  - The client classification task is what separates them
- When we have label skew, the client classification task is beneficial
  - We prove that it maximizes a lower bound to the mutual information between the representations and the unknown ground truth label

The effects of non-i.i.d.-ness on SimCLR



- What are the effects of data non-i.i.d.-ness?
  - Use local $I(z_1; z_2|s)$ or global $I(z_1; z_2)$?
  - The client classification task is what separates them
- When we have label skew, the client classification task is beneficial
  - We prove that it maximizes a lower bound to the mutual information between the representations and the unknown ground truth label
- When we have covariate shift, it can be detrimental
  - It encourages storing in the representations irrelevant, for the downstream task, information

Experimental results

Experimental results

Results in the unsupervised case

	CIFAR 10			CIFAR 100		
Method	Label skew	Covariate shift	Joint shift	Label skew	Covariate shift	Joint shift
Local SimCLR	$79.4_{\pm 0.2}$	$\mathbf{74.3_{\pm 0.3}}$	$71.0_{\pm 0.4}$	$42.2_{\pm 0.2}$	$\mathbf{41.2_{\pm 0.2}}$	$38.1_{\pm 0.3}$
Federated SimCLR	$\mathbf{85.0_{\pm 0.2}}$	$73.8_{\pm 0.2}$	$\mathbf{74.8_{\pm 0.5}}$	$\mathbf{48.5_{\pm 0.1}}$	$39.5_{\pm 0.2}$	$\mathbf{43.1_{\pm 0.2}}$
Spectral CL	$76.5_{\pm 1.1}$	$\mathbf{73.5_{\pm 0.4}}$	$68.2_{\pm 0.6}$	$33.3_{\pm 6.0}$	$\mathbf{33.6_{\pm 2.3}}$	$\mathbf{29.6_{\pm 6.2}}$
Spectral CL + UV	$\mathbf{87.8_{\pm 0.3}}$	$71.7_{\pm 0.5}$	$\mathbf{76.6_{\pm 0.6}}$	$\mathbf{41.0_{\pm 6.4}}$	$29.3_{\pm 4.8}$	$21.5_{\pm 6.2}$
SimSiam	$\mathbf{40.0_{\pm 0.5}}$	$\mathbf{39.9_{\pm 0.3}}$	$\mathbf{39.6_{\pm 0.3}}$	$16.9_{\pm 0.3}$	$16.6_{\pm 0.4}$	$16.9_{\pm 0.4}$
SimSiam + UV	$35.4_{\pm 0.4}$	$35.4_{\pm 0.2}$	$34.5_{\pm 0.3}$	$16.5_{\pm 0.2}$	$16.5_{\pm 0.3}$	$16.3_{\pm 0.5}$
Supervised	$89.6_{\pm 0.1}$	$78.3_{\pm 0.4}$	$76.3_{\pm 1.1}$	$59.2_{\pm 0.2}$	$47.9_{\pm 0.2}$	$43.9_{\pm 0.3}$

Experimental results

Results in the unsupervised case

	CIFAR 10			CIFAR 100		
Method	Label skew	Covariate shift	Joint shift	Label skew	Covariate shift	Joint shift
Local SimCLR	$79.4_{\pm 0.2}$	$\mathbf{74.3_{\pm 0.3}}$	$71.0_{\pm 0.4}$	$42.2_{\pm 0.2}$	$\mathbf{41.2_{\pm 0.2}}$	$38.1_{\pm 0.3}$
Federated SimCLR	$\mathbf{85.0_{\pm 0.2}}$	$73.8_{\pm 0.2}$	$\mathbf{74.8_{\pm 0.5}}$	$\mathbf{48.5_{\pm 0.1}}$	$39.5_{\pm 0.2}$	$\mathbf{43.1_{\pm 0.2}}$
Spectral CL	$76.5_{\pm 1.1}$	$\mathbf{73.5_{\pm 0.4}}$	$68.2_{\pm 0.6}$	$33.3_{\pm 6.0}$	$\mathbf{33.6_{\pm 2.3}}$	$\mathbf{29.6_{\pm 6.2}}$
Spectral CL + UV	$\mathbf{87.8_{\pm 0.3}}$	$71.7_{\pm 0.5}$	$\mathbf{76.6_{\pm 0.6}}$	$\mathbf{41.0_{\pm 6.4}}$	$29.3_{\pm 4.8}$	$21.5_{\pm 6.2}$
SimSiam	$\mathbf{40.0_{\pm 0.5}}$	$\mathbf{39.9_{\pm 0.3}}$	$\mathbf{39.6_{\pm 0.3}}$	$16.9_{\pm 0.3}$	$16.6_{\pm 0.4}$	$16.9_{\pm 0.4}$
SimSiam + UV	$35.4_{\pm 0.4}$	$35.4_{\pm 0.2}$	$34.5_{\pm 0.3}$	$16.5_{\pm 0.2}$	$16.5_{\pm 0.3}$	$16.3_{\pm 0.5}$
Supervised	$89.6_{\pm 0.1}$	$78.3_{\pm 0.4}$	$76.3_{\pm 1.1}$	$59.2_{\pm 0.2}$	$47.9_{\pm 0.2}$	$43.9_{\pm 0.3}$

Results in the semi-supervised case

	CIFAR 10			CIFAR 100		
Method	Label skew	Covariate shift	Joint shift	Label Skew	Covariate shift	Joint shift
Local SimCLR	$74.5_{\pm 0.3}$	$\mathbf{49.1_{\pm 1.3}}$	$45.8_{\pm 1.4}$	$30.3_{\pm 0.2}$	$15.1_{\pm 0.4}$	$13.1_{\pm 0.3}$
Federated SimCLR	$\mathbf{78.0_{\pm 0.2}}$	$\mathbf{50.3_{\pm 1.1}}$	$\mathbf{49.9_{\pm 1.4}}$	$\mathbf{34.5_{\pm 0.3}}$	$14.8_{\pm 0.3}$	$\mathbf{14.6_{\pm 0.3}}$
Spectral CL	$74.2_{\pm 0.3}$	$48.0_{\pm 0.7}$	$45.4_{\pm 1.5}$	$30.1_{\pm 0.2}$	$14.1_{\pm 0.4}$	$12.3_{\pm 0.3}$
Spectral CL + UV	$\mathbf{79.6_{\pm 0.3}}$	$\mathbf{49.7_{\pm 1.0}}$	$\mathbf{49.8_{\pm 1.1}}$	$\mathbf{34.0_{\pm 0.2}}$	$13.7_{\pm 0.3}$	$\mathbf{13.6_{\pm 0.4}}$
SimSiam	$75.3_{\pm 0.4}$	$46.8_{\pm 0.7}$	$40.5_{\pm 0.9}$	$30.7_{\pm 0.2}$	$13.4_{\pm 0.3}$	$12.8_{\pm 0.3}$
SimSiam + UV	$\mathbf{80.4_{\pm 0.2}}$	$\mathbf{50.0_{\pm 1.2}}$	$\mathbf{44.3_{\pm 1.0}}$	$\mathbf{34.3_{\pm 0.1}}$	$13.6_{\pm 0.3}$	$\mathbf{14.0_{\pm 0.4}}$
Supervised	$75.1_{\pm 0.2}$	$48.1_{\pm 0.9}$	$42.7_{\pm 1.7}$	$29.6_{\pm 0.3}$	$12.6_{\pm 0.2}$	$12.2_{\pm 0.1}$

Thank you

Qualcomm

Follow us on:

For more information, visit us at:

qualcomm.com & qualcomm.com/blog