Lyon, Avril 17– 21

DLMI2023

CREATIS

UNIVERSITÉ DE LYON

ÉTS Le génie pour l'industrie

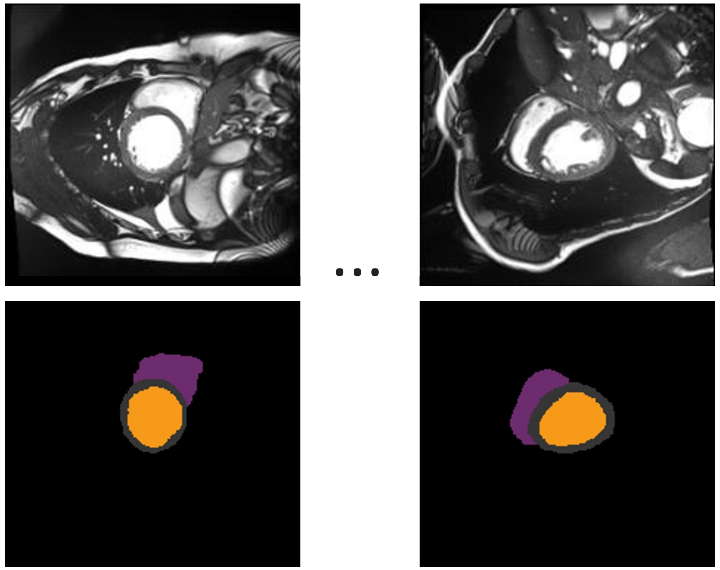
Semi-supervised Learning for Medical Image Segmentation

Christian Desrosiers

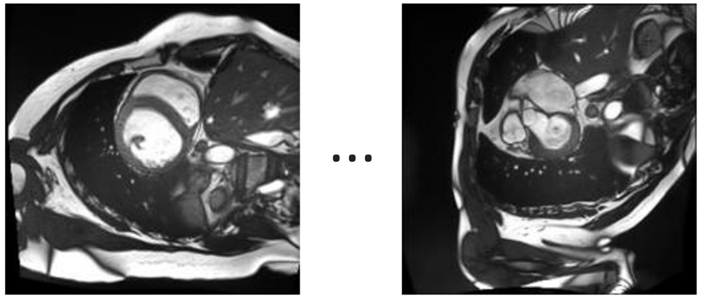
ÉTS, Montreal

Learning with unlabeled images

Labeled images (few)

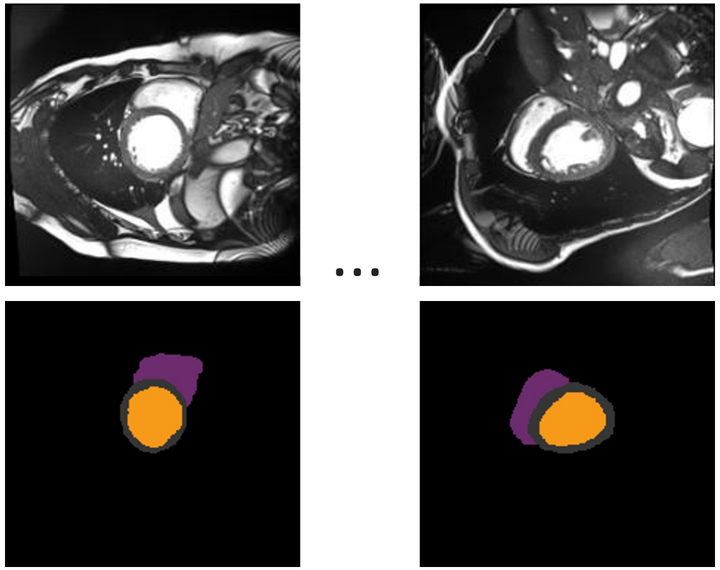


Unlabeled images (many)

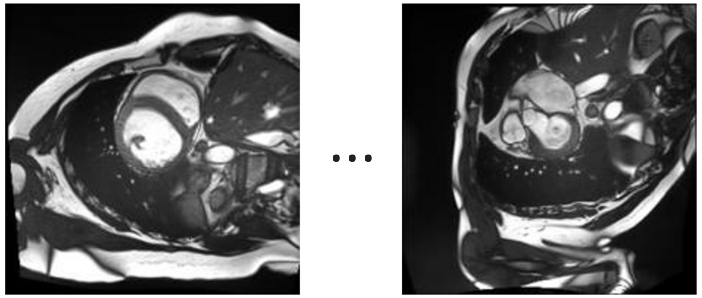


Learning with unlabeled images

Labeled images (few)



Unlabeled images (many)



How can we use this information
to learn segmentation ?

Outline

1) Adversarial learning
2) Consistency regularization
3) Unsupervised representation learning

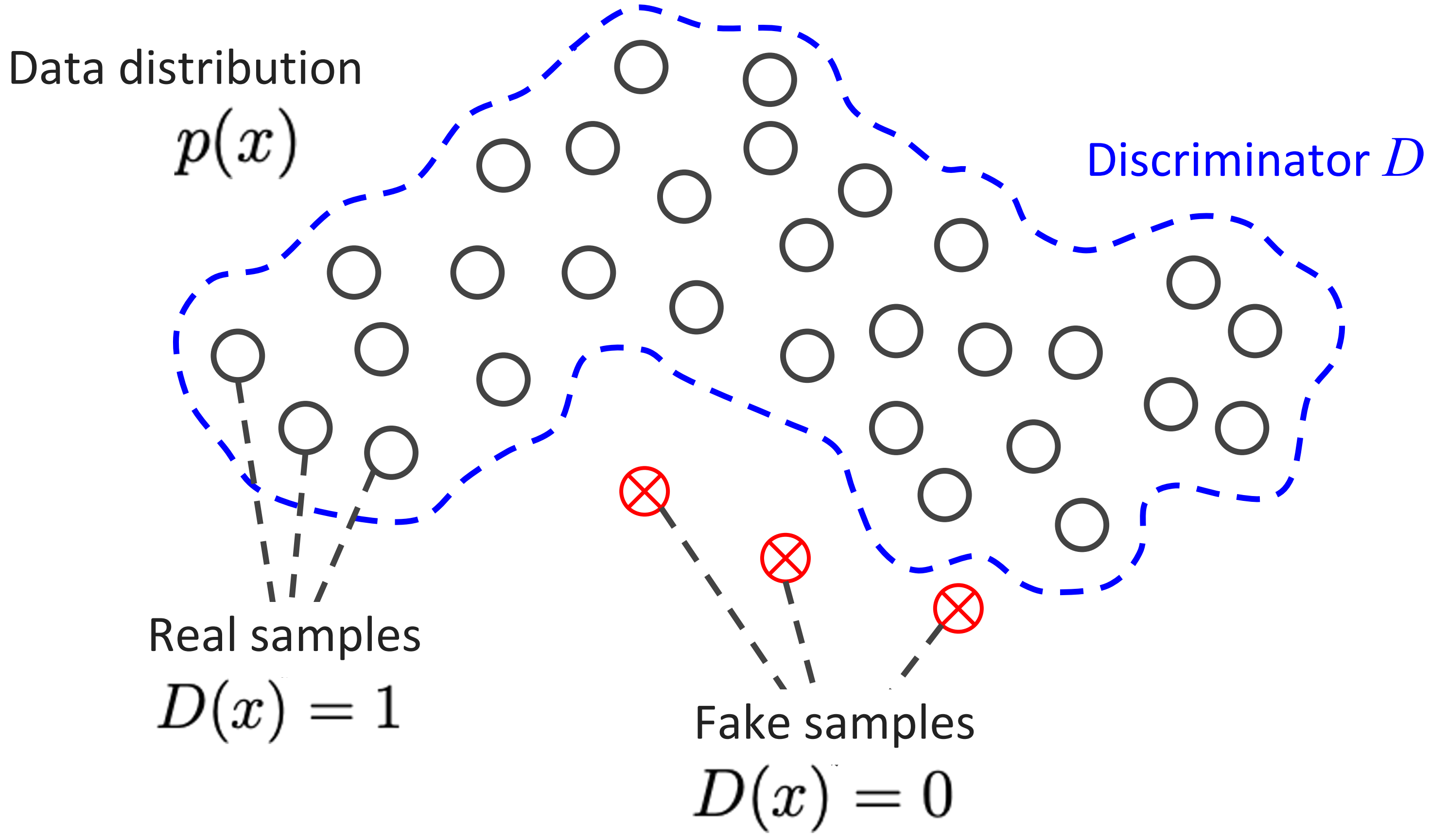
Adversarial learning

for semi-supervised segmentation

Adversarial learning

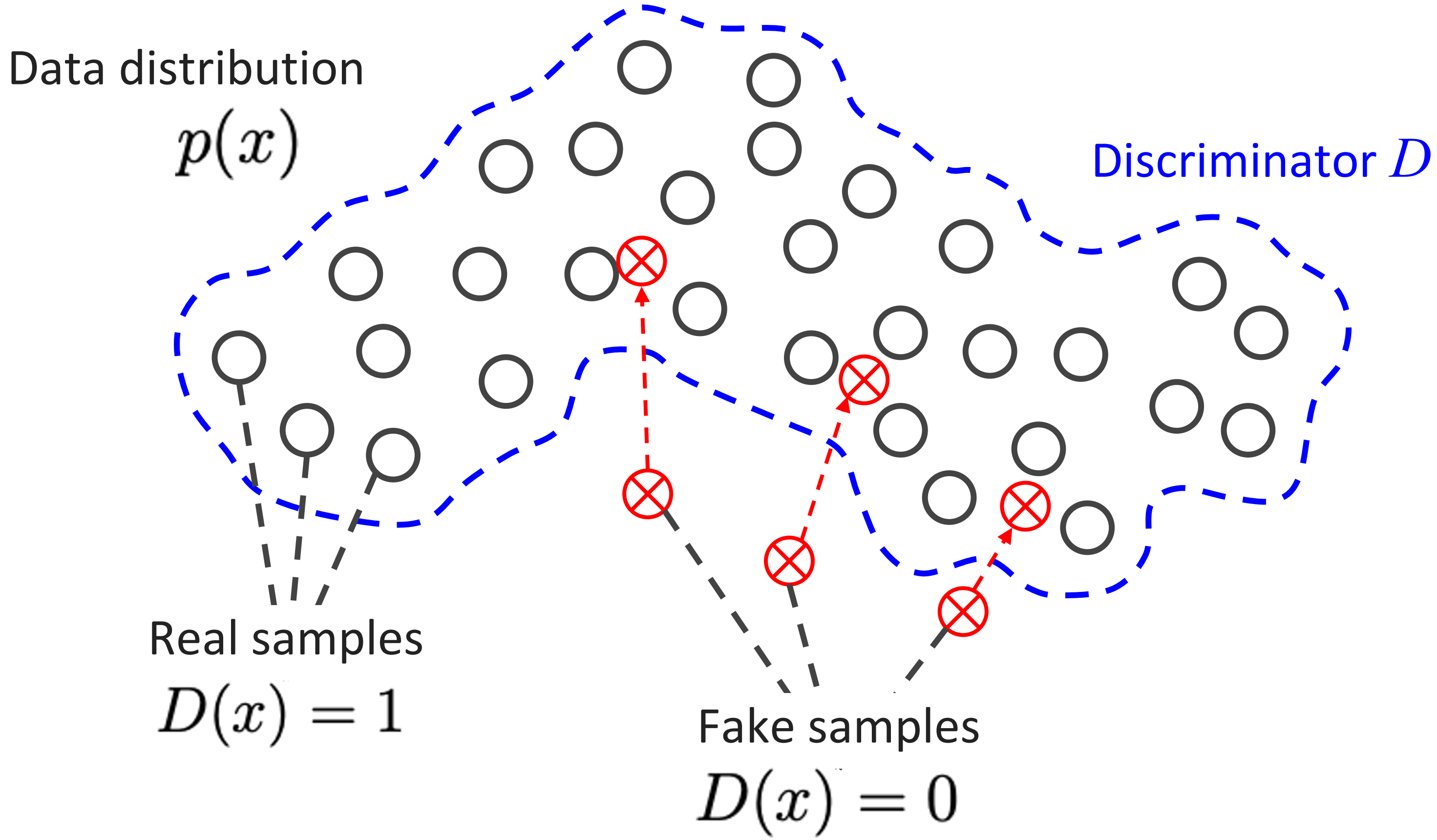
Basic idea:

Learn the data distribution using a classifier (the discriminator)



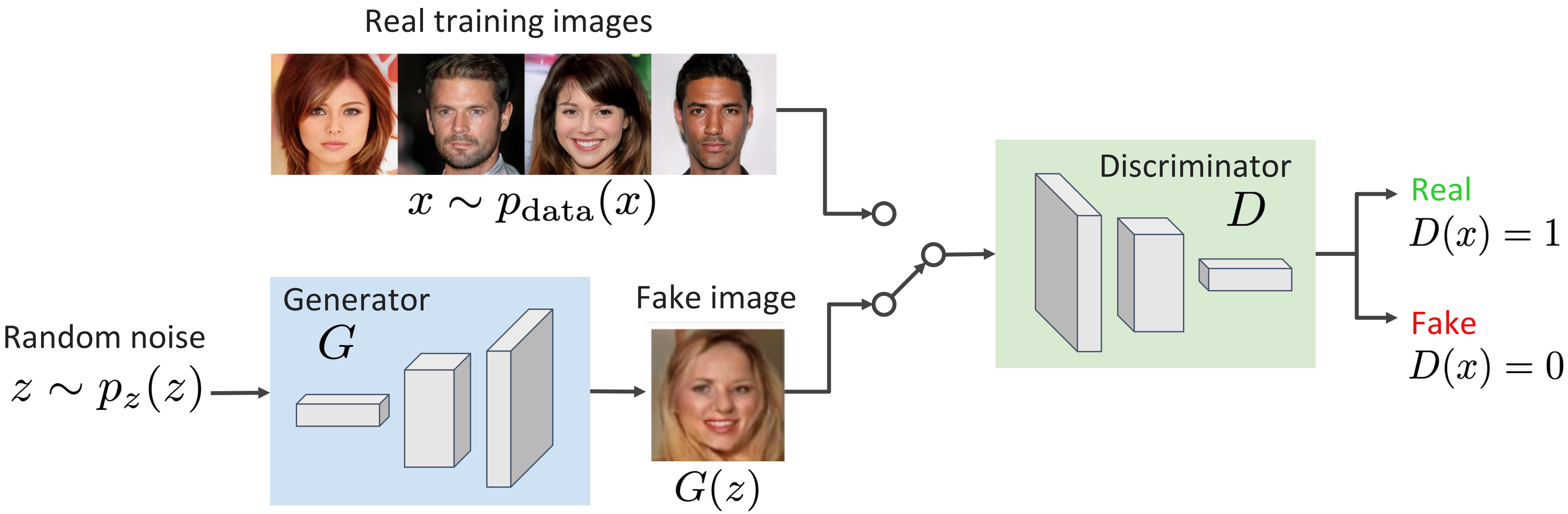
Adversarial learning

Basic idea:
Learn the data distribution using a classifier (the discriminator)

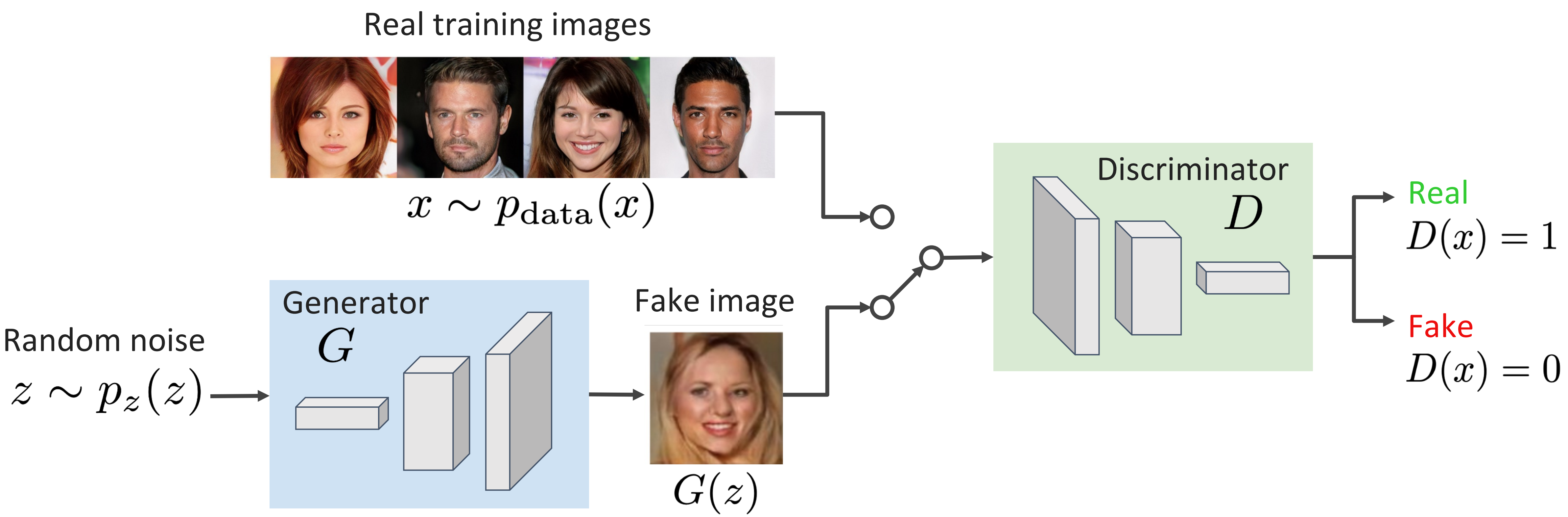


Objective: Generate samples in the distribution of real data

Generative adversarial network (GAN)

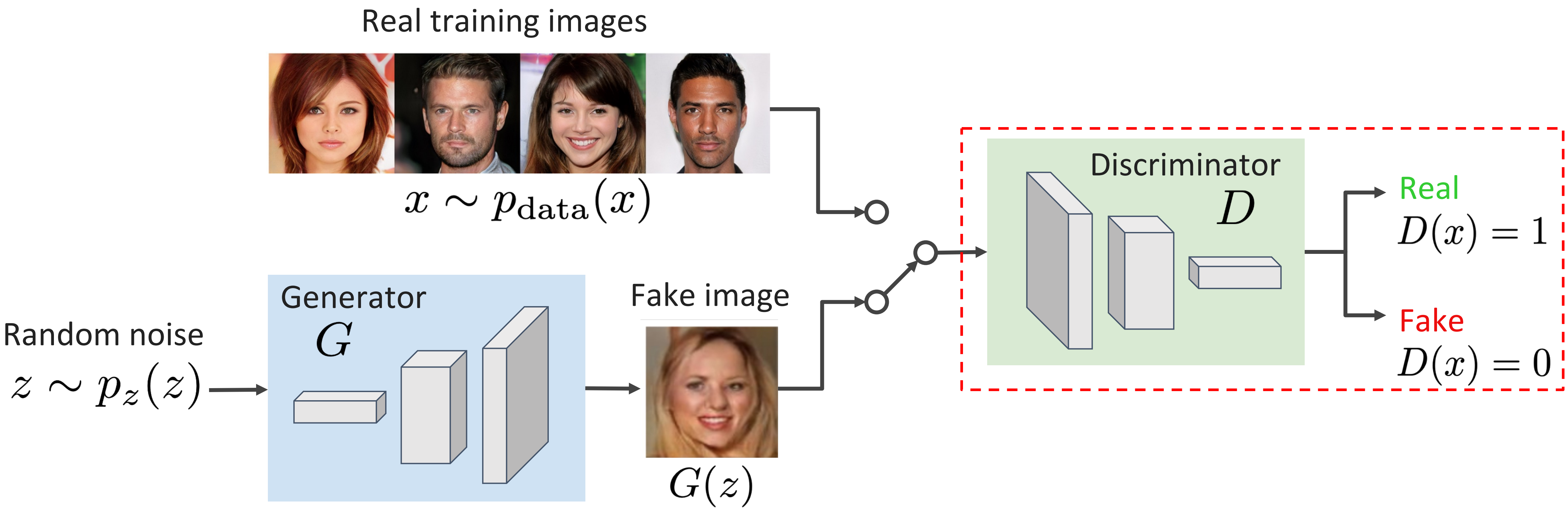


Generative adversarial network (GAN)



How to make sure that generated images look real ?

Generative adversarial network (GAN)



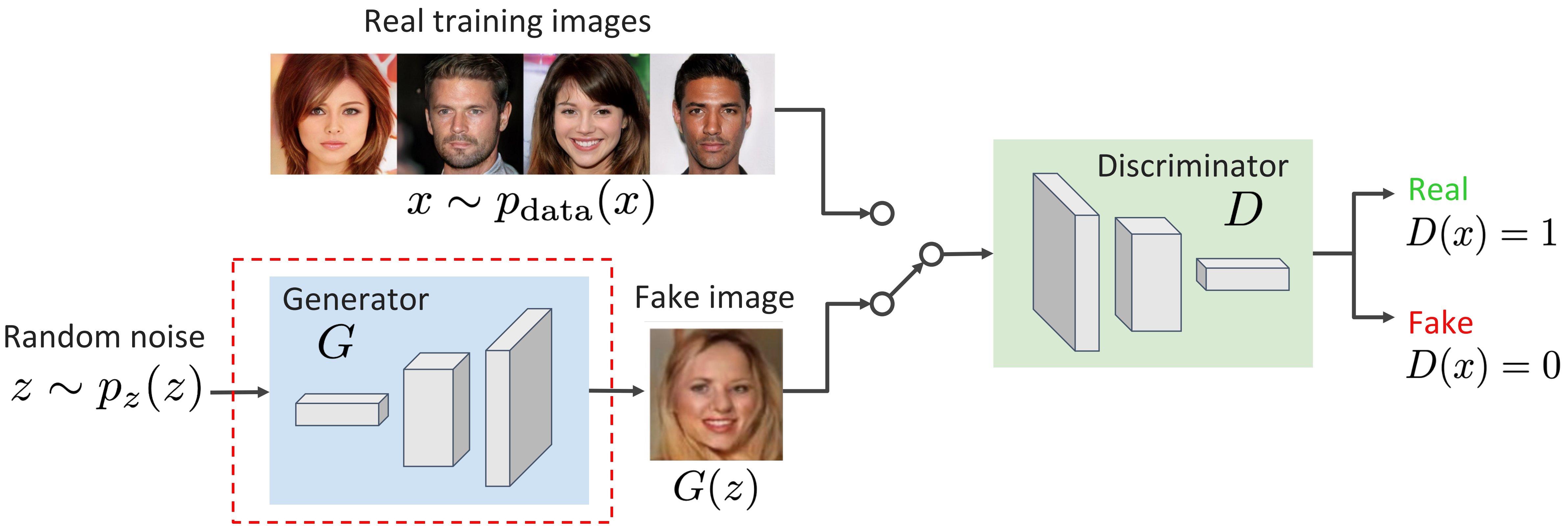
Training the discriminator (cross-entropy):

$$\max_D \; \underbrace{\mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\log D(x)\right]}_{\text{Output '1' for real images}} + \underbrace{\mathbb{E}_{z\sim p_z(z)}\left[\log(1 - D(G(z)))\right]}_{\text{Output '0' for generated images}}$$

Output ‘1’ for real images

Output ‘0’ for generated images

Generative adversarial network (GAN)

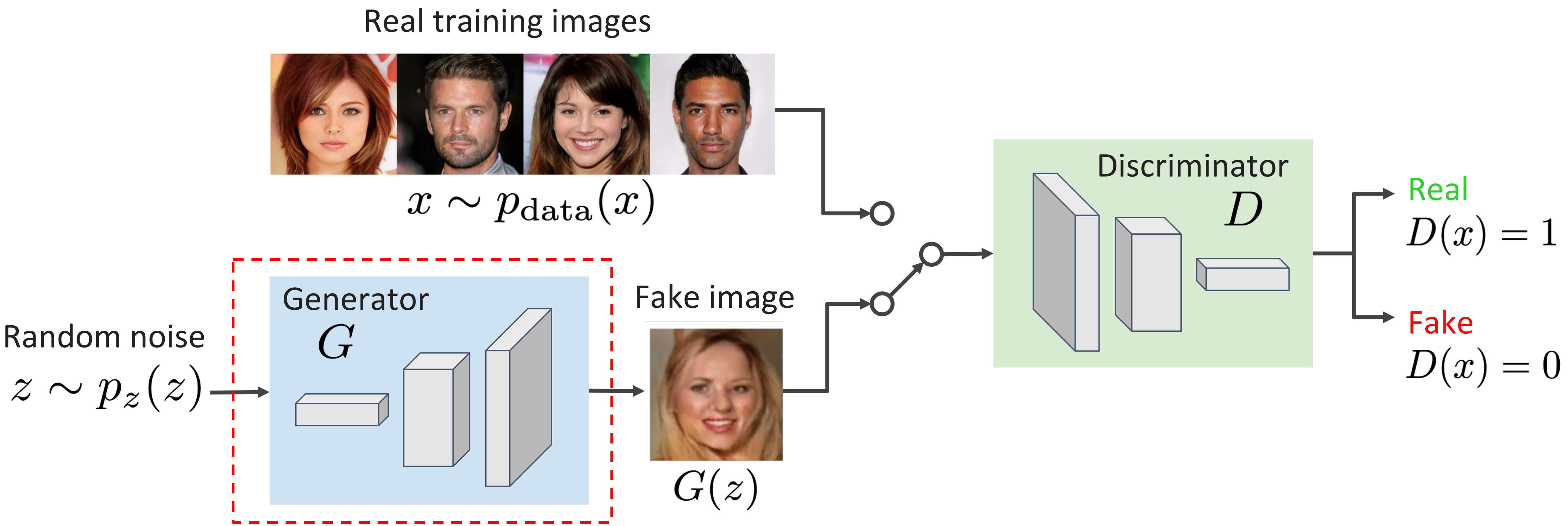


Training the generator:

$$\min_G \; \mathbb{E}_{z \sim p_z(z)} \left[\log(1 - D(G(z))) \right]$$

Fool the discriminator into predicting '1' for fake images

Generative adversarial network (GAN)



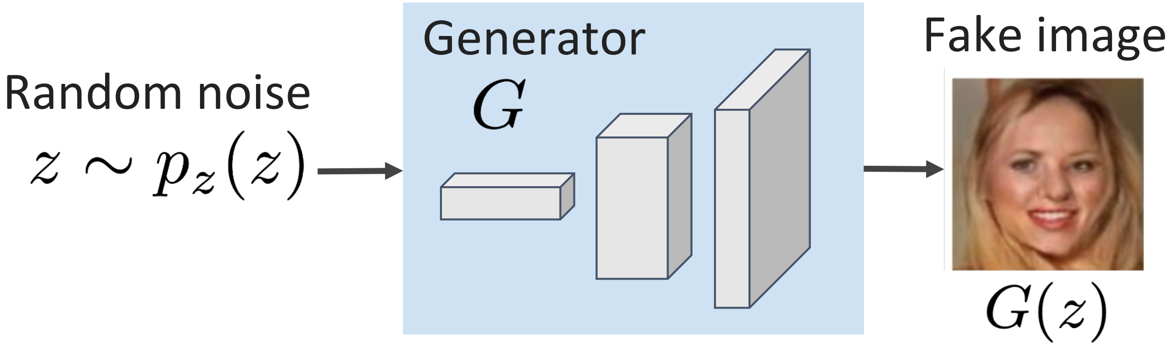
Training the whole architecture:

$$\min_G \max_D \; \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log D(x)\right] + \mathbb{E}_{z \sim p_z(z)}\left[\log(1 - D(G(z)))\right]$$

Corresponds to a minimax problem (*more on this later...*)

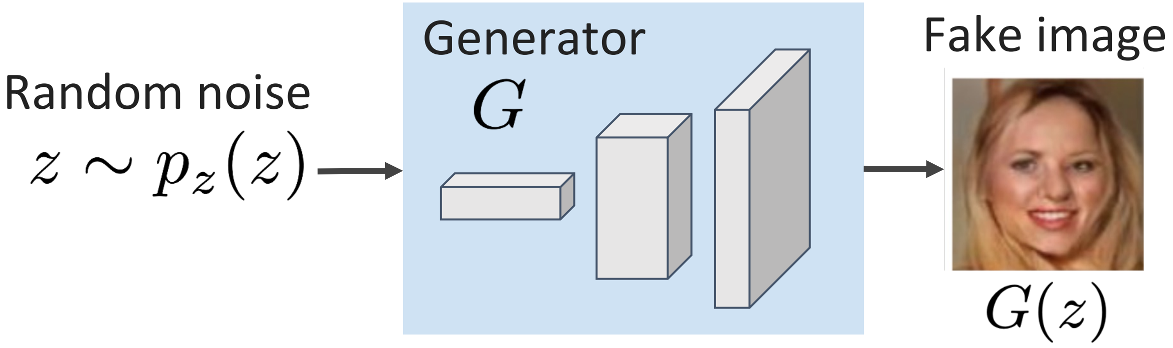
GANs for segmentation

GAN for image generation:

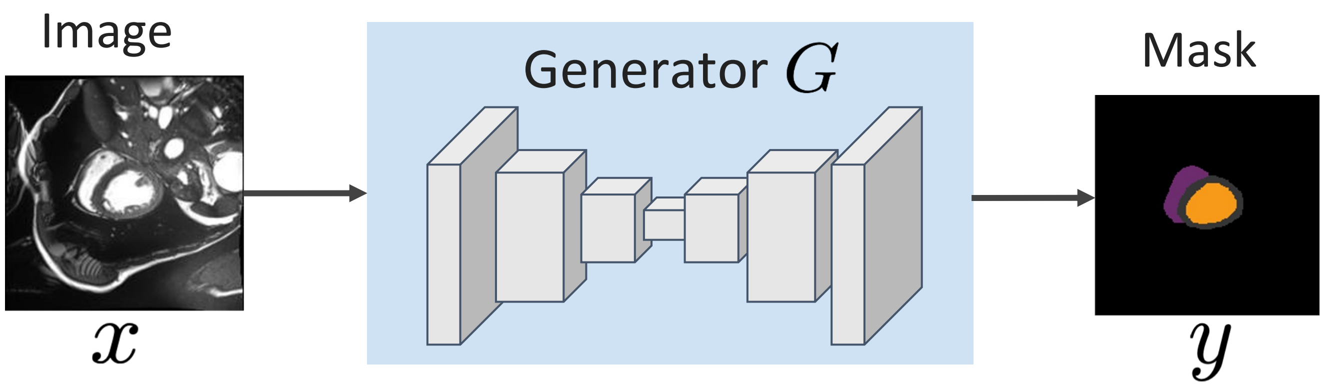


GANs for segmentation

GAN for image generation:

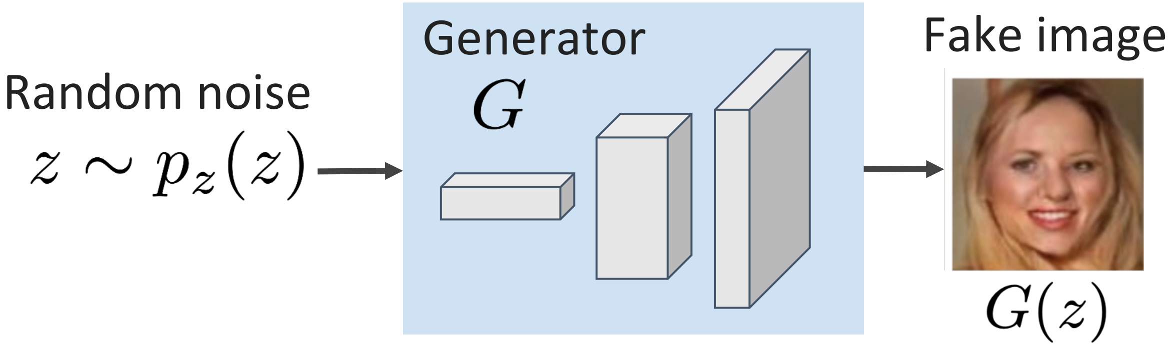


GAN for image segmentation:

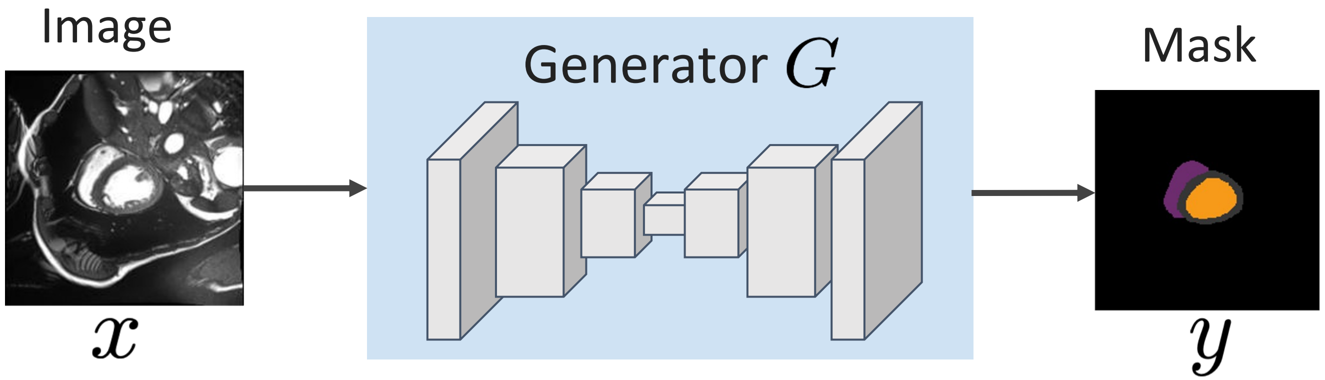


GANs for segmentation

GAN for image generation:



GAN for image segmentation:



We are now modeling the distribution of *segmentation masks*

The generator is a segmentation network (encoder-decoder)

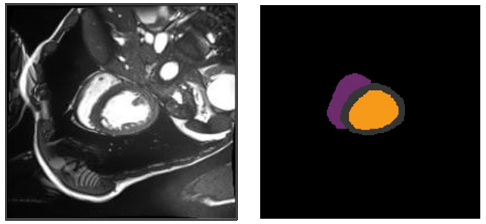
Adversarial semi-supervised segmentation

Basic idea: Learn to generate segmentation masks which can't be differentiated from ground-truth (GT)

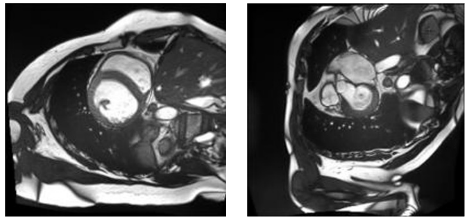
Labeled images $\mathcal{X}_l, \mathcal{Y}_l$

Unlabeled images $\mathcal{X}_u$

Data =



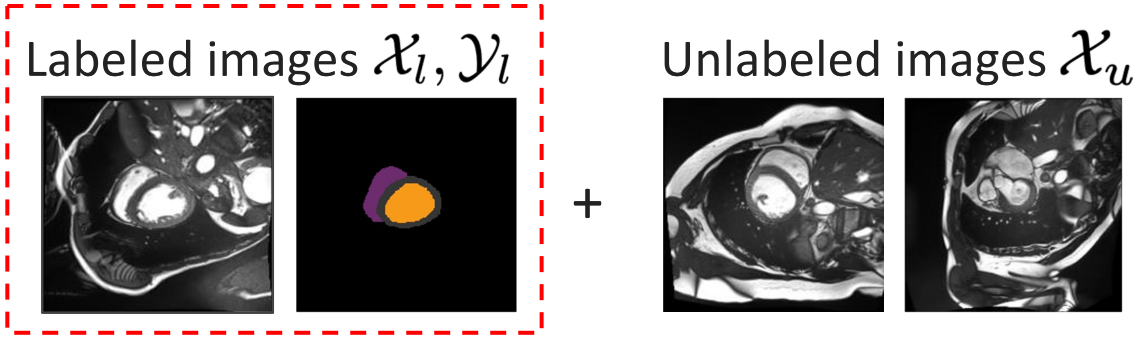
+

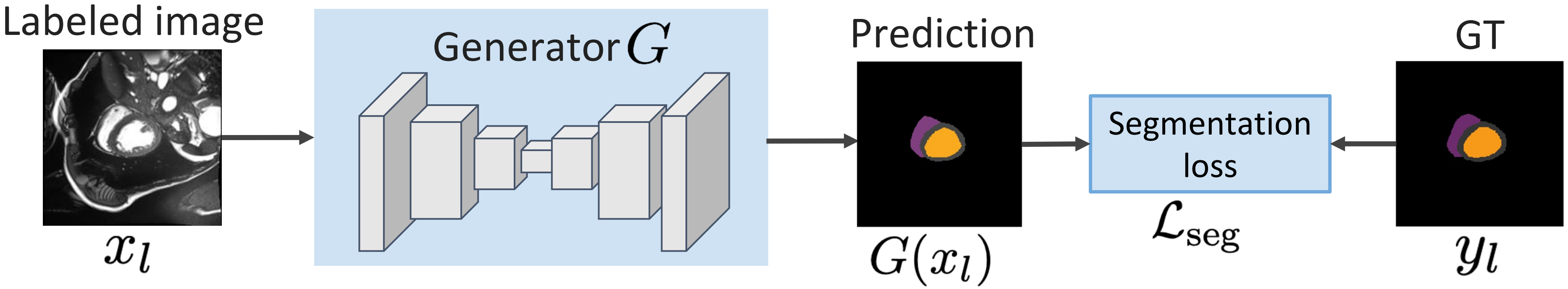


Adversarial semi-supervised segmentation

Basic idea: Learn to generate segmentation masks which can't be differentiated from ground-truth (GT)

Data = Labeled images $\mathcal{X}_l, \mathcal{Y}_l$ + Unlabeled images $\mathcal{X}_u$

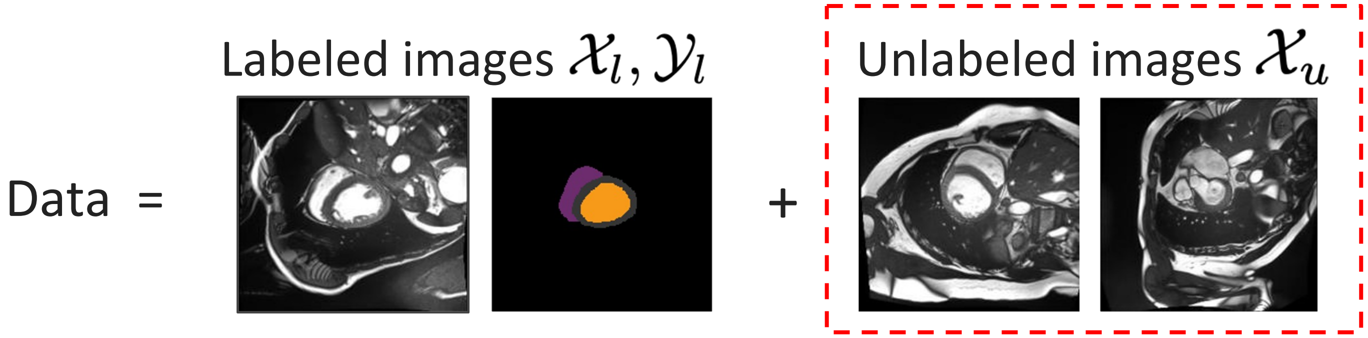


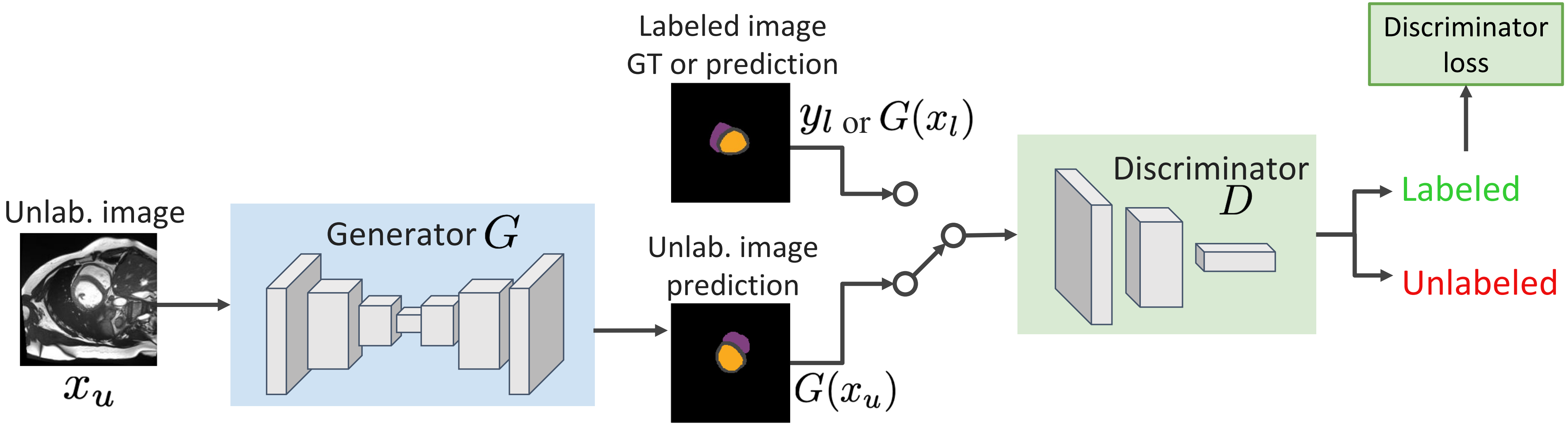


$$\mathcal{L}_{\text{sup}}(G) \;=\; \mathbb{E}_{(x_l, y_l) \sim \mathcal{X}_l, \mathcal{Y}_l}\left[\mathcal{L}_{\text{seg}}(G(x_l), y_l)\right]$$

Adversarial semi-supervised segmentation

Basic idea: Learn to generate segmentation masks which can't be differentiated from ground-truth (GT)



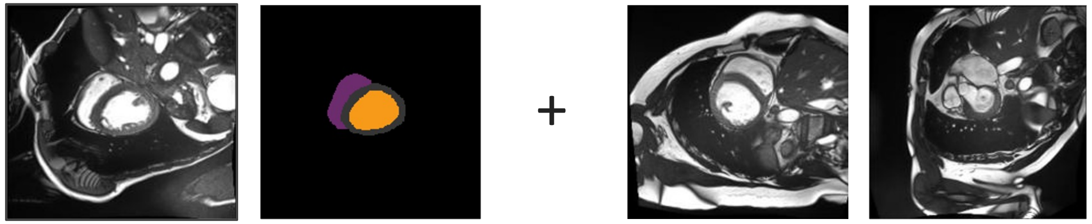


$$\mathcal{L}_{\text{adv}}(G, D) \;=\; \mathbb{E}_{x_u \sim \mathcal{X}_u}\big[\mathcal{L}_{\text{dis}}\big(D(G(x_u)), 0\big)\big] \;+\; \mathbb{E}_{x_l \sim \mathcal{X}_l}\big[\mathcal{L}_{\text{dis}}\big(D(G(x_l)), 1\big)\big]$$

Adversarial semi-supervised segmentation

Basic idea: Learn to generate segmentation masks which can't be differentiated from ground-truth (GT)

Labeled images $\mathcal{X}_l, \mathcal{Y}_l$ Unlabeled images $\mathcal{X}_u$

Data = 

Both labeled and unlabeled:

Controls the trade-off between the two losses

$$\min_G \max_D \; \mathcal{L}(G,D) \;=\; \frac{1}{|\mathcal{X}_l|}\sum_{l=1}^{|\mathcal{X}_l|}\mathcal{L}_{\text{seg}}(G(x_l), y_l) \;-\; \frac{\lambda}{|\mathcal{X}_l|+|\mathcal{X}_u|}\left(\sum_{l=1}^{|\mathcal{X}_l|}\mathcal{L}_{\text{dis}}\big(D(G(x_l)),1\big) \;+\; \sum_{u=1}^{|\mathcal{X}_u|}\mathcal{L}_{\text{dis}}\big(D(G(x_u)),0\big)\right)$$

Supervised loss

Adversarial loss

Adversarial semi-supervised segmentation

Adversarial network for semi-supervised segmentation of histological images

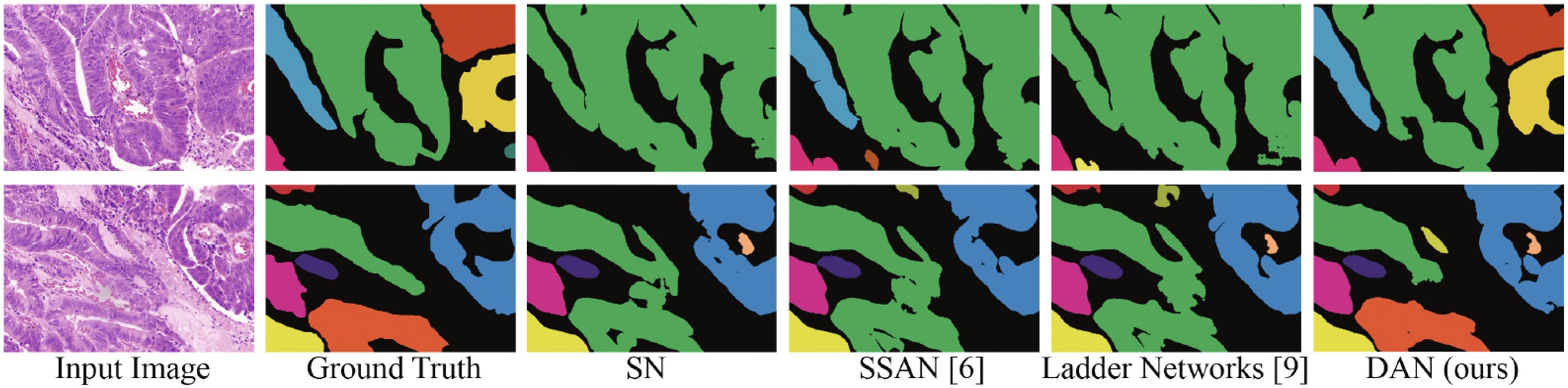
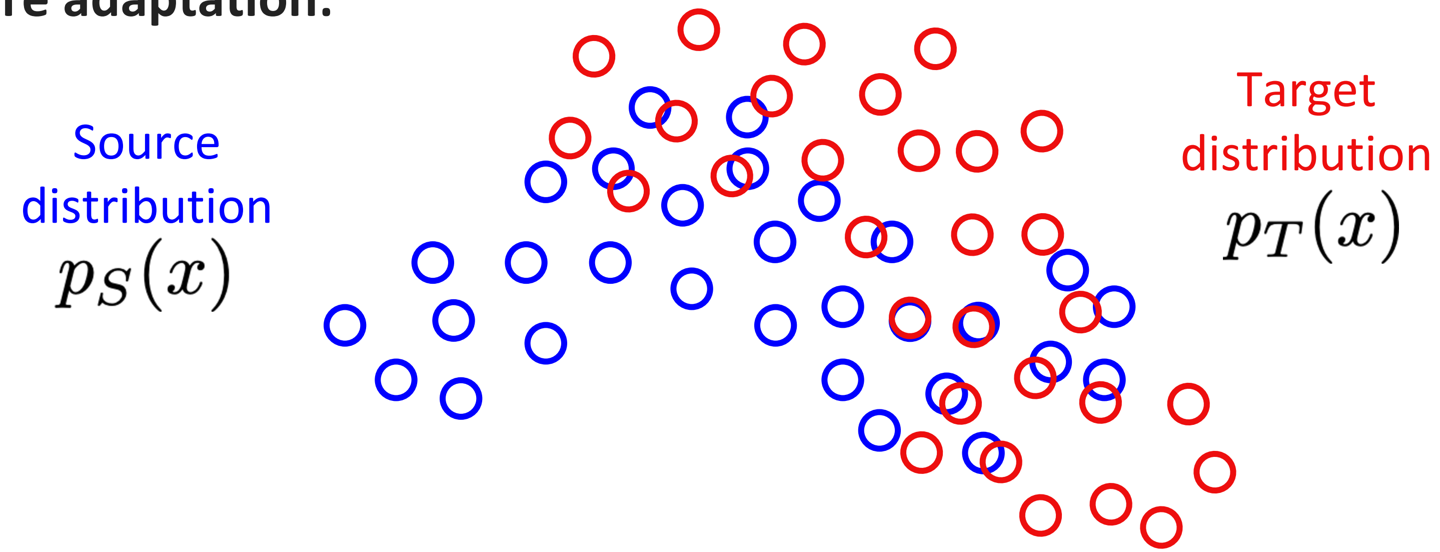


Image from Zhang, Y., et al. "Deep adversarial networks for biomedical image segmentation utilizing unannotated images." *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*. 2017.

Domain adaptation

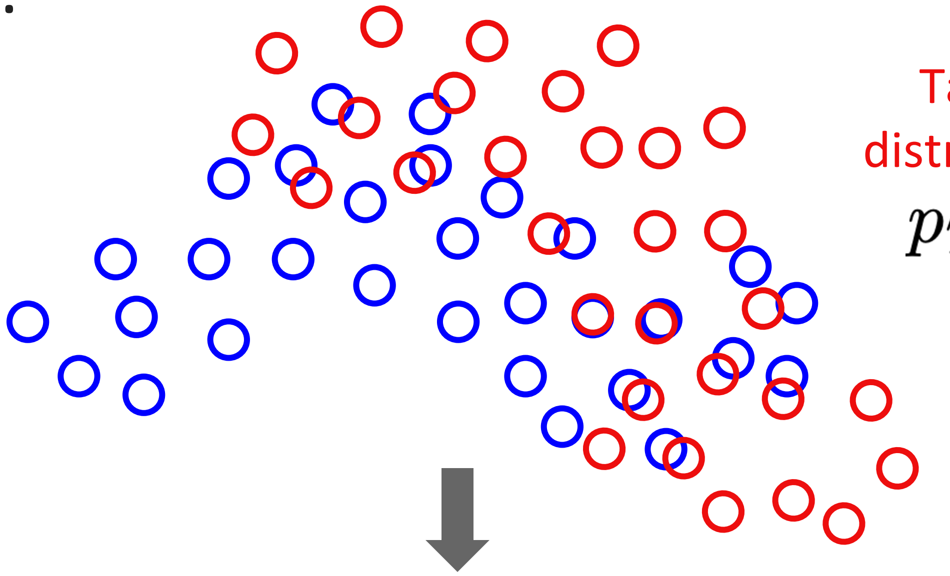
Before adaptation:



Domain adaptation

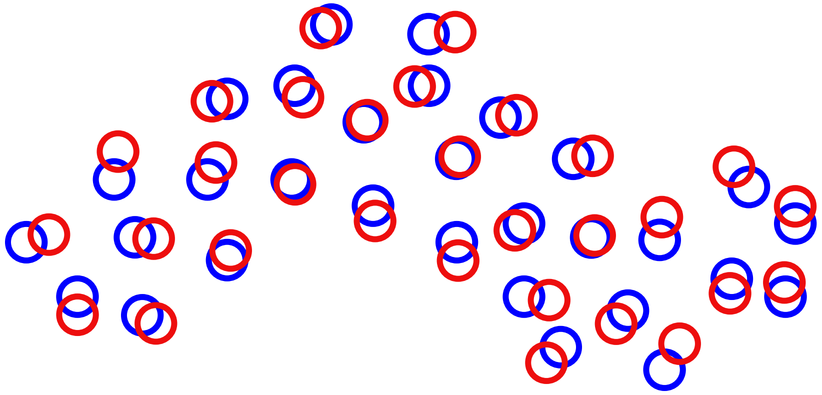
Before adaptation:

Source distribution
$p_S(x)$



Target distribution
$p_T(x)$

After adaptation:



$$p_S(x) = p_T(x)$$

Objective

Align the distributions (input, output or representation) so that a model trained on Source data also works on Target data

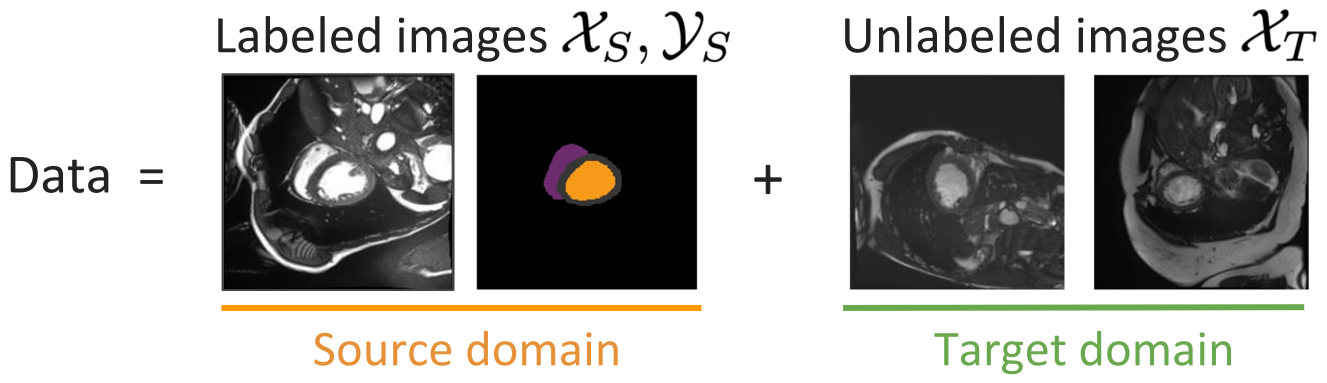
Adversarial domain adaptation

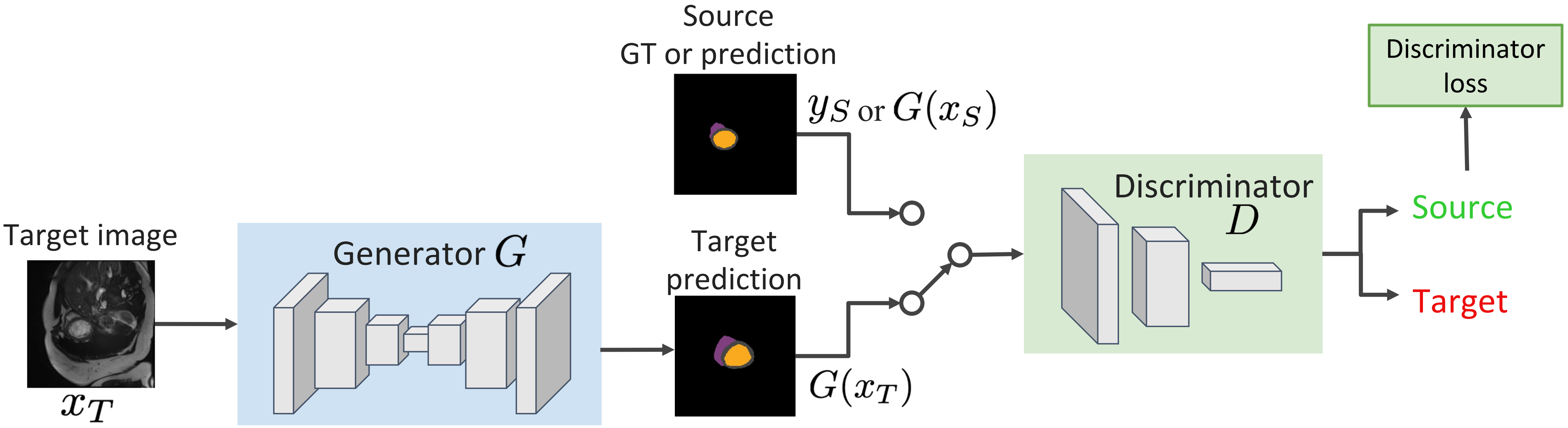
Labeled images $\mathcal{X}_S, \mathcal{Y}_S$ Unlabeled images $\mathcal{X}_T$

Data = 

Source domain + Target domain

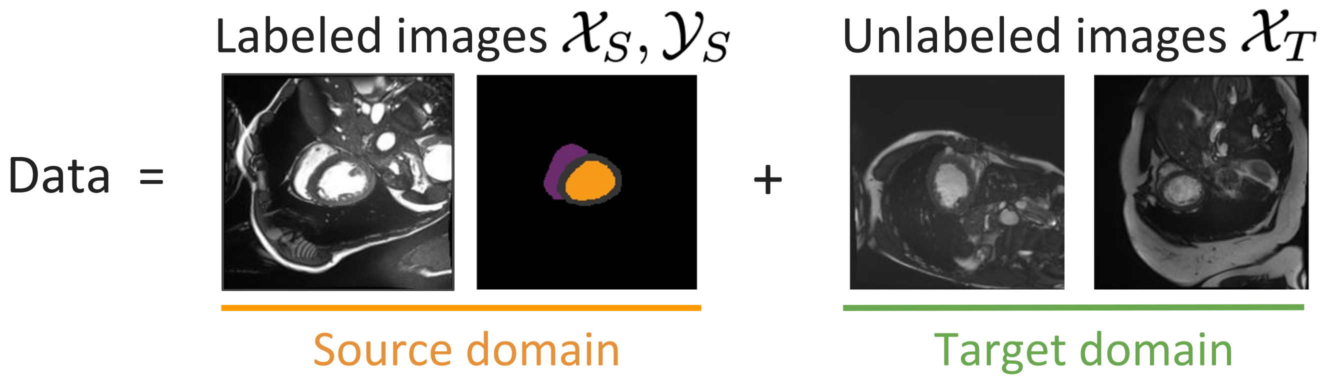
Adversarial domain adaptation

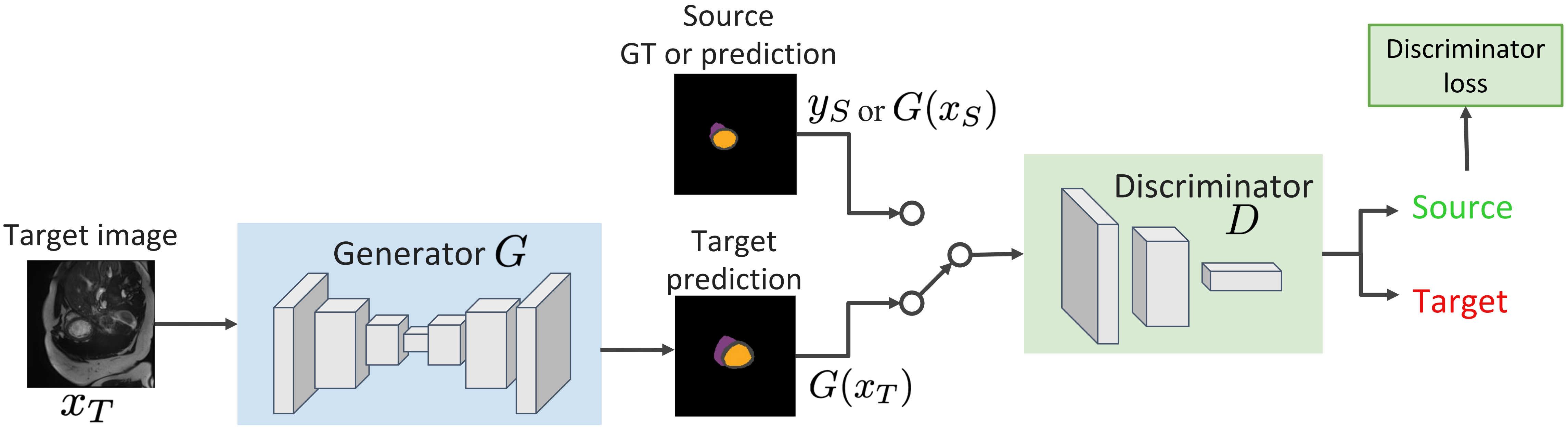




Tsai, Y.-H., et al. "Learning to adapt structured output space for semantic segmentation." IEEE Conf. on Computer Vision and Pattern Recognition. 2018.

Adversarial domain adaptation



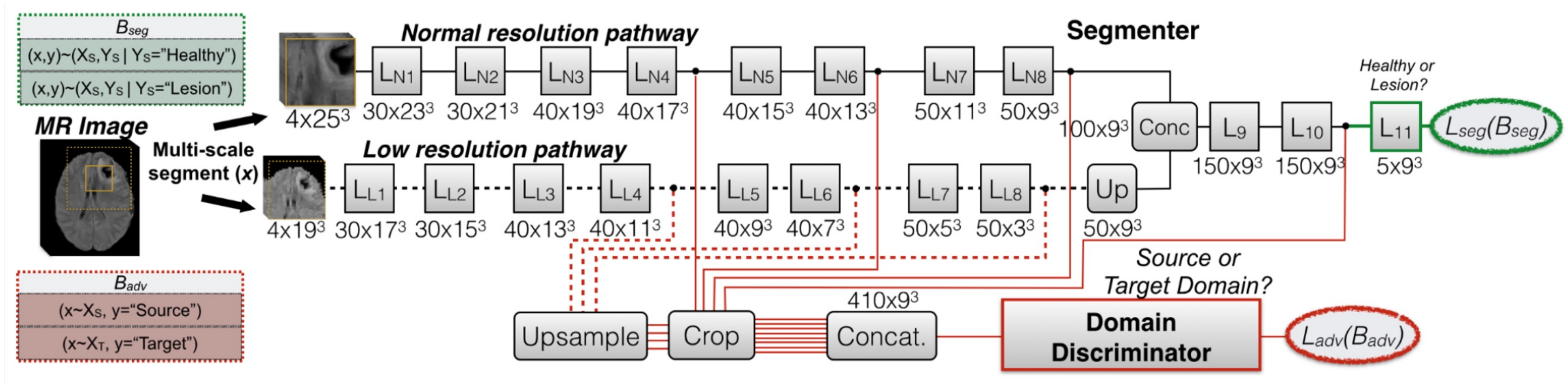


Like semi-supervised segmentation except target images are from a different domain

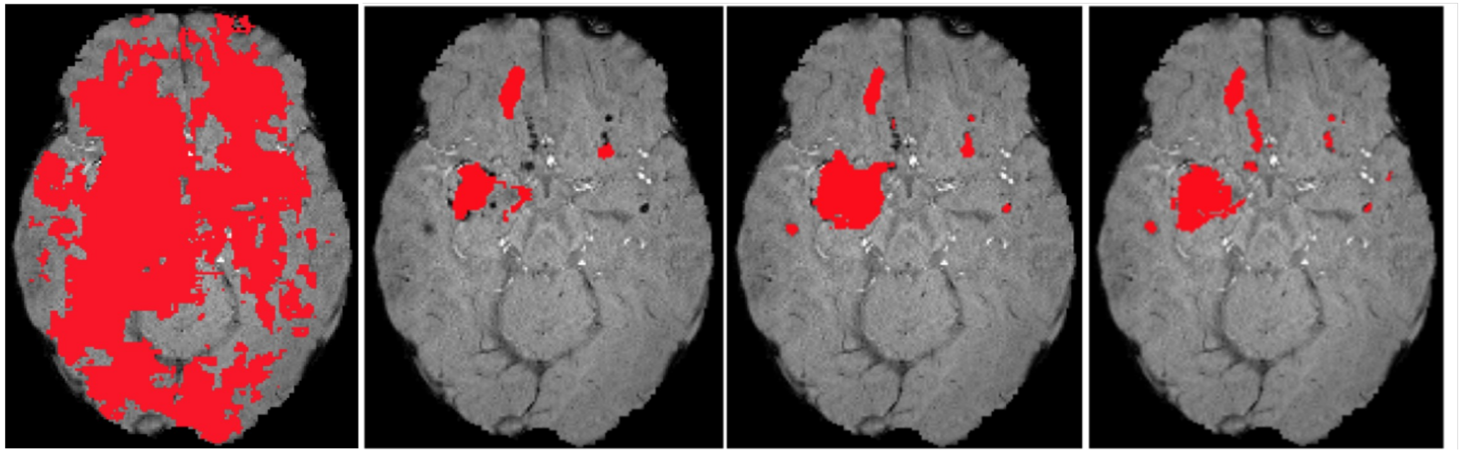
Tsai, Y.-H., et al. "Learning to adapt structured output space for semantic segmentation." IEEE Conf. on Computer Vision and Pattern Recognition. 2018.

Adversarial domain adaptation

Adversarial domain adaptation for brain lesion segmentation



	No adaptation (All sequences)	No adaptation (No GE/SWI)	Adaptation S→T	Ground-truth



Source domain (Database 1):

- GE, FLAIR, T2, MPRAGE, PD

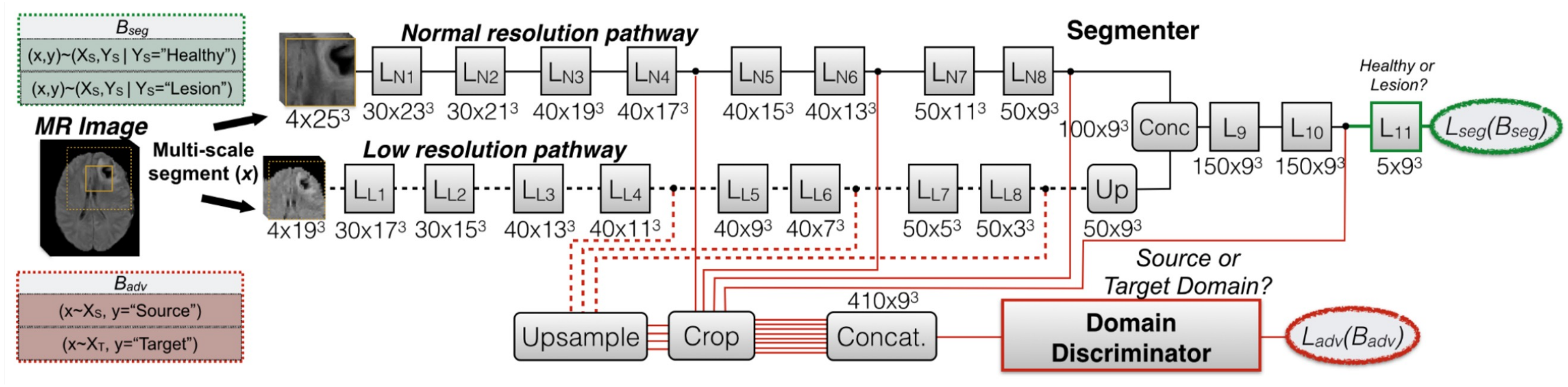
Target domain (Database 2):

- SWI, FLAIR, T2, MPRAGE, PD

Image from Kamnitsas, K., et al. "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks." *Int. Conf. on Information Processing in Medical Imaging*, 2017.

Adversarial domain adaptation

Adversarial domain adaptation for brain lesion segmentation



Adaptation done on *multi-scale feature representation*

Source domain (Database 1):

- GE, FLAIR, T2, MPRAGE, PD

Target domain (Database 2):

- SWI, FLAIR, T2, MPRAGE, PD

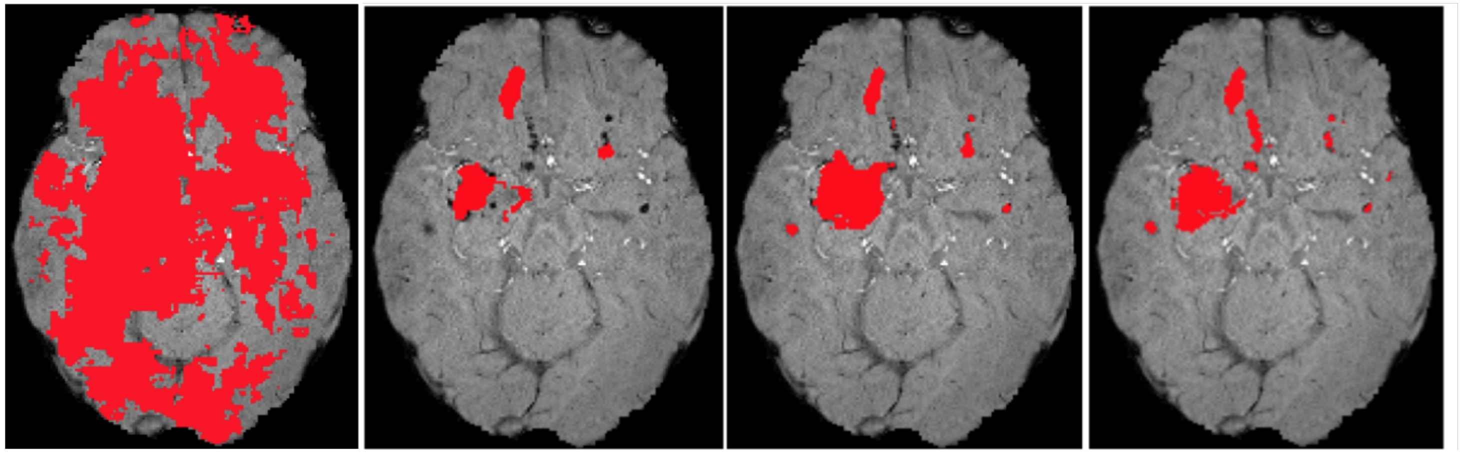


Image from Kamnitsas, K., et al. "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks." *Int. Conf. on Information Processing in Medical Imaging*, 2017.

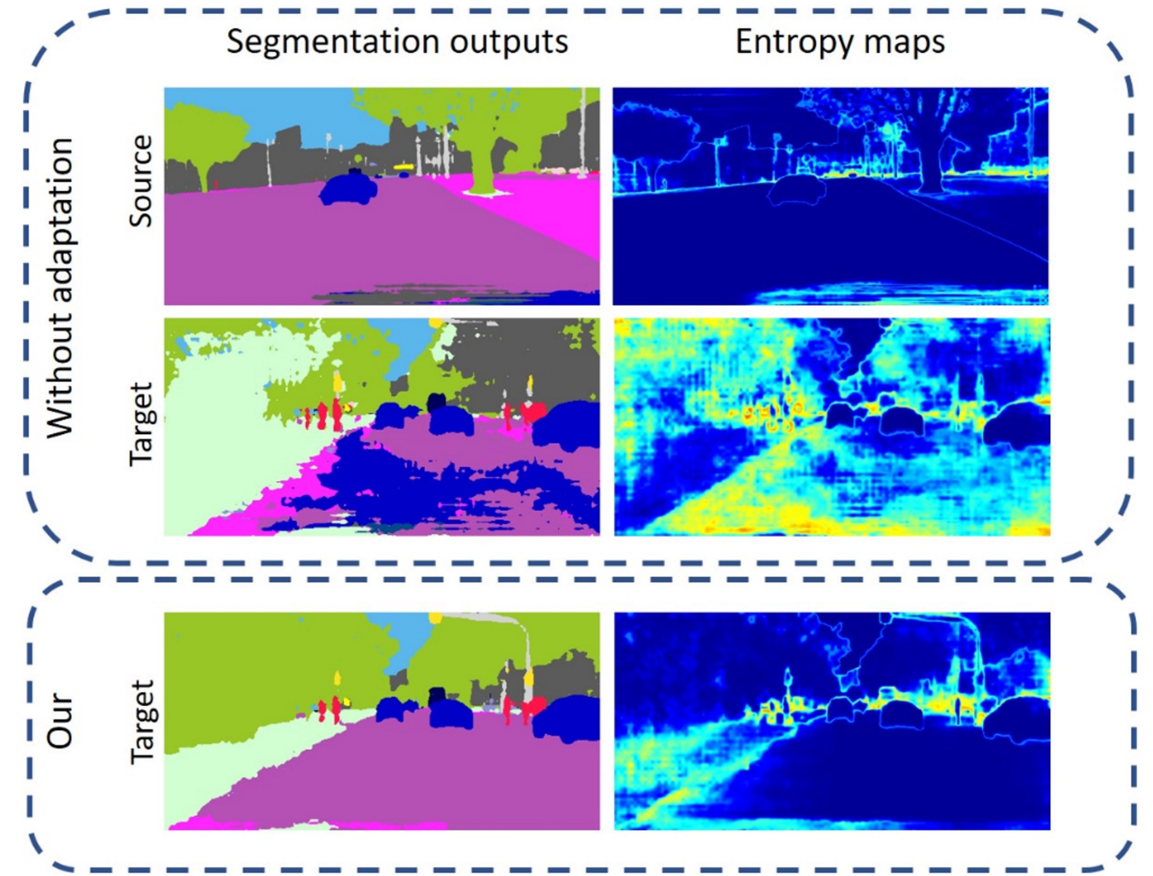
Adversarial domain adaptation

Adaptation on feature *representation* or *softmax* output. <u>What else</u> ?

Vu, T.-H., et al. "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation." IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2019.

Adversarial domain adaptation

Adaptation on feature *representation* or *softmax* output. What else ?



Adversarial entropy minimization

- The discriminator must differentiate between source and target examples using the entropy spatial maps
- Forces the segmentation model to be consistent in its **confidence** across different semantic regions

Vu, T.-H., et al. "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation." IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2019.

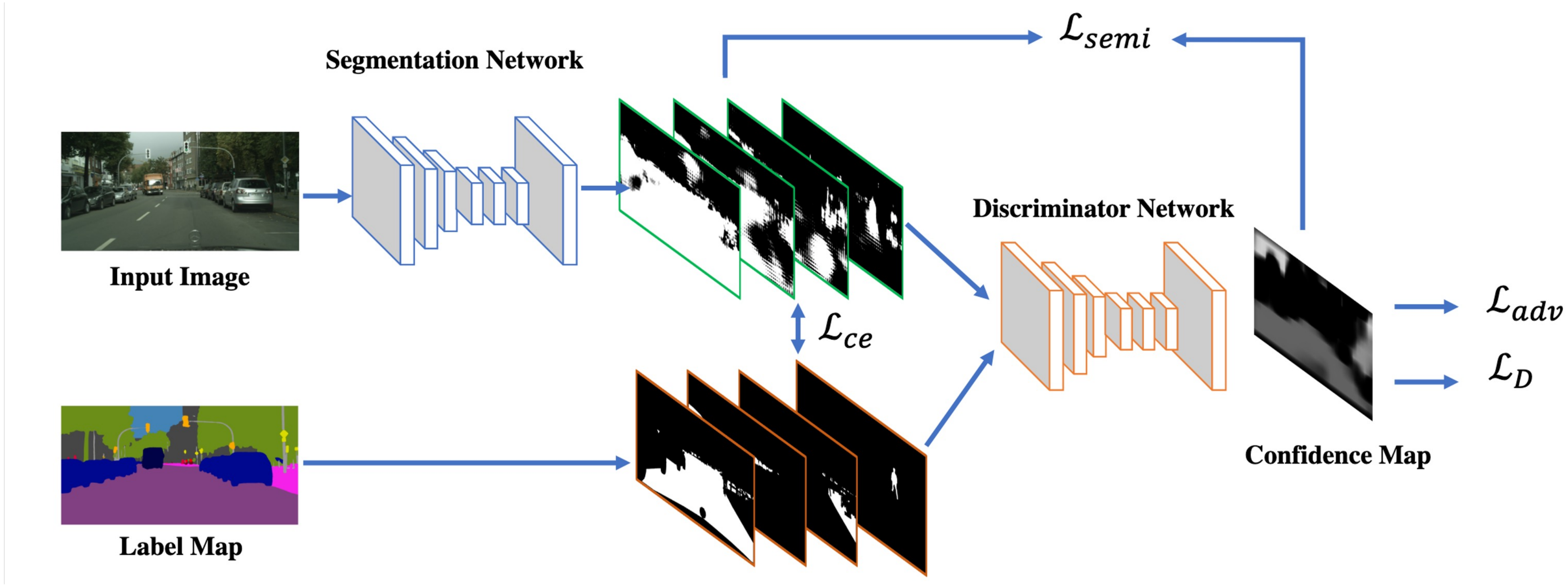
Adversarial model for self-training

How can we leverage discriminator predictions at the pixel-level ?

Hung, W.-C. et al. "Adversarial learning for semi-supervised semantic segmentation." *British Machine Vision Conference* (*BMVC*), 2018.

Adversarial model for self-training

How can we leverage discriminator predictions at the pixel-level ?

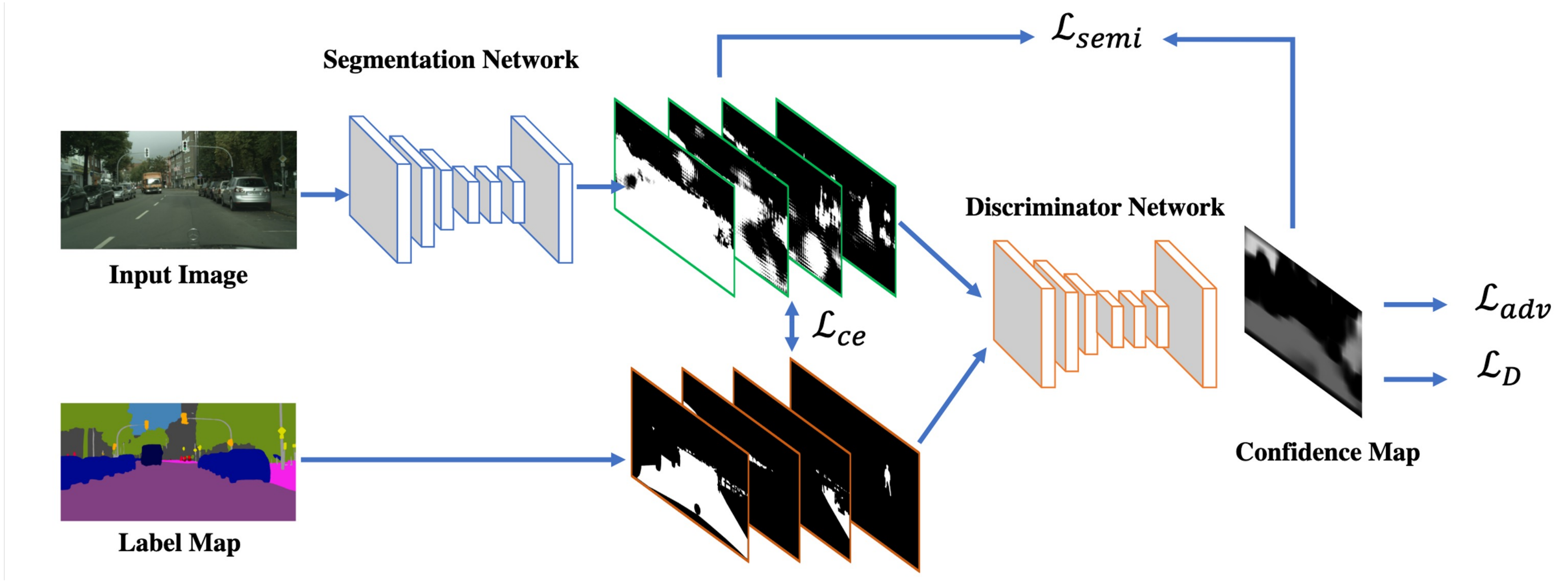


- The discriminator must discriminate between prediction and ground-truth (GT) at each pixel
- Consider the discriminator GT-class probabilities as confidence scores
- Use high-confidence predictions on unlabeled images as pseudo-labels for self-training

Hung, W.-C. et al. "Adversarial learning for semi-supervised semantic segmentation." *British Machine Vision Conference* (*BMVC*), 2018.

Adversarial model for self-training

How can we leverage discriminator predictions at the pixel-level ?



Select pixels with confidence above a given threshold

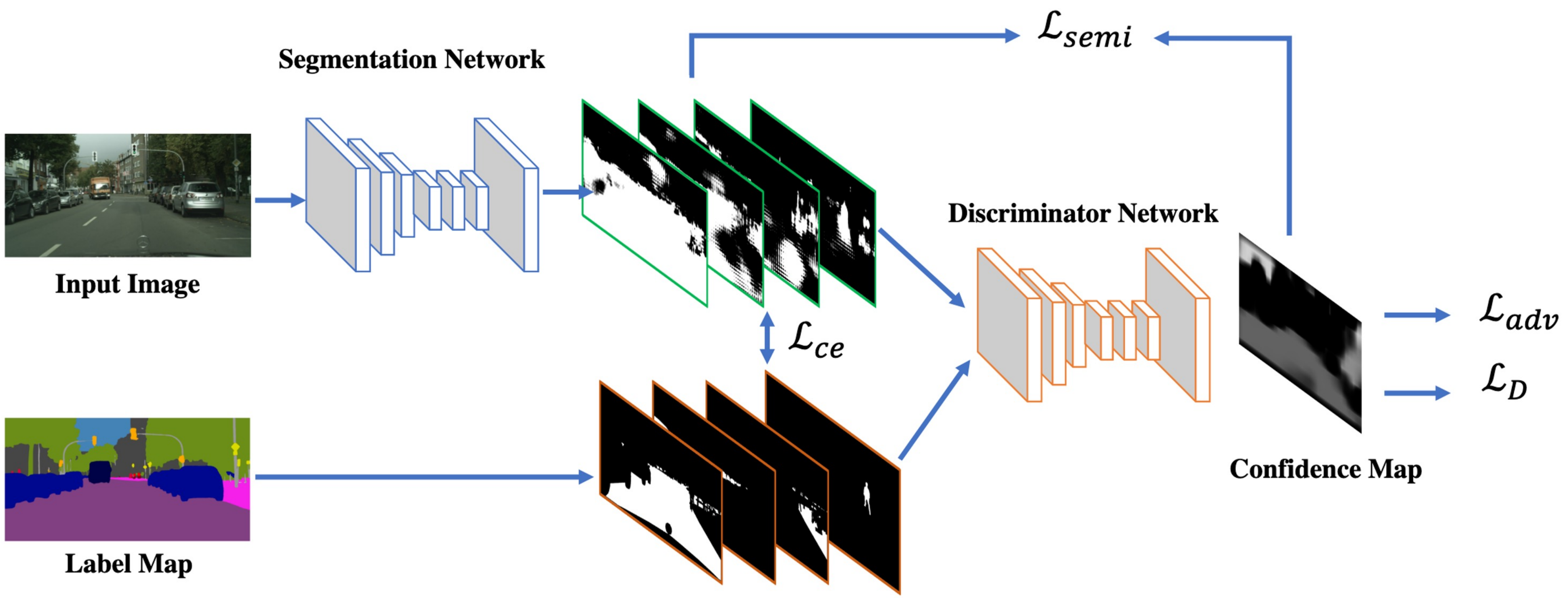
$$\mathcal{L}_{semi} = -\sum_{h,w}\sum_{c\in C} I(D(S(\mathbf{X}_n))^{(h,w)} > T_{semi}) \cdot \hat{\mathbf{Y}}_n^{(h,w,c)} \log(S(\mathbf{X}_n)^{(h,w,c)})$$

$$\hat{\mathbf{Y}}_n^{(h,w,c^*)} = 1 \text{ if } c^* = \arg\max_c S(\mathbf{X}_n)^{(h,w,c)}$$

Hung, W.-C. et al. "Adversarial learning for semi-supervised semantic segmentation." *British Machine Vision Conference* (*BMVC*), 2018.

Adversarial model for self-training

How can we leverage discriminator predictions at the pixel-level ?



Use class with highest probability as pseudo-label

$$\mathcal{L}_{semi} = -\sum_{h,w}\sum_{c \in C} I(D(S(\mathbf{X}_n))^{(h,w)} > T_{semi}) \cdot \hat{\mathbf{Y}}_n^{(h,w,c)} \log(S(\mathbf{X}_n)^{(h,w,c)})$$

$$\hat{\mathbf{Y}}_n^{(h,w,c^*)} = 1 \text{ if } c^* = \arg\max_c S(\mathbf{X}_n)^{(h,w,c)}$$

Hung, W.-C. et al. "Adversarial learning for semi-supervised semantic segmentation." *British Machine Vision Conference* (*BMVC*), 2018.

Cycle GANs for domain adaptation

How can we learn a model to segment target images without paired images or GT ?

Source domain

Image

Ground-truth

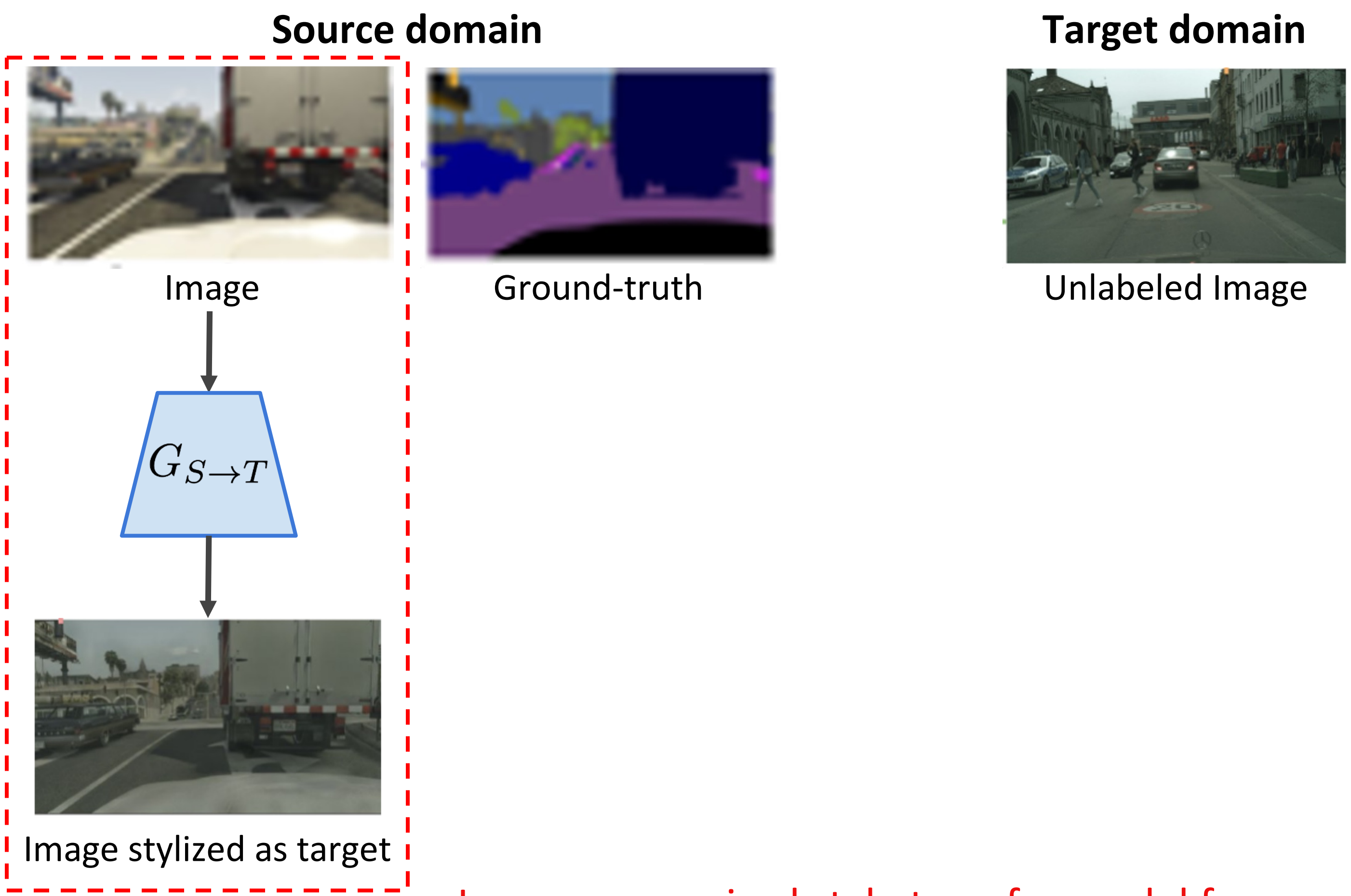
Target domain

Unlabeled Image

Hoffman, J., et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." *Int. Conf. on Machine Learning* (ICML). 2018.

Cycle GANs for domain adaptation

How can we learn a model to segment target images without paired images or GT ?

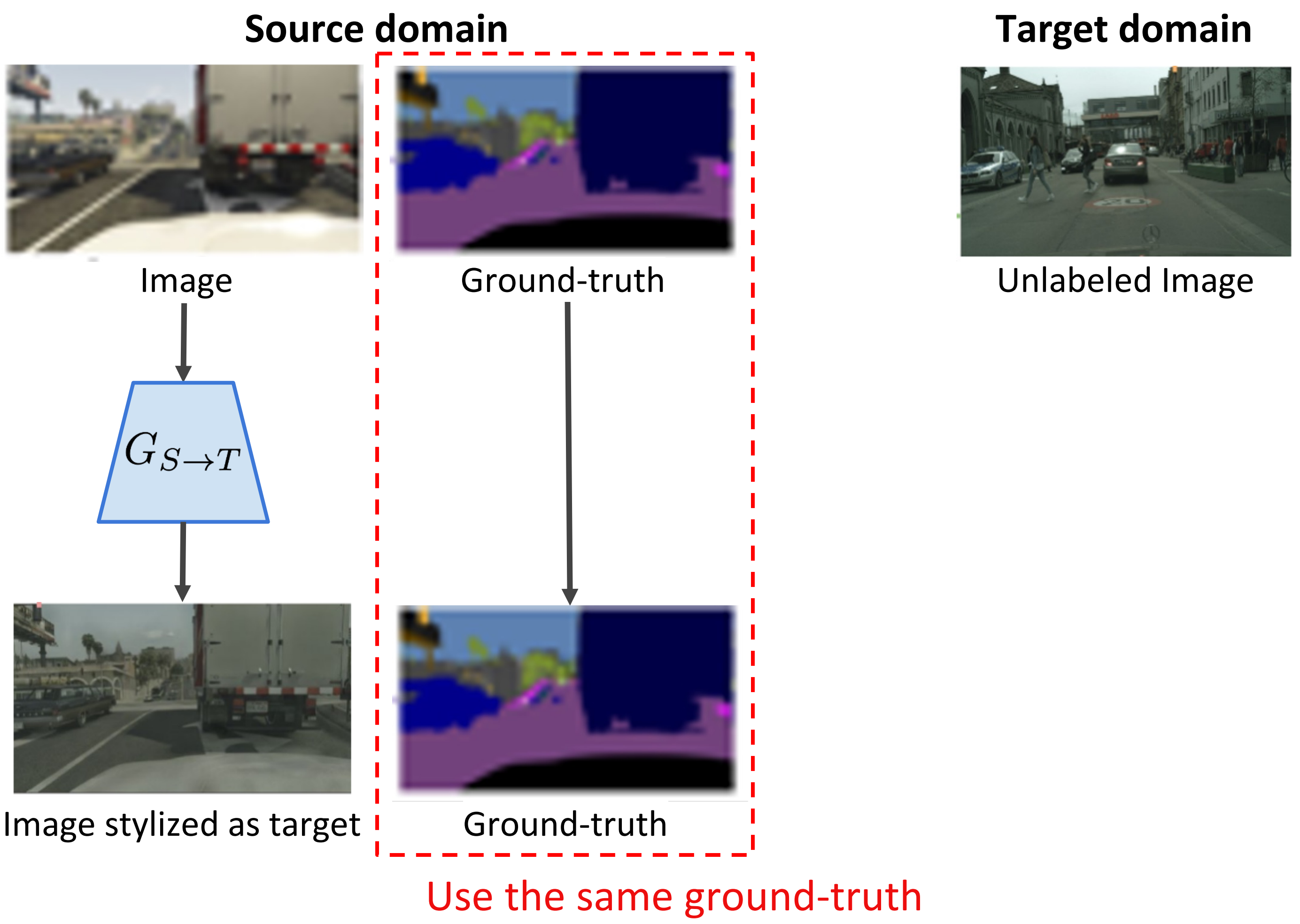


Learn an unpaired style transfer model from source to target domain

Hoffman, J., et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." *Int. Conf. on Machine Learning* (ICML). 2018.

Cycle GANs for domain adaptation

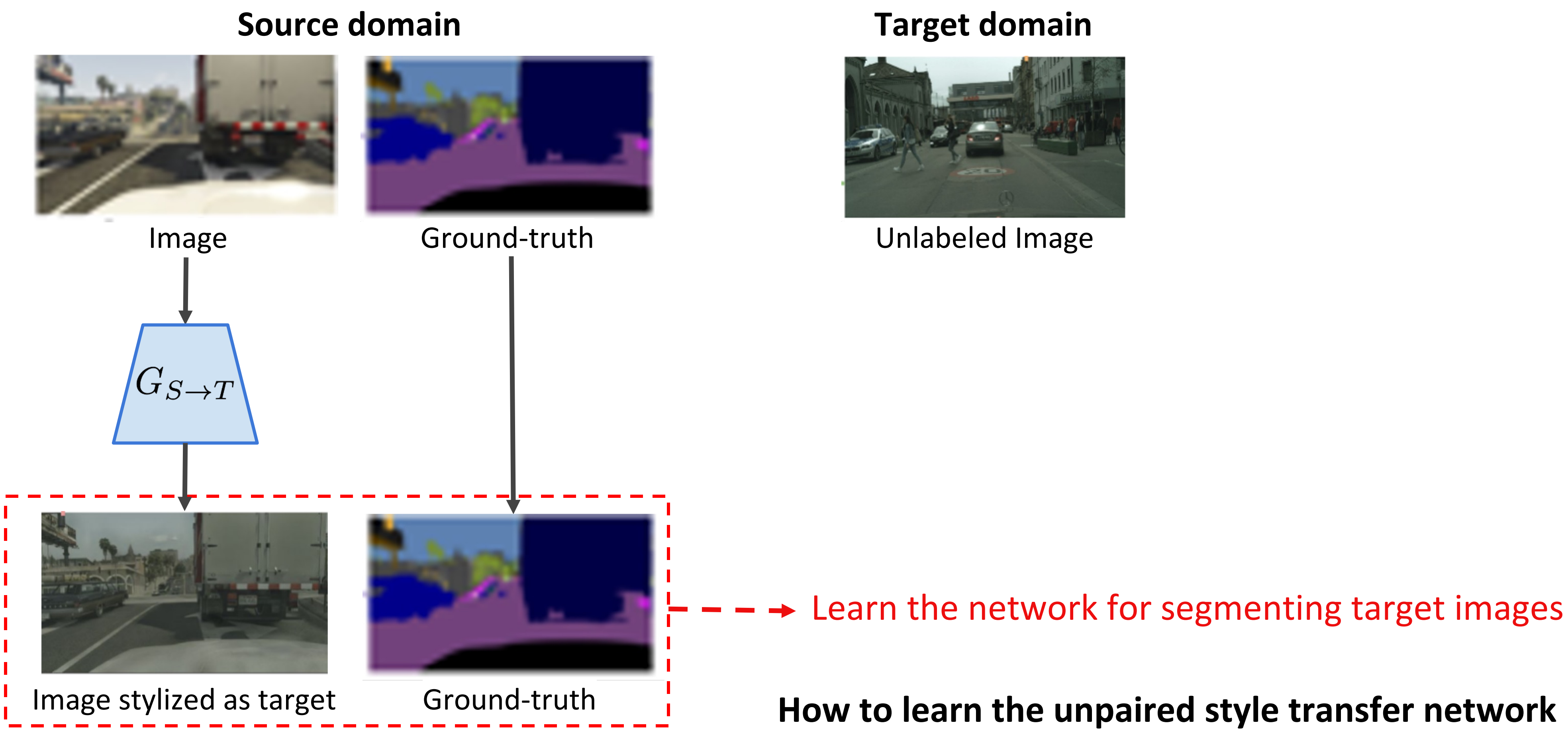
How can we learn a model to segment target images without paired images or GT ?



Hoffman, J., et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." *Int. Conf. on Machine Learning* (ICML). 2018.

Cycle GANs for domain adaptation

How can we learn a model to segment target images without paired images or GT ?

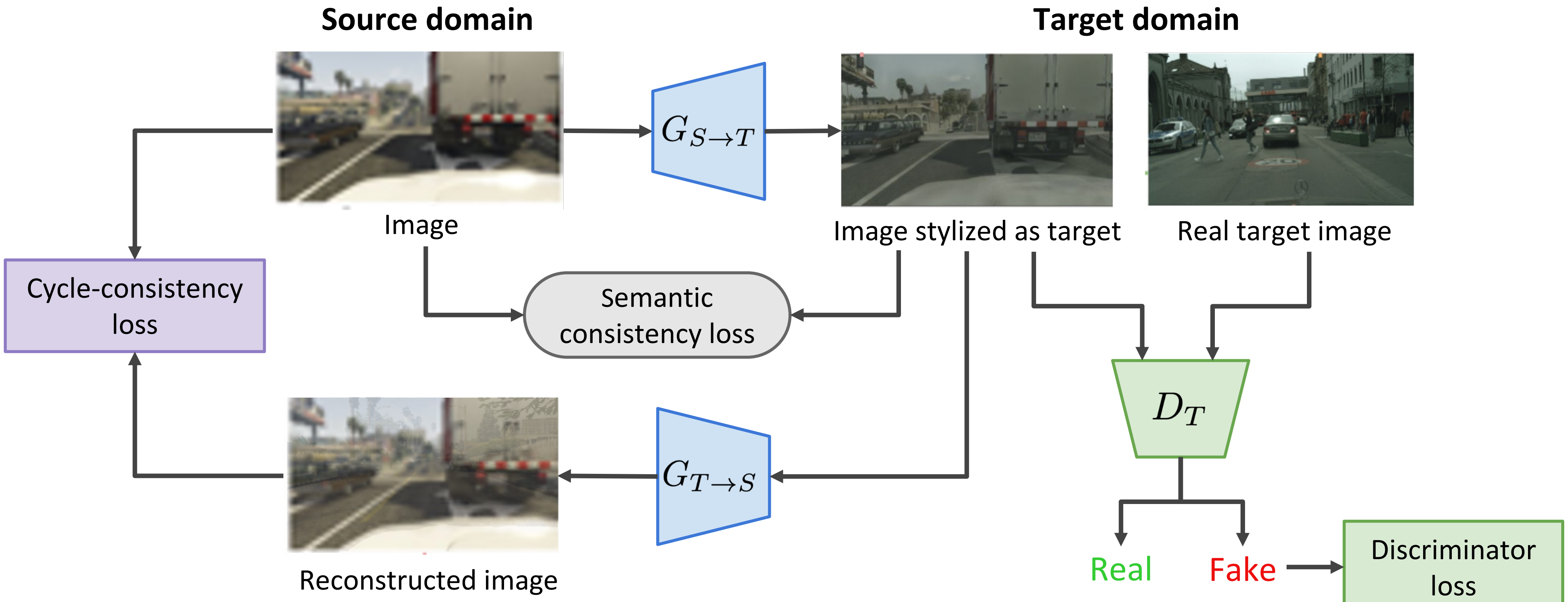


How to learn the unpaired style transfer network ?

Hoffman, J., et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." *Int. Conf. on Machine Learning* (ICML). 2018.

Cycle GANs for domain adaptation

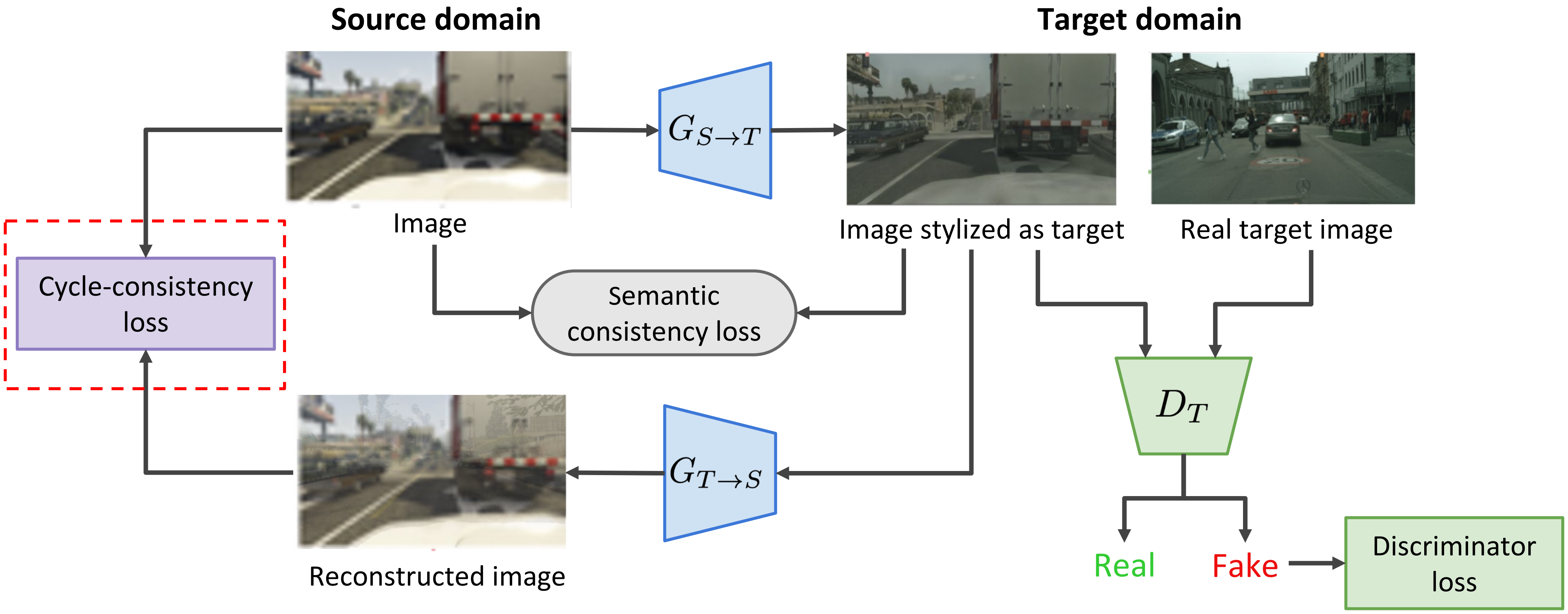
How can we learn a model to segment target images without paired images or GT ?



Hung, W.-C. et al. "Adversarial learning for semi-supervised semantic segmentation." *British Machine Vision Conference* (*BMVC*), 2018.

Cycle GANs for domain adaptation

How can we learn a model to segment target images without paired images or GT ?

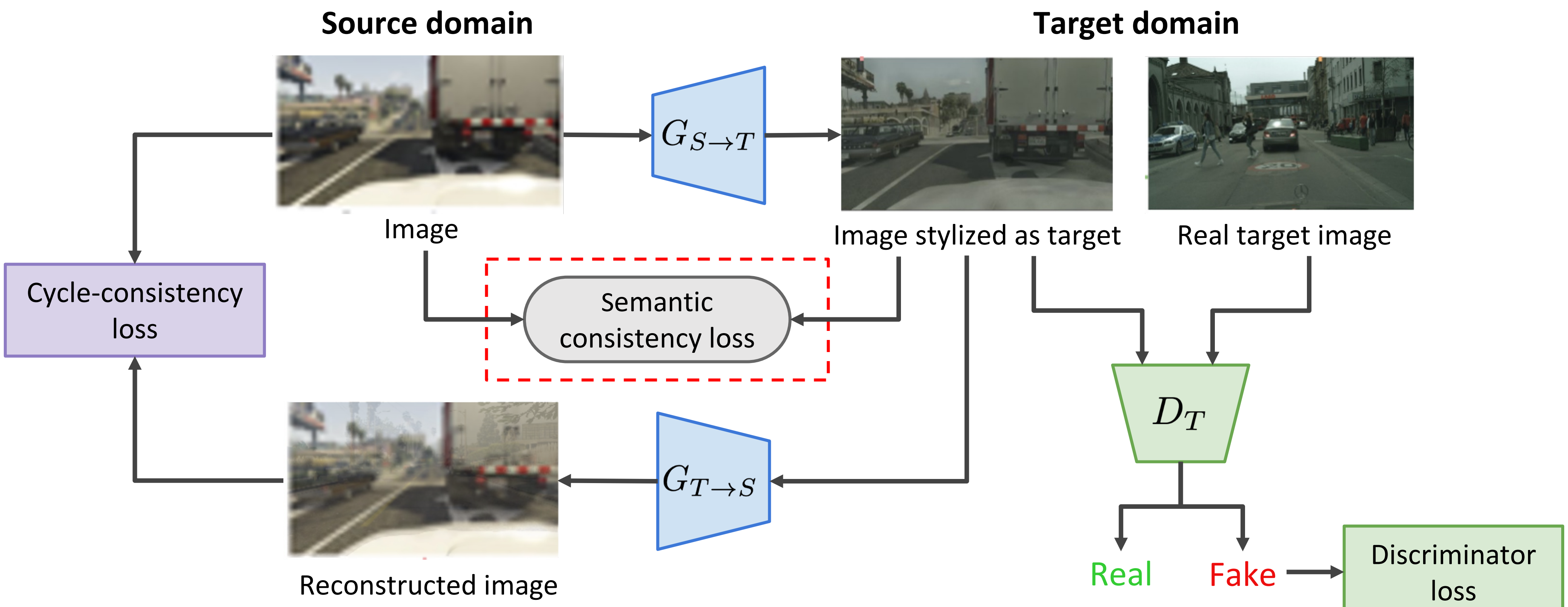


Cycle consistency loss: $L_{\text{cycle}}(G_{S\to T}, G_{T\to S}) = \mathbb{E}_{x\sim p_S(x)}\left[\|x - G_{T\to S}(G_{S\to T}(x))\|_1\right]$

Hoffman, J., et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." *Int. Conf. on Machine Learning* (ICML). 2018.

Cycle GANs for domain adaptation

How can we learn a model to segment target images without paired images or GT ?

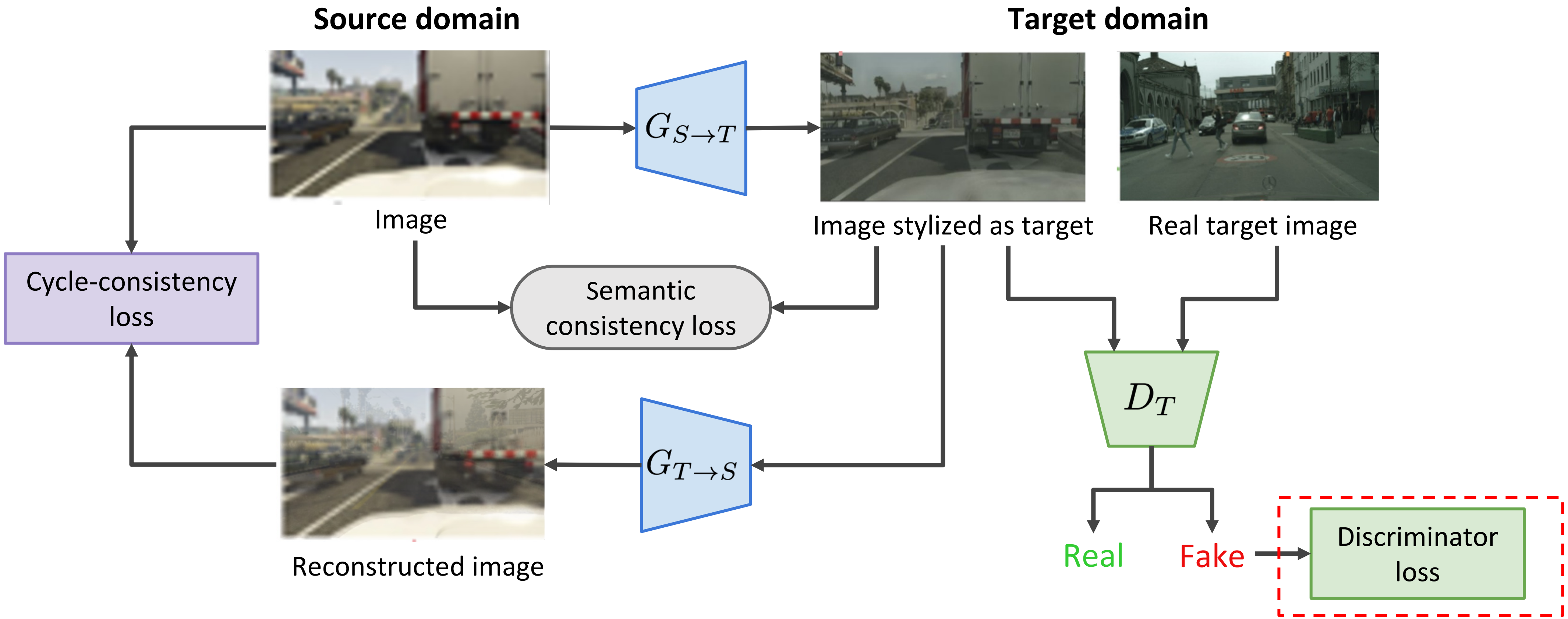


Semantic consistency loss: Segmentation for the source image and its stylized target version should be <u>consistent</u>

Hoffman, J., et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." *Int. Conf. on Machine Learning* (ICML). 2018.

Cycle GANs for domain adaptation

How can we learn a model to segment target images without paired images or GT ?

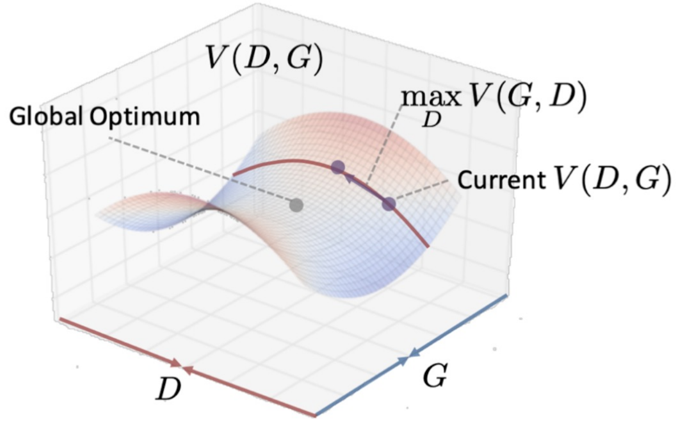


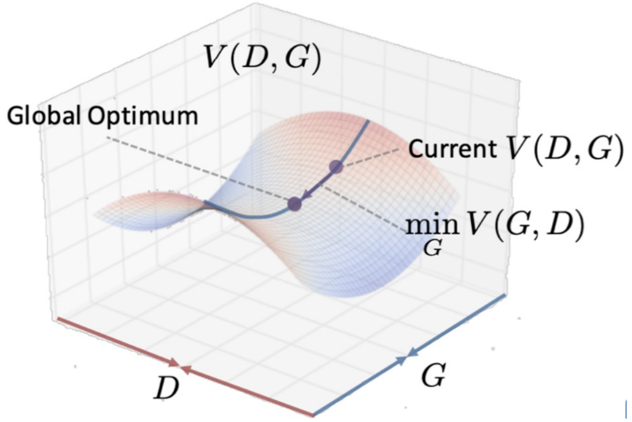
Discriminator loss: Target images generated from source should look like real target ones

Hoffman, J., et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." *Int. Conf. on Machine Learning* (ICML). 2018.

Challenges of adversarial learning

1) Unstable optimization of minimax problem

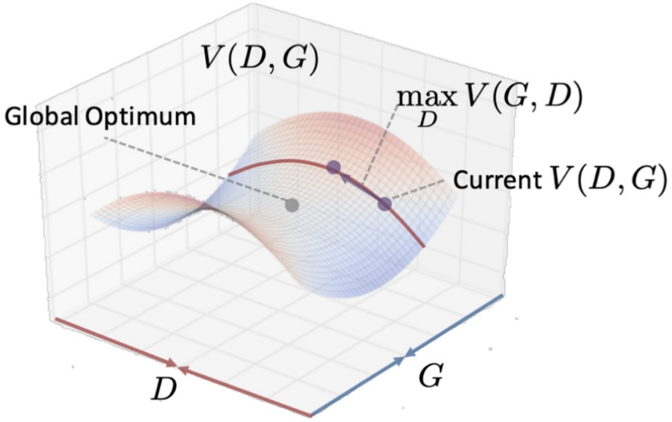


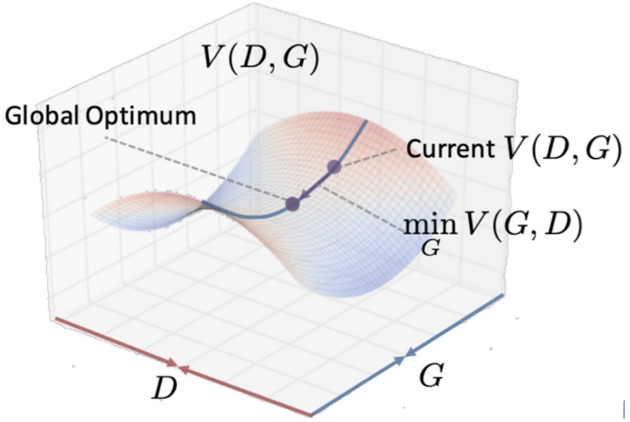


Ghosh, A., et al. "Multi-agent diverse generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
Chang, Mark. "Generative Adversarial Networks", published online, 2016

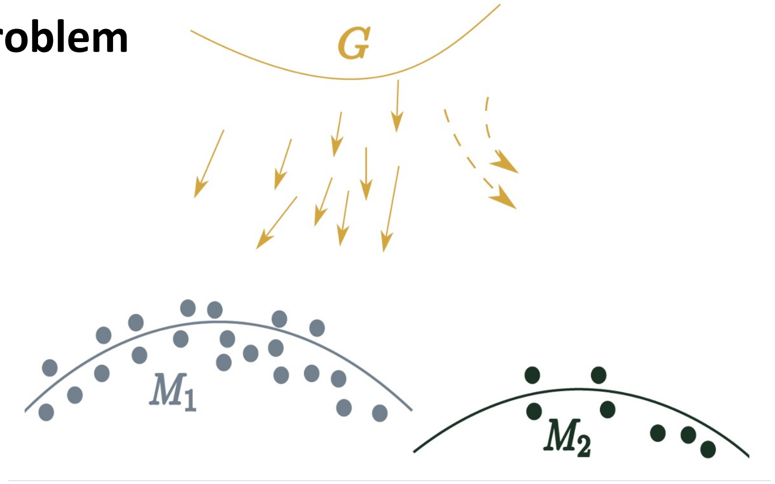
Challenges of adversarial learning

1) Unstable optimization of minimax problem





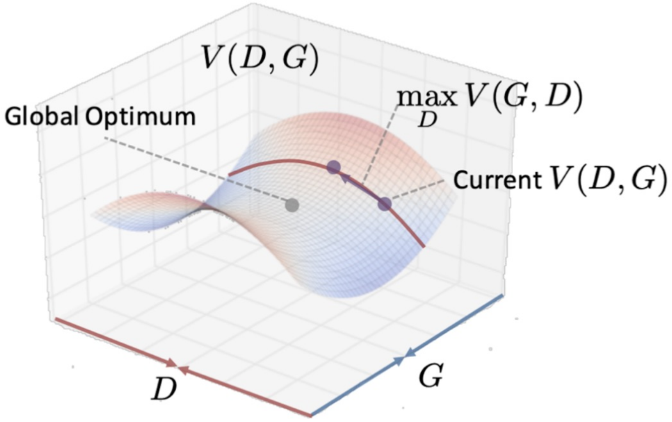
2) Mode collapse problem

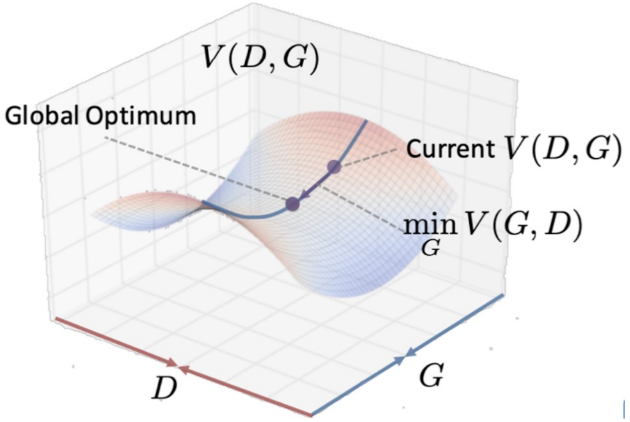


Ghosh, A., et al. "Multi-agent diverse generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
Chang, Mark. “Generative Adversarial Networks”, published online, 2016

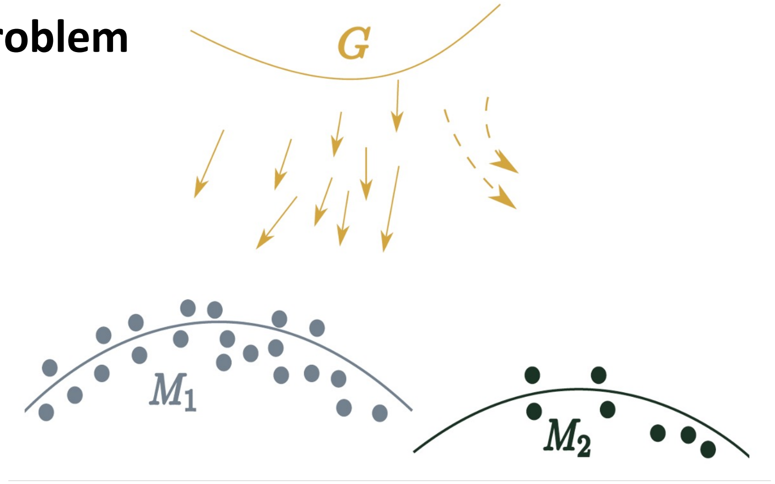
Challenges of adversarial learning

1) Unstable optimization of minimax problem





2) Mode collapse problem



Various solutions:

- Spectral normalization (Miyato *et al.*, 2018)
- Wasserstein GANs (Arjovsky *et al.*, 2017)
- LSGANs (Mao *et al.*, 2017)
- etc.

Ghosh, A., et al. "Multi-agent diverse generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
Chang, Mark. "Generative Adversarial Networks", published online, 2016

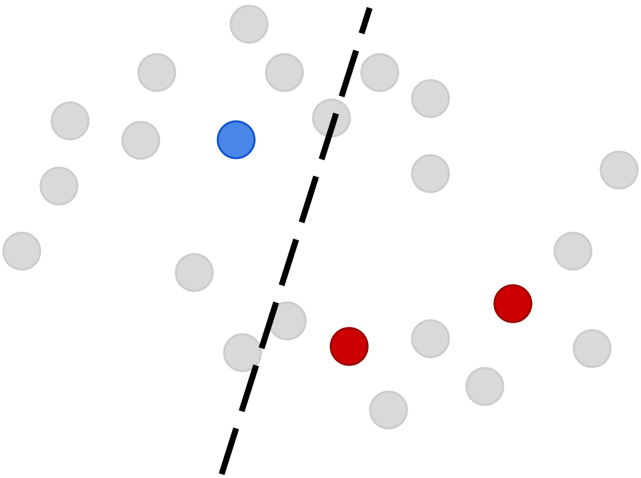
Consistency regularization

for semi-supervised segmentation

Consistency regularization for SSL

How to better use unlabeled data ?

Vanilla supervised learning

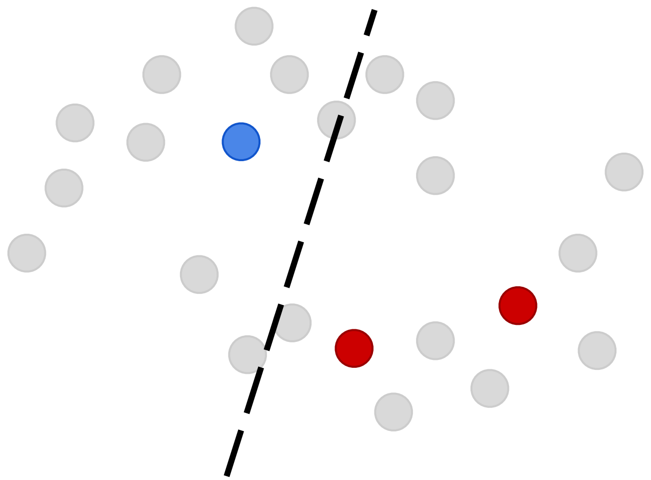


- Consider only *labeled* samples
- Overfits when few training samples

Consistency regularization for SSL

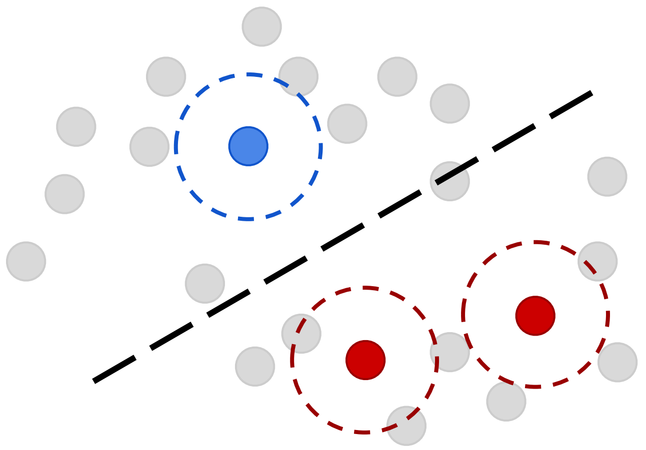
How to better use <u>unlabeled data</u> ?

Vanilla supervised learning



- Consider only *labeled* samples
- Overfits when few training samples

Data augmentation

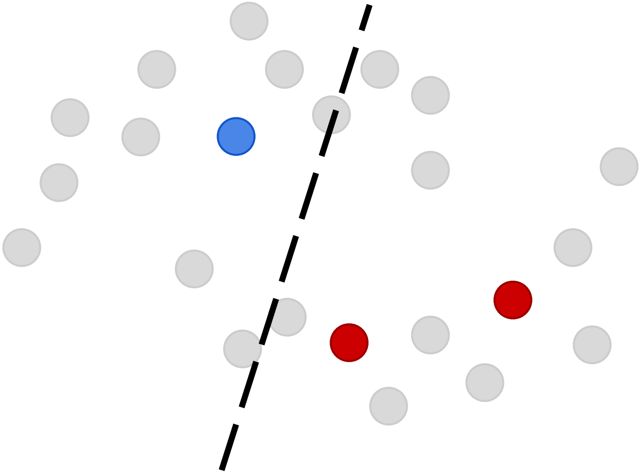


- Transform *labeled* samples to augment the training set
- Better generalization, but not enough for semi-supervised learning

Consistency regularization for SSL

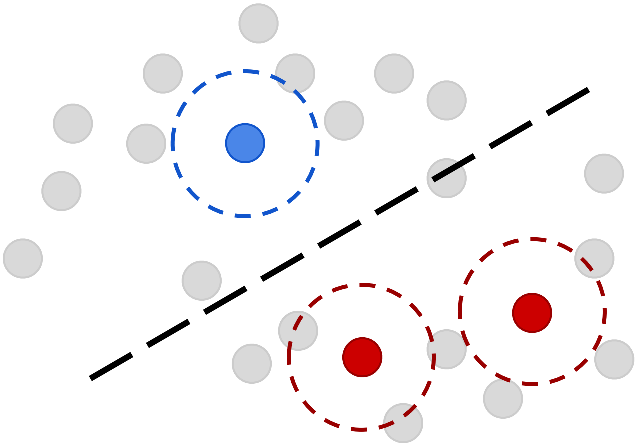
How to better use unlabeled data ?

Vanilla supervised learning



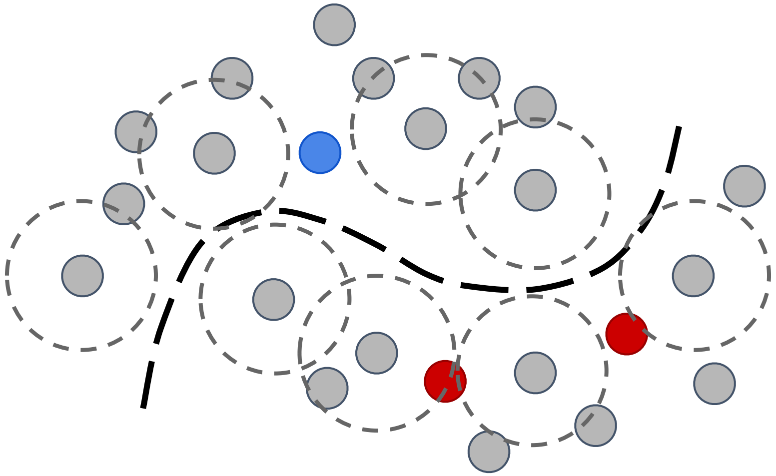
- Consider only *labeled* samples
- Overfits when few training samples

Data augmentation



- Transform *labeled* samples to augment the training set
- Better generalization, but not enough for semi-supervised learning

Consistency regularization



- Perturb *unlabeled* samples with noise or guided transformations
- Impose the network to have consistent outputs for perturbed samples

SSL methods using consistency regularization

Basic transformation consistency (Γ-model)

$$\mathcal{L}(\theta; \mathcal{D}_l, \mathcal{D}_u) \;=\; \frac{1}{|\mathcal{D}_l|} \sum_{(x,y)\in\mathcal{D}_l} \ell_{\text{sup}}\big(f(x), y\big) \;+\; \frac{\lambda}{|\mathcal{D}_u|} \sum_{x\in\mathcal{D}_u} \mathbb{E}_{T\sim p_T} \Big[\ell_{\text{reg}}\big(T(f(x)),\, f(T(x))\big)\Big]$$

SSL methods using consistency regularization

Basic transformation consistency (Γ-model)

Standard supervised loss

$$\mathcal{L}(\theta; \mathcal{D}_l, \mathcal{D}_u) = \frac{1}{|\mathcal{D}_l|} \sum_{(x,y)\in\mathcal{D}_l} \ell_{\text{sup}}\big(f(x), y\big) + \frac{\lambda}{|\mathcal{D}_u|} \sum_{x\in\mathcal{D}_u} \mathbb{E}_{T\sim p_T}\Big[\ell_{\text{reg}}\big(T(f(x)),\ f(T(x))\big)\Big]$$

Cross-entropy, Dice, etc.

SSL methods using consistency regularization

Basic transformation consistency (Γ-model)

Transformation consistency loss

$$\mathcal{L}(\theta; \mathcal{D}_l, \mathcal{D}_u) = \frac{1}{|\mathcal{D}_l|} \sum_{(x,y)\in\mathcal{D}_l} \ell_{\text{sup}}\big(f(x), y\big) + \frac{\lambda}{|\mathcal{D}_u|} \sum_{x\in\mathcal{D}_u} \mathbb{E}_{T\sim p_T}\Big[\ell_{\text{reg}}\big(T(f(x)),\, f(T(x))\big)\Big]$$

Random transformation: rotation, flip, crop, etc.

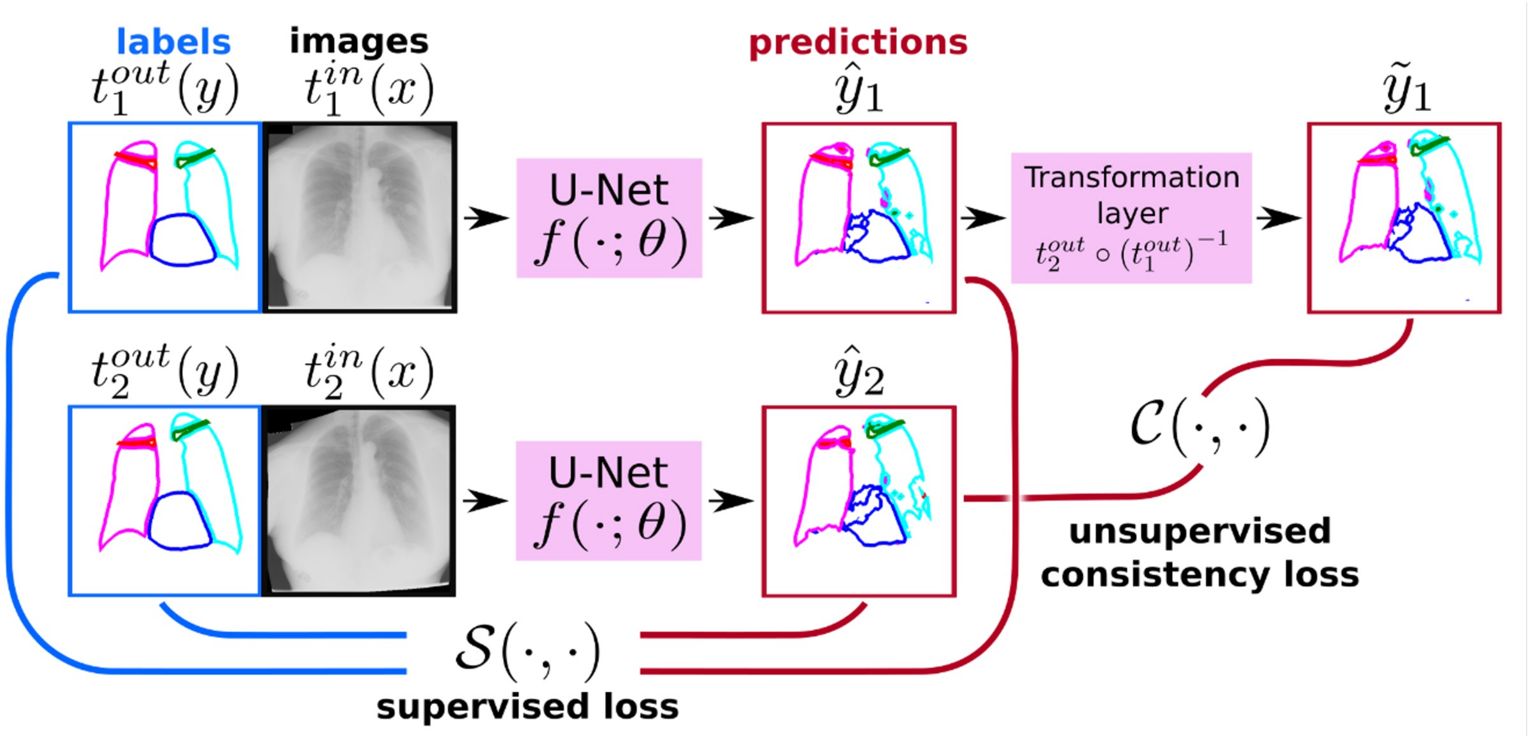
Regularization loss imposing *transformation equivariance*

L2 regularization loss:

$$\ell_{\text{reg}}\big(T(f(x)),\, f(T(x))\big) = \big\|T(f(x)) - f(T(x))\big\|_2^2$$

SSL methods using consistency regularization

Application to chest X-ray segmentation:

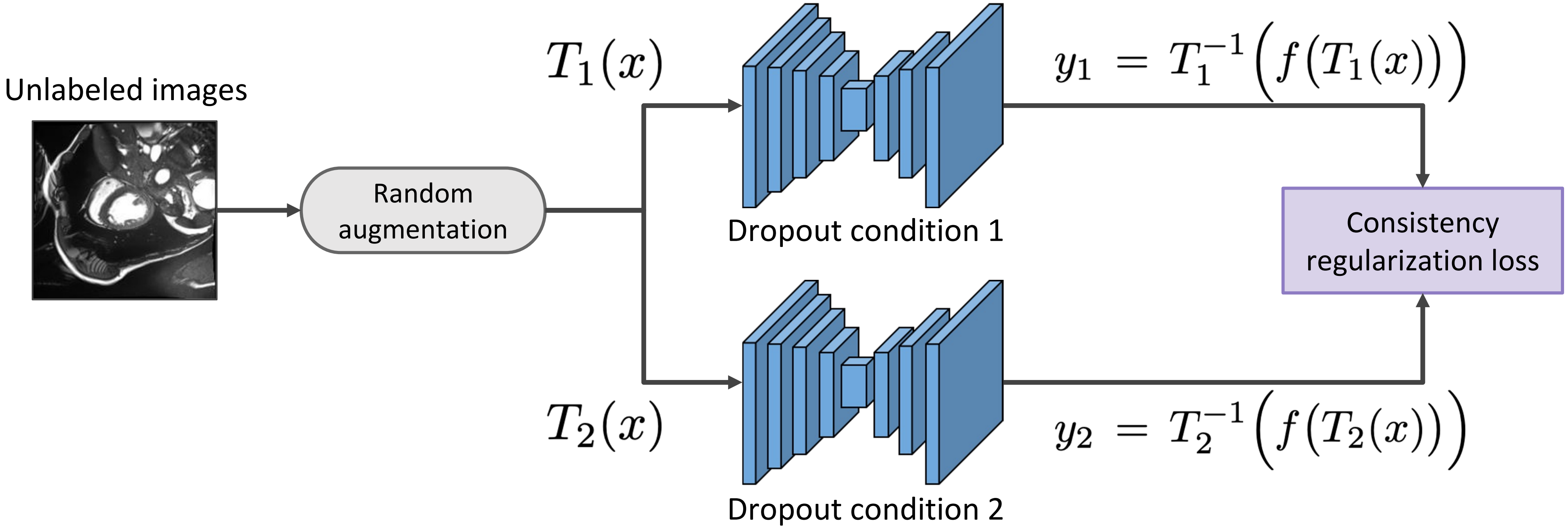


Transformations are random elastic deformations

Bortsova, G., et al.. "Semi-supervised medical image segmentation via learning consistency under transformations." MICCAI, 2019.

SSL methods using consistency regularization

Self-ensembling (Π-model):

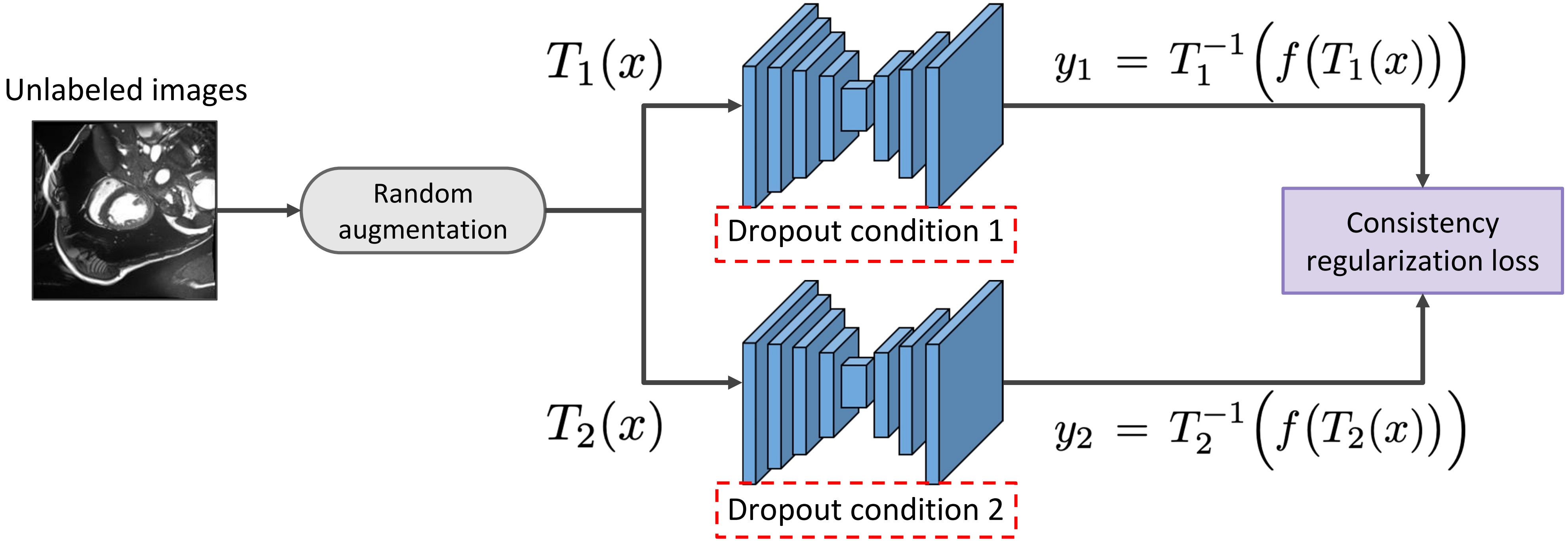


Key idea:

- Applying different dropouts on the same network gives an ensemble of models
- Also leverages random image transformations

SSL methods using consistency regularization

Self-ensembling (Π-model):

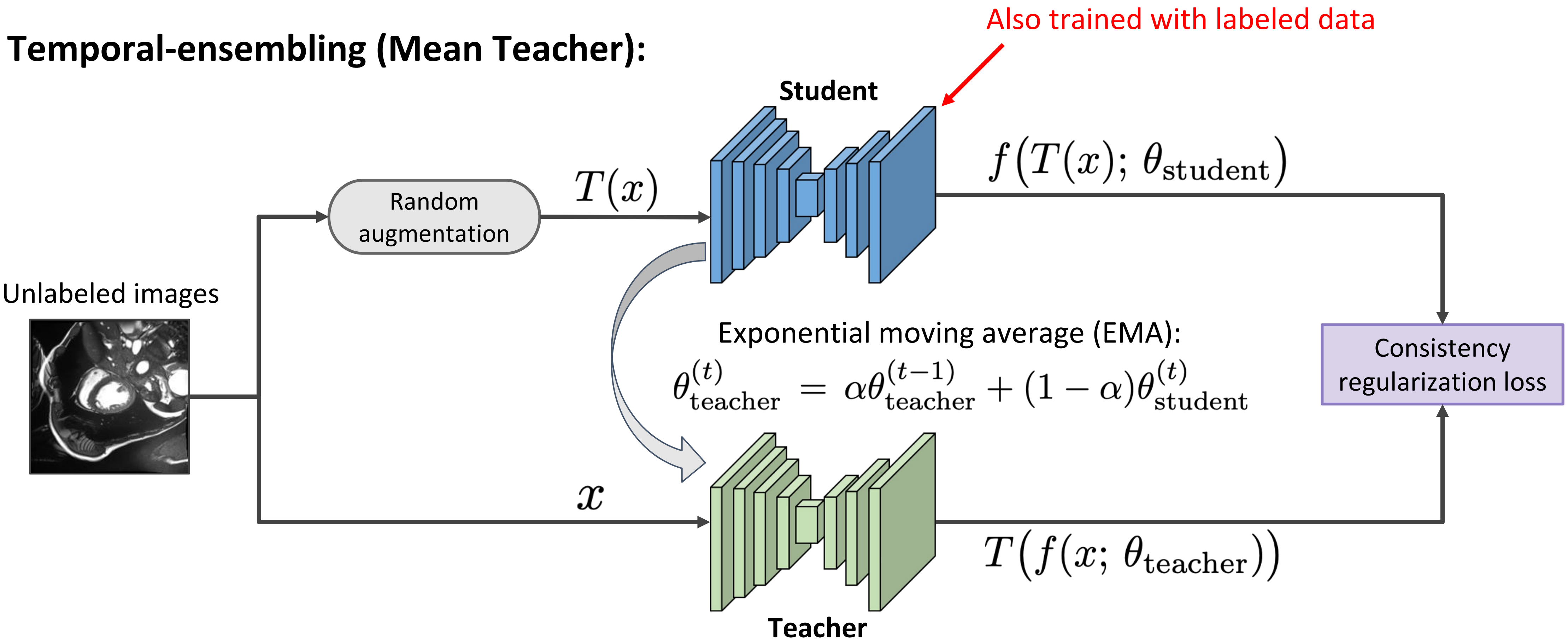


Key idea:

- Applying <u>different dropouts</u> on the <u>same network</u> gives an ensemble of models
- Also leverages random image transformations

SSL methods using consistency regularization

Temporal-ensembling (Mean Teacher):

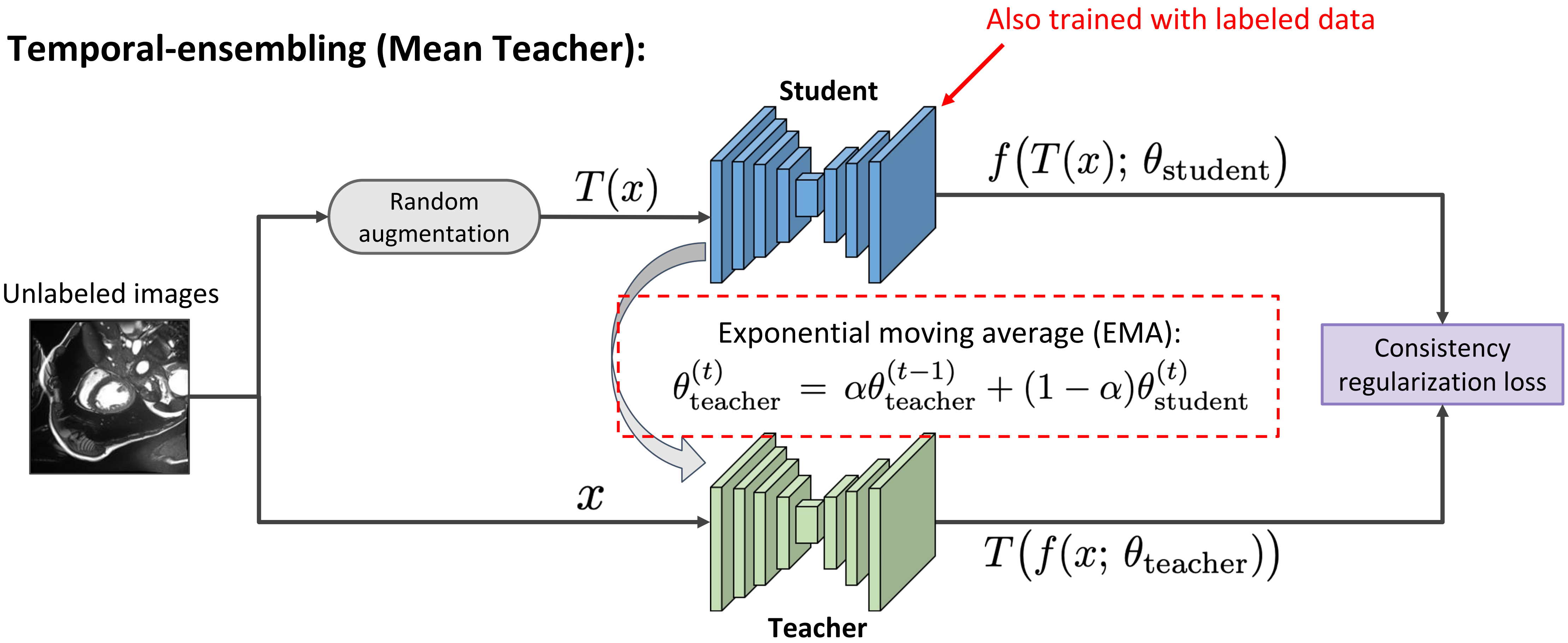


Key idea:

- Consistency between the predictions of a Teacher and a Student network
- The Teacher's weights are an EMA of the Student's at previous training iterations ($\alpha \approx 1$)
- *Note*: original Temporal Ensembling computes the EMA on outputs for each sample

SSL methods using consistency regularization

Temporal-ensembling (Mean Teacher):

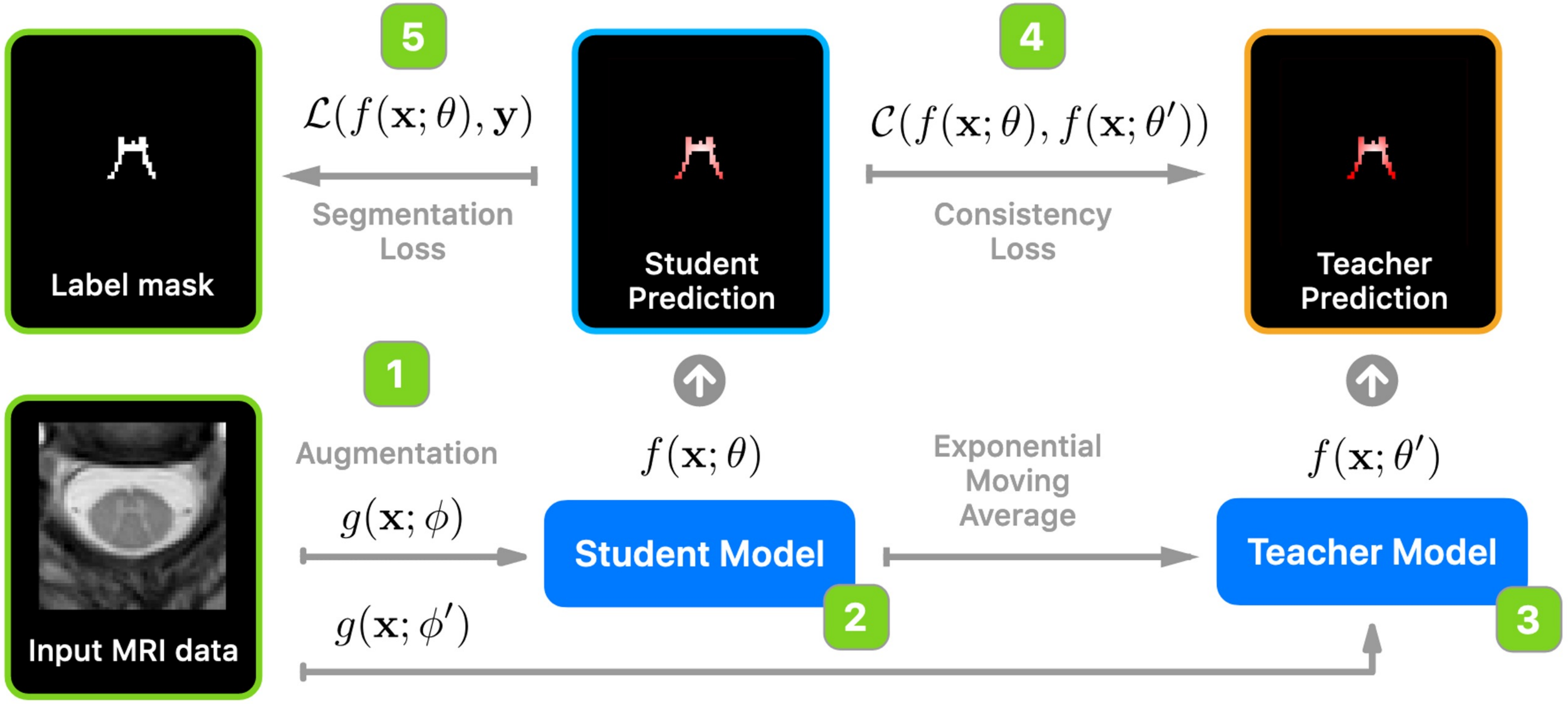


Key idea:

- Consistency between the predictions of a Teacher and a Student network
- The Teacher's weights are an EMA of the Student's at previous training iterations
- *Note*: original Temporal Ensembling computes the EMA on outputs for each sample

SSL methods using consistency regularization

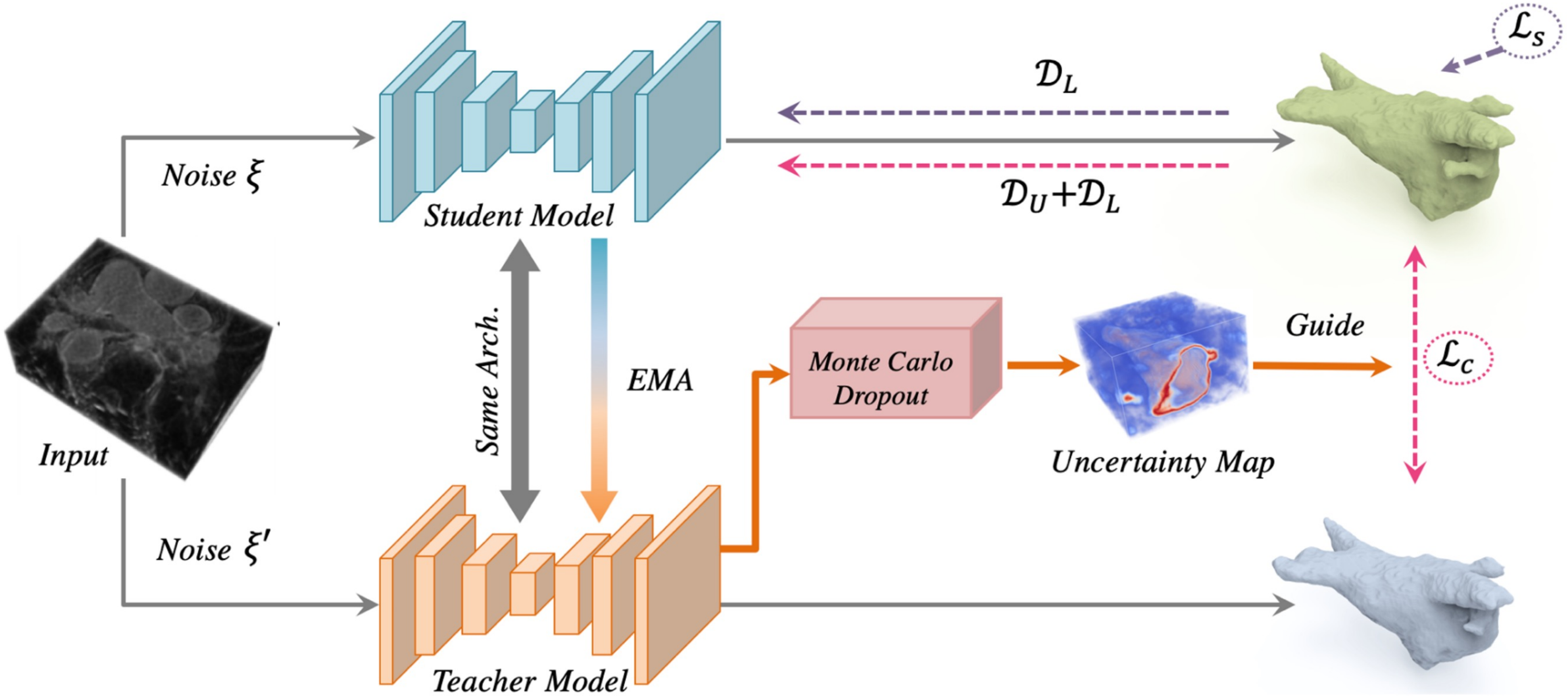
Application of Mean Teacher to segmenting MRI spinal cord gray matter



Perone, C.S. and Cohen-Adad, J. "Deep semi-supervised segmentation with weight-averaged consistency targets." Deep learning in medical image analysis and multimodal learning for clinical decision support, 2019 (*extended version in Neuroimage*)

SSL methods using consistency regularization

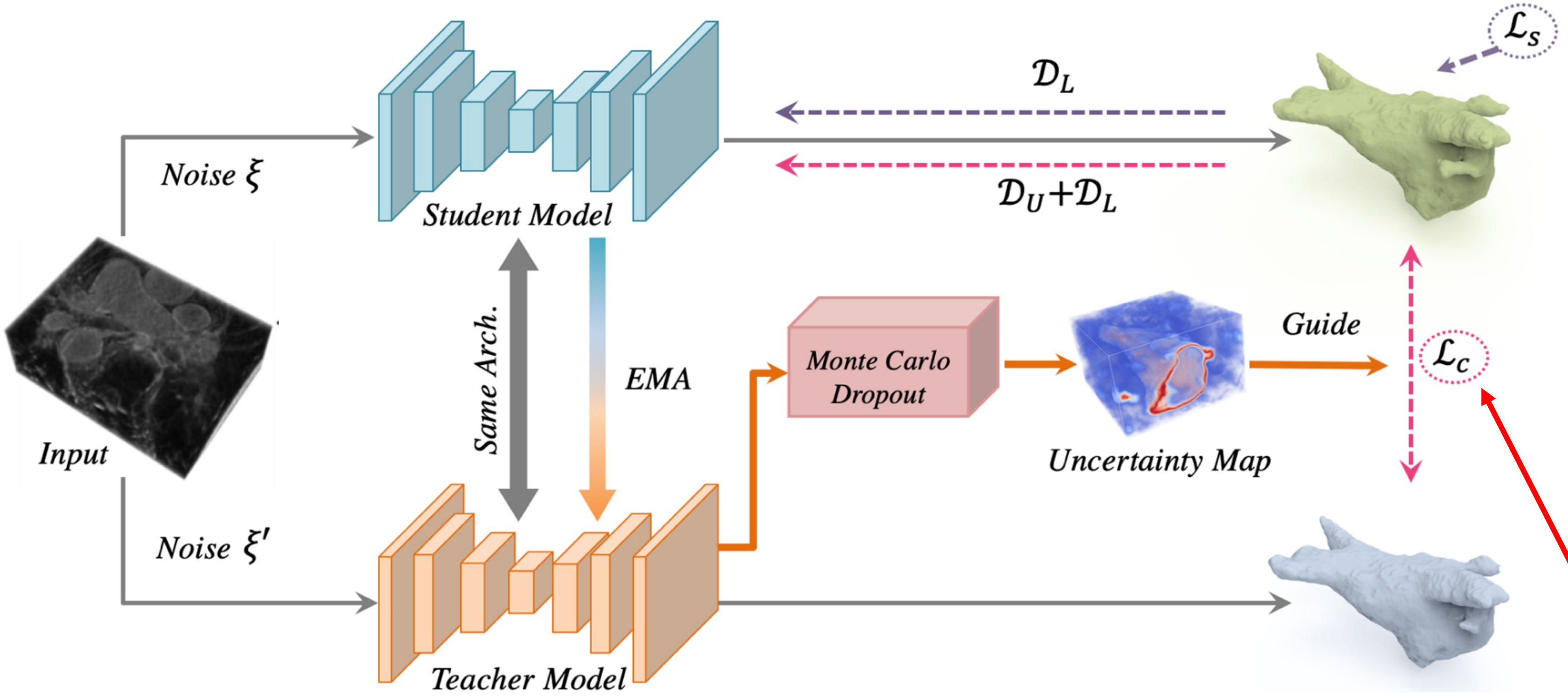
Uncertainty-aware self-ensembling



Yu, L et al.. "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation." MICCAI, 2019.

SSL methods using consistency regularization

Uncertainty-aware self-ensembling



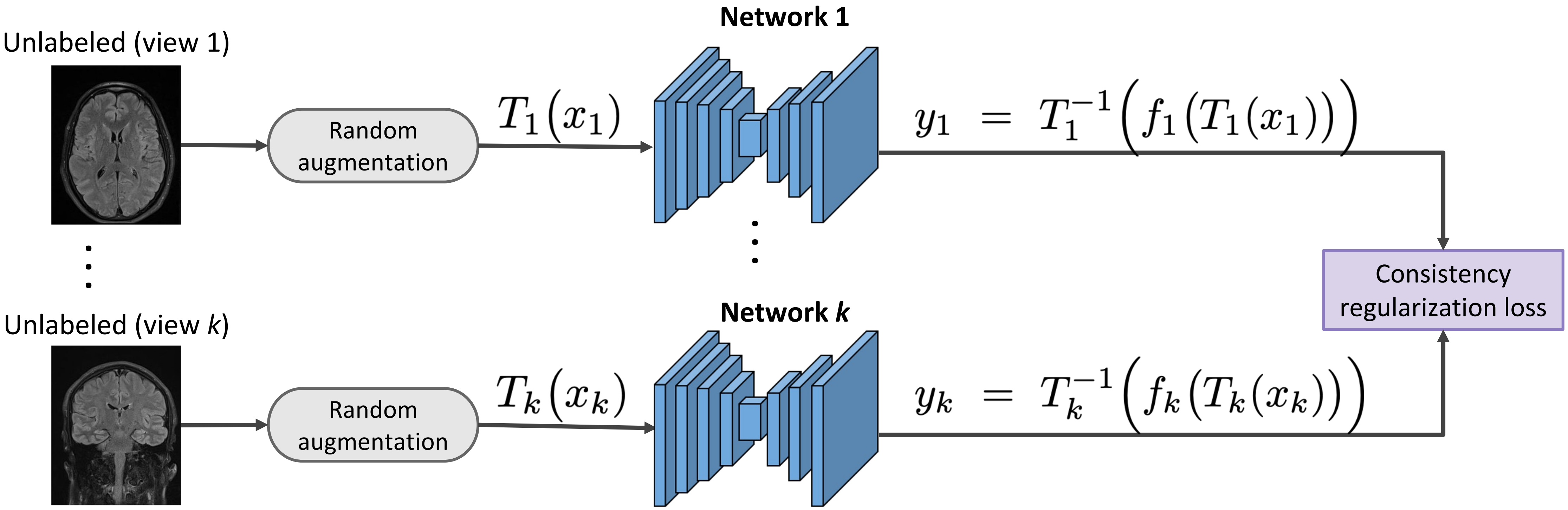
Enforces consistency only in low-uncertainty regions of the image

$$\mathcal{L}_c(f', f) = \frac{\sum_v \mathbb{I}(u_v < H)\,\|f'_v - f_v\|^2}{\sum_v \mathbb{I}(u_v < H)},$$

Yu, L et al.. "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation." MICCAI, 2019.

SSL methods using consistency regularization

Muti-view co-training

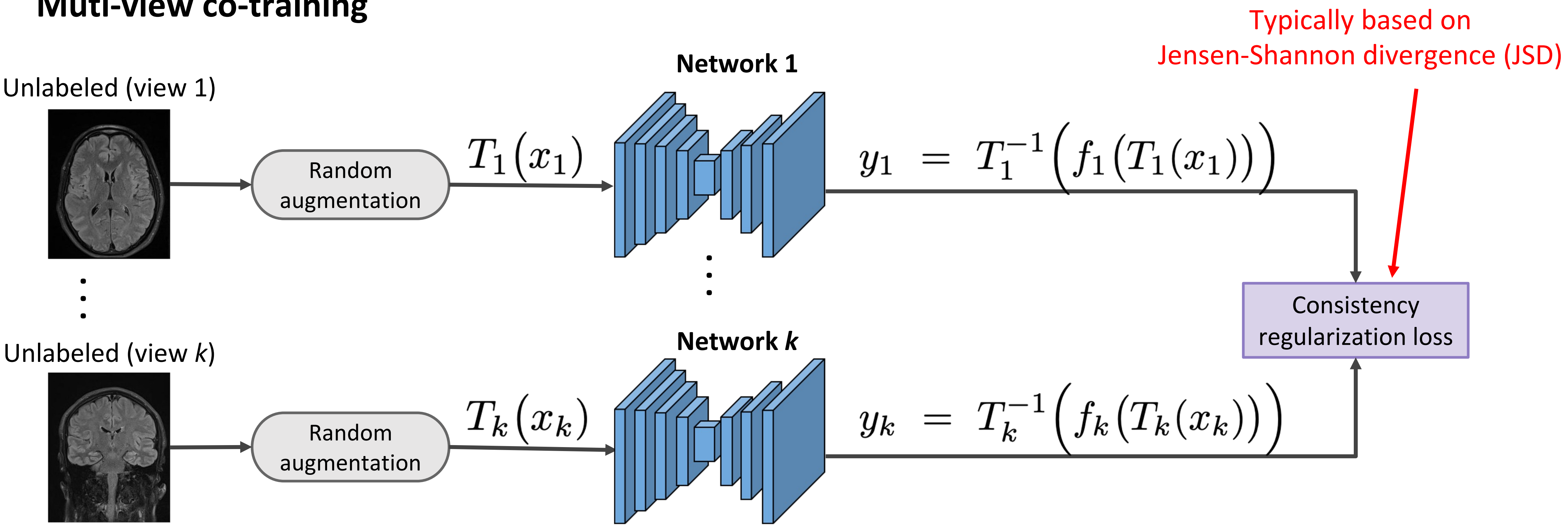


Key idea:

- Supposes the existence of separate, complementary views of the data
- Use high-confidence predictions for a given view as pseudo-labels in other views

SSL methods using consistency regularization

Muti-view co-training

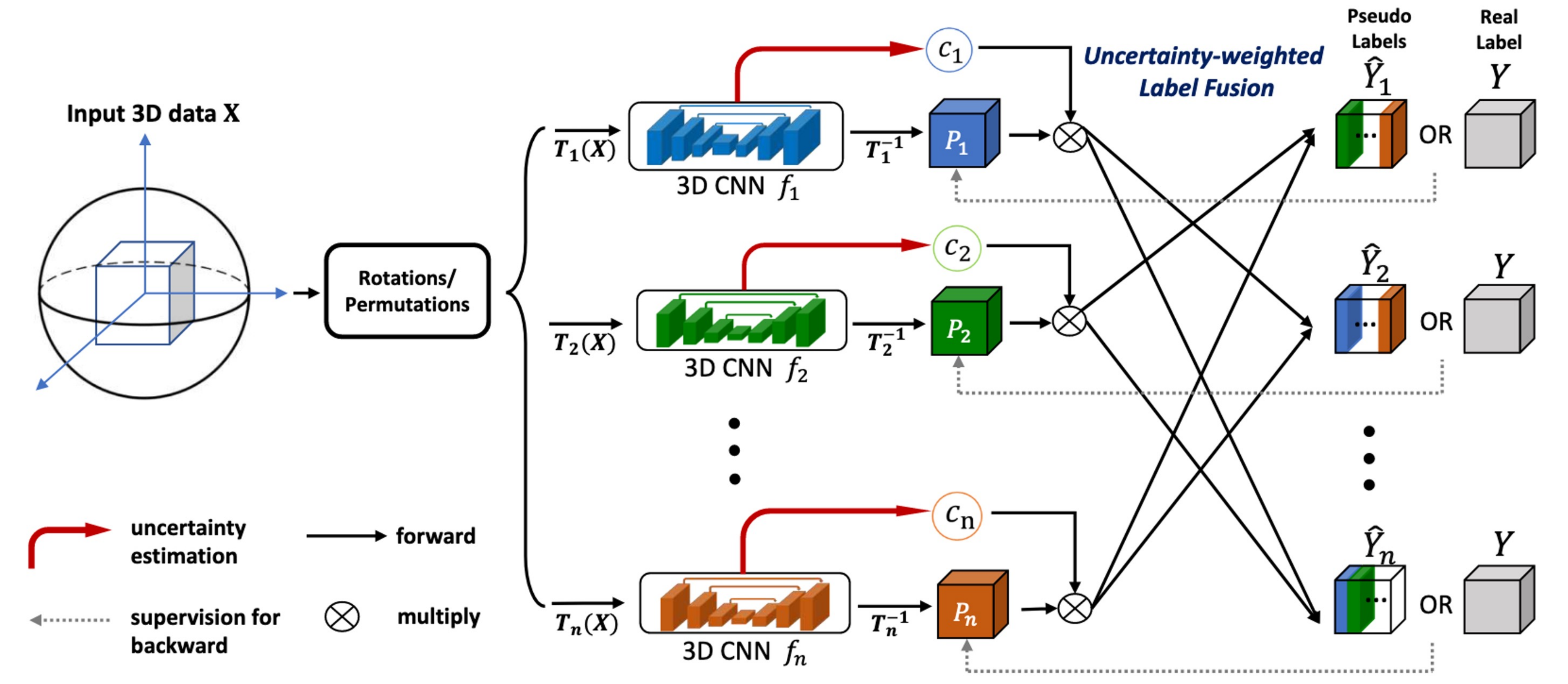


Key idea:

- Supposes the existence of separate, complementary views of the data
- Use high-confidence predictions for a given view as pseudo-labels in other views

SSL methods using consistency regularization

Application of multi-view co-training for pancreas and liver tumor segmentation



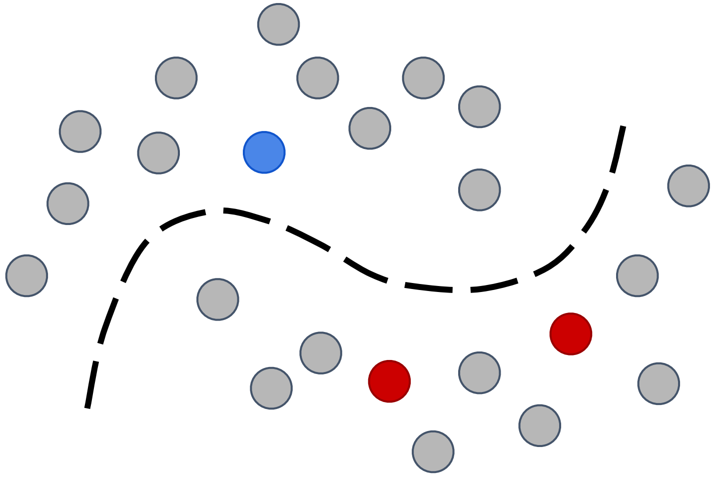
Xia, Y., et al. "3D semi-supervised learning with uncertainty-aware multi-view co-training." WACV 2020.

Unsupervised representation learning

for weakly-supervised segmentation

Unsupervised representation learning (URL)

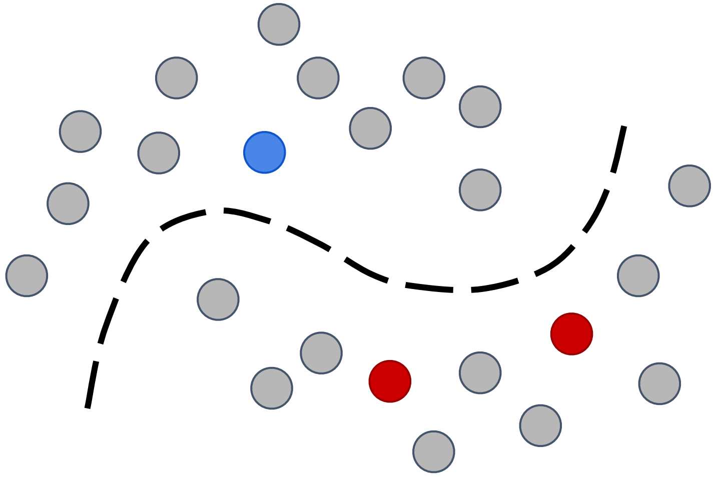
Traditional SSL



- Train a model simultaneously with both labeled and unlabeled data

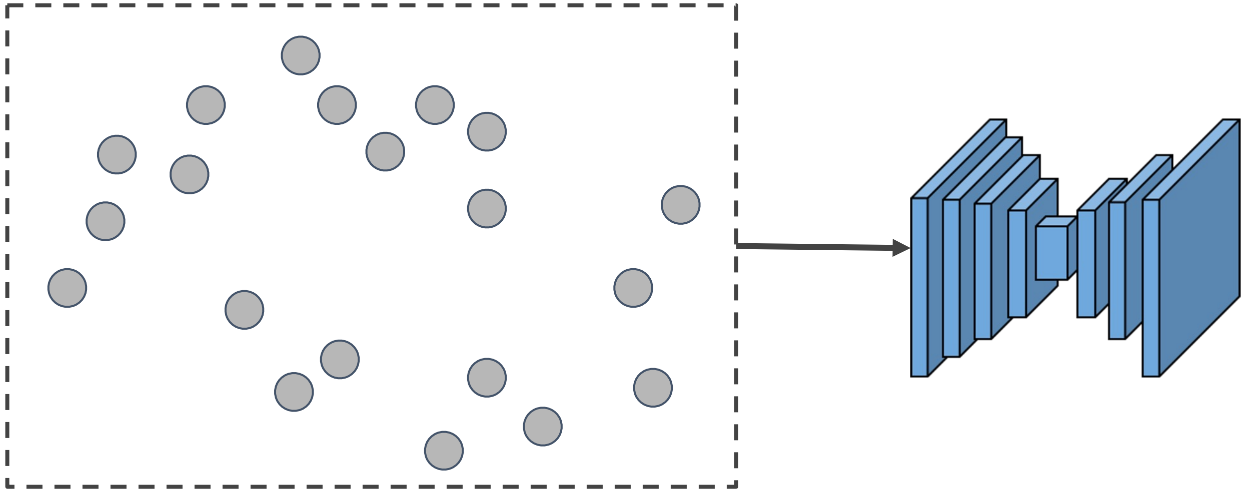
Unsupervised representation learning (URL)

Traditional SSL



- Train a model simultaneously with both labeled and unlabeled data

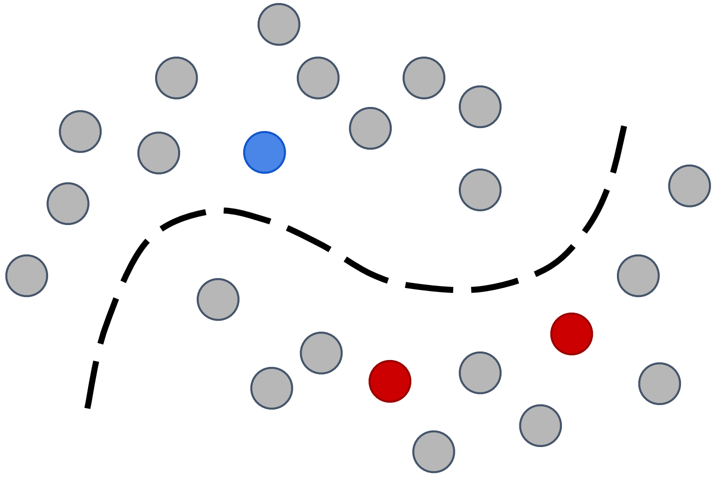
Unsupervised representation learning



- In an upstream step, use <u>only unlabeled data</u> to learn a representation useful to downstream tasks
- **Examples**:
  - Self-supervised learning
  - Contrastive learning

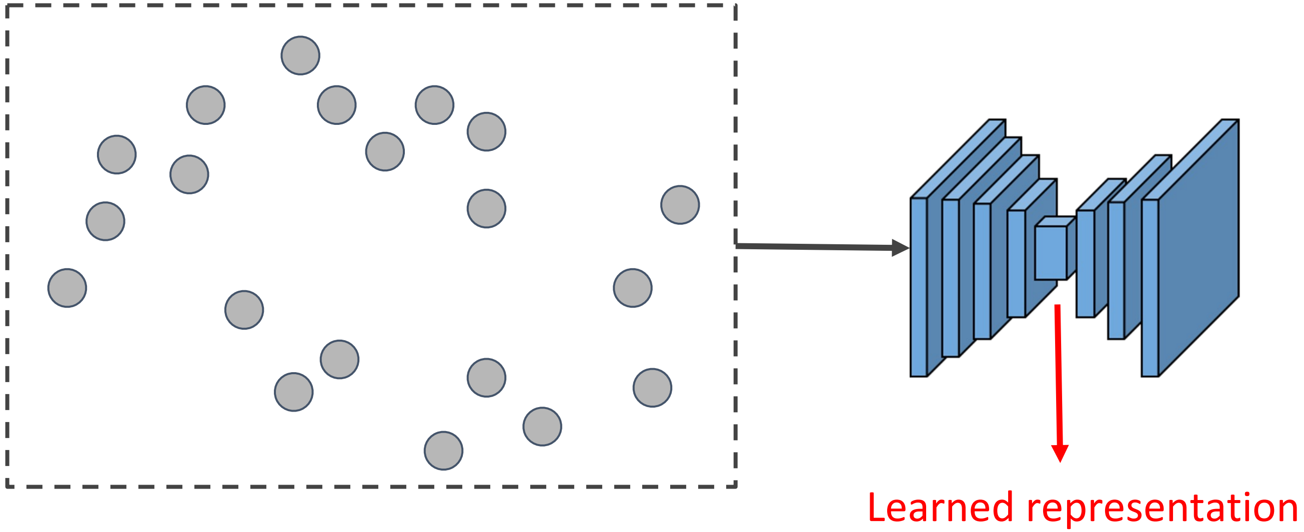
Unsupervised representation learning (URL)

Traditional SSL



- Train a model simultaneously with both labeled and unlabeled data

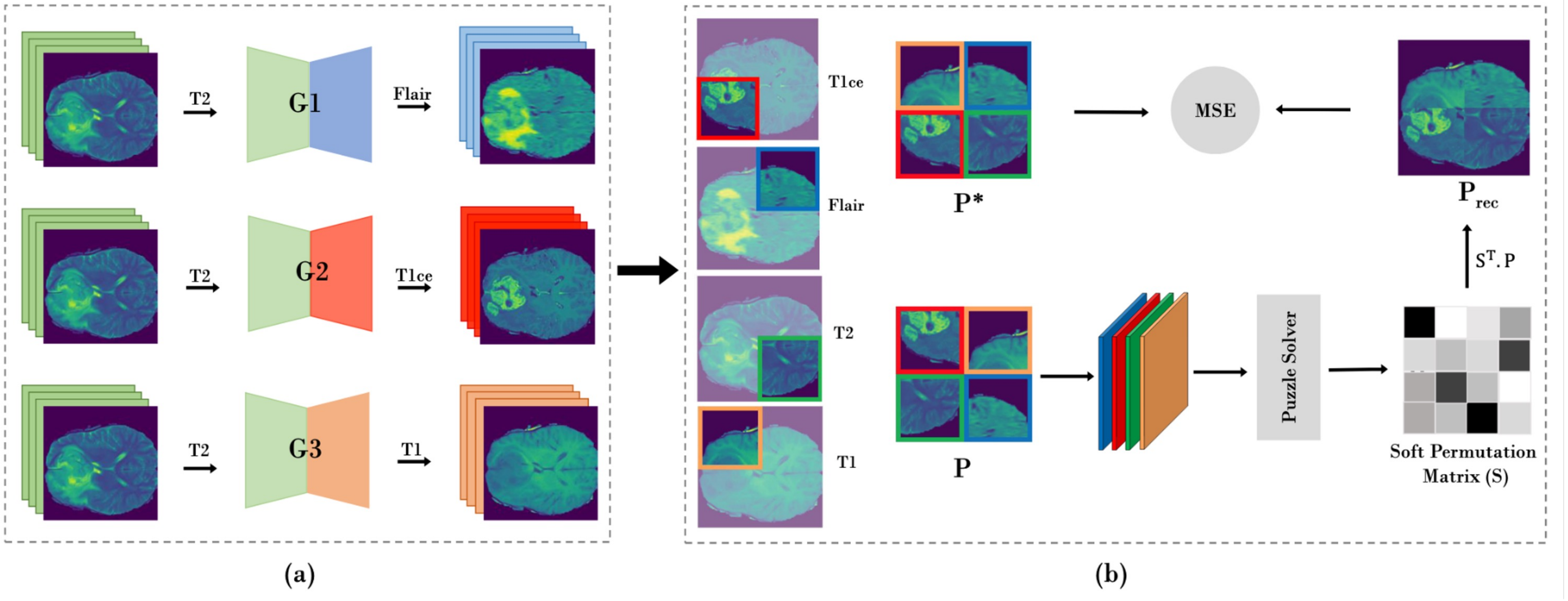
Unsupervised representation learning



- In an upstream step, use only unlabeled data to learn a representation useful to downstream tasks
- **Examples**:
  - Self-supervised learning
  - Contrastive learning

Approaches for URL

Self-supervised learning:



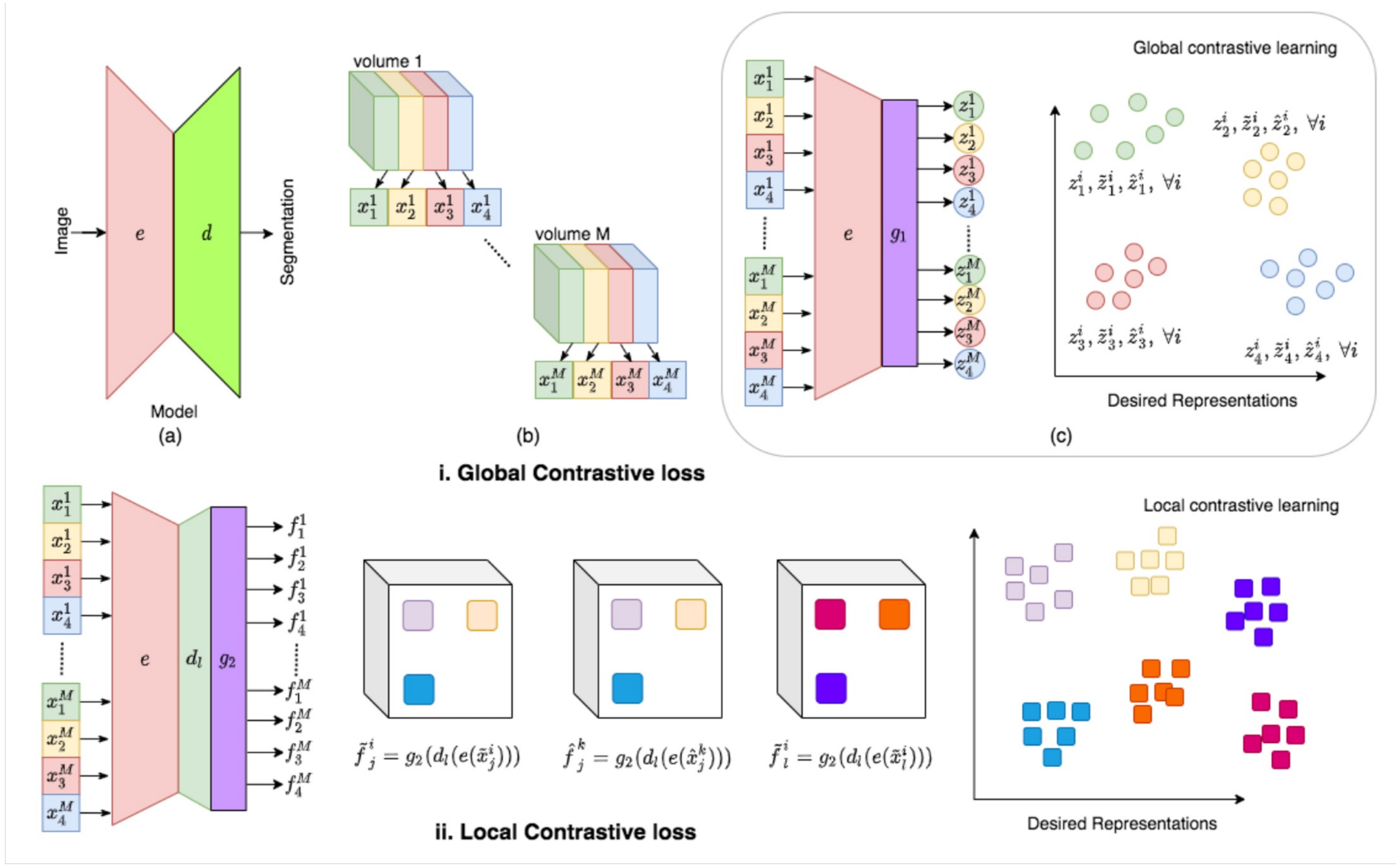
Basic idea:

- Learn to solve a pretext task which does not require annotations
- Example: find the correct order of permuted patches (*see above*)

Taleb, A., et al. "Multimodal self-supervised learning for medical image analysis." arXiv preprint (2019).

Approaches for URL

Contrastive learning:



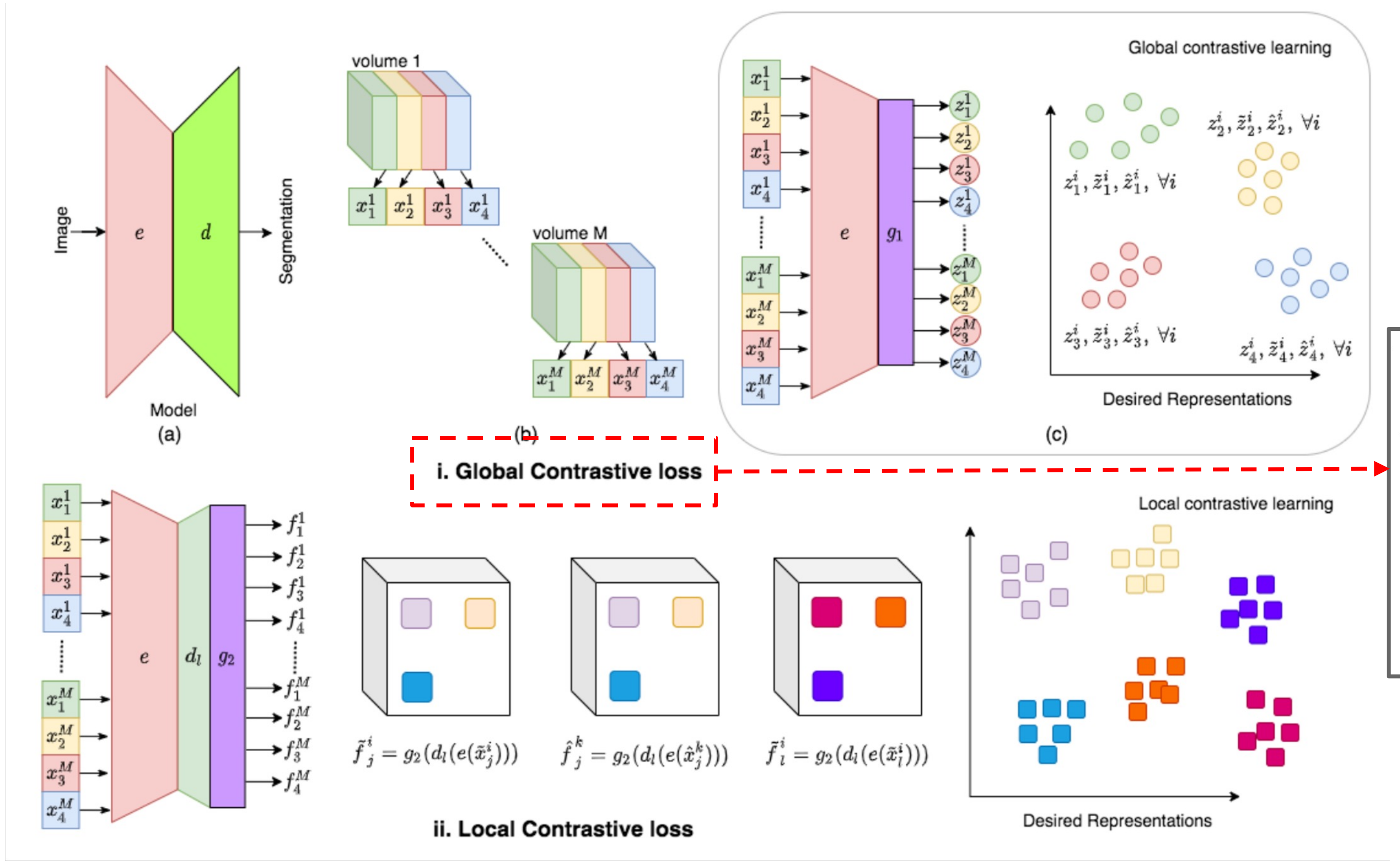
Basic idea:

- Train with pairs of images that <u>match</u> (e.g., same position in volume, same image under different transformations, etc.) or not
- Find a representation that is similar for matching pairs and different for non-matching ones

Chaitanya, K., et al. "Contrastive learning of global and local features for medical image segmentation with limited annotations." NeurIPS 2020.

Approaches for URL

Contrastive learning:



Global contrastive loss

$$-\log \frac{e^{\mathrm{sim}(\tilde{z},\hat{z})/\tau}}{e^{\mathrm{sim}(\tilde{z},\hat{z})/\tau} + \sum_{\bar{x}\in\Lambda^-} e^{\mathrm{sim}(\tilde{z},g_1(e(\bar{x})))/\tau}}$$

$$\tilde{z} = g_1(e(\tilde{x})),\ \hat{z} = g_1(e(\hat{x})).$$

Matching pair

Non-matching pairs

Basic idea:

- Train with pairs of images that <u>match</u> (e.g., same position in volume, same image under different transformations, etc.) or not
- Find a representation that is similar for matching pairs and different for non-matching ones

Chaitanya, K., et al. "Contrastive learning of global and local features for medical image segmentation with limited annotations." NeurIPS 2020.

Lyon, Avril 17– 21

DLMI2023

CREATIS

UNIVERSITÉ DE LYON

ÉTS
Le génie pour l'industrie

Thank you

Any questions ?