Harnessing Out-Of-Distribution Examples via Augmenting Content and Style

Zhuo Huang[1] Xiaobo Xia[1] Li Shen[2] Bo Han[3]
Mingming Gong[4] Chen Gong[5] Tongliang Liu[1]

[1]Sydney AI Centre, The University of Sydney [2]JD Explore Academy
[3] Hong Kong Baptist University [4] The University of Melbourne
[5]Jiangsu Key Lab of Image and Video Understanding for Social Security, School of Computer Science and Engineering, Nanjing University of Science and Technology

ICLR 2023

Motivation

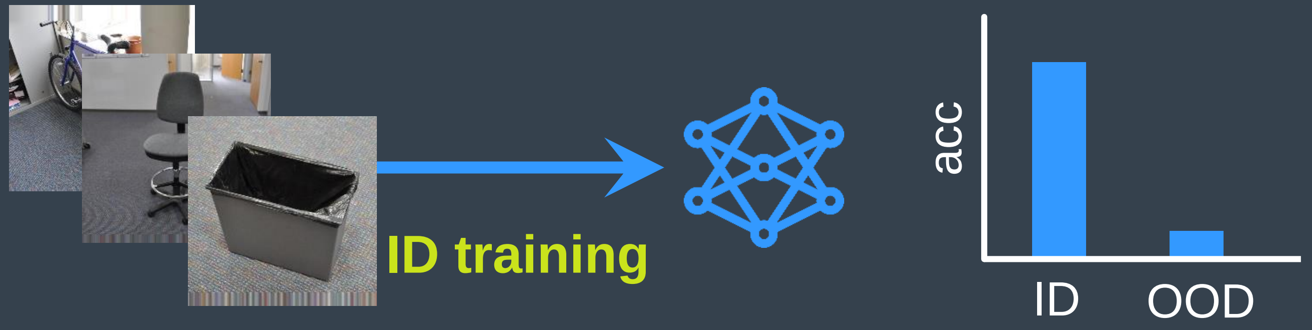
- When learning models meet Out-Of-Distribution (OOD) data, what happens?

Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?

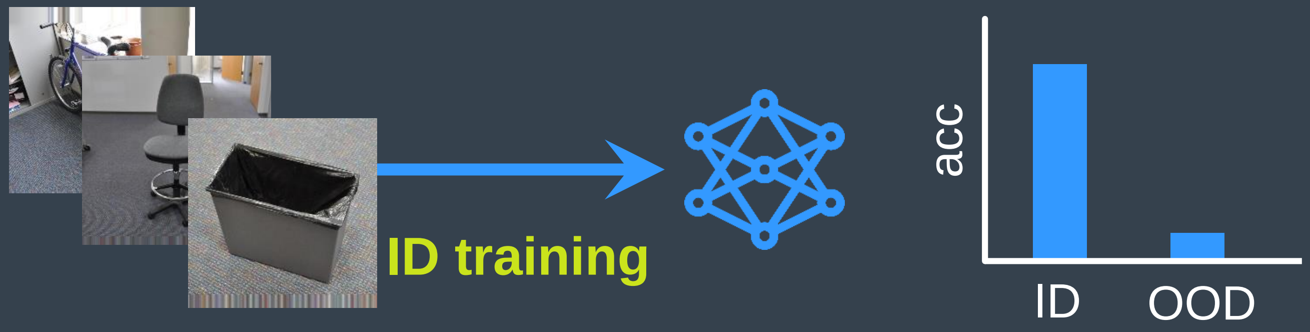
Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



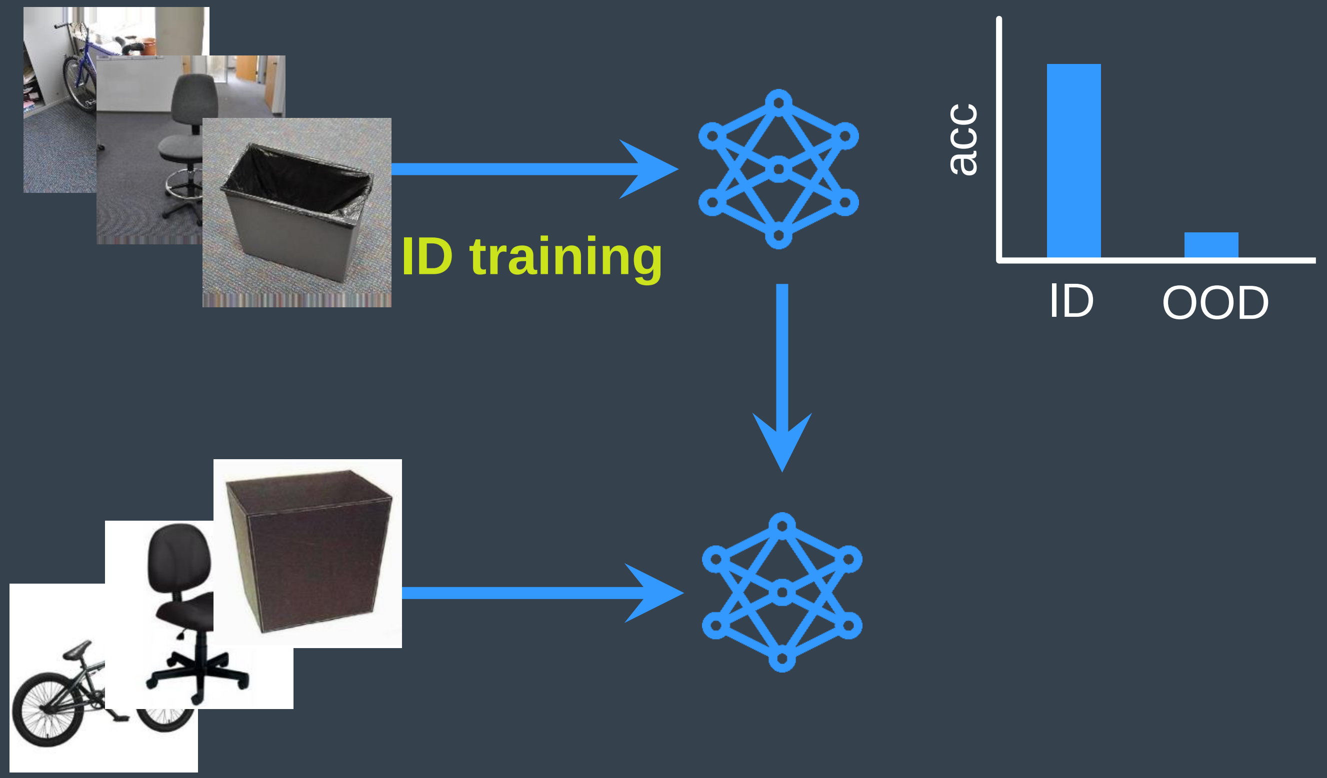
Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



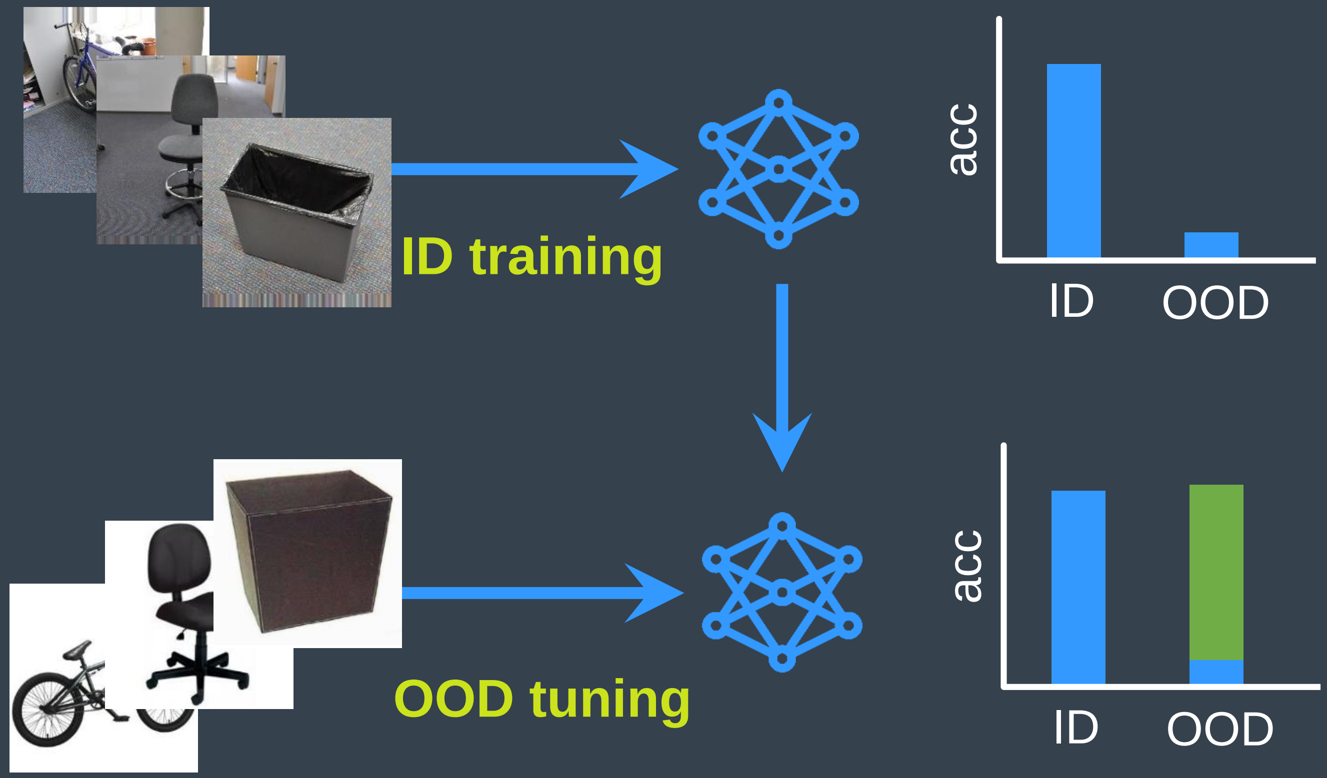
Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



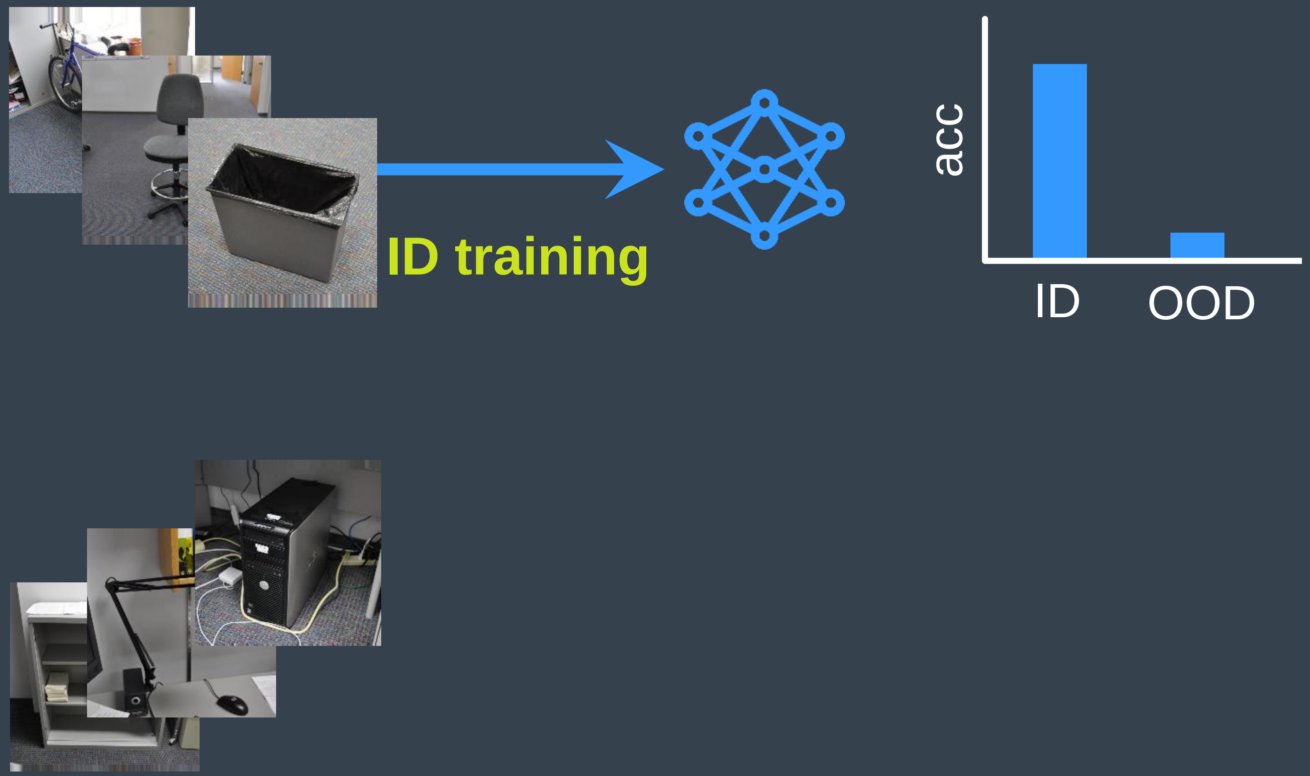
Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



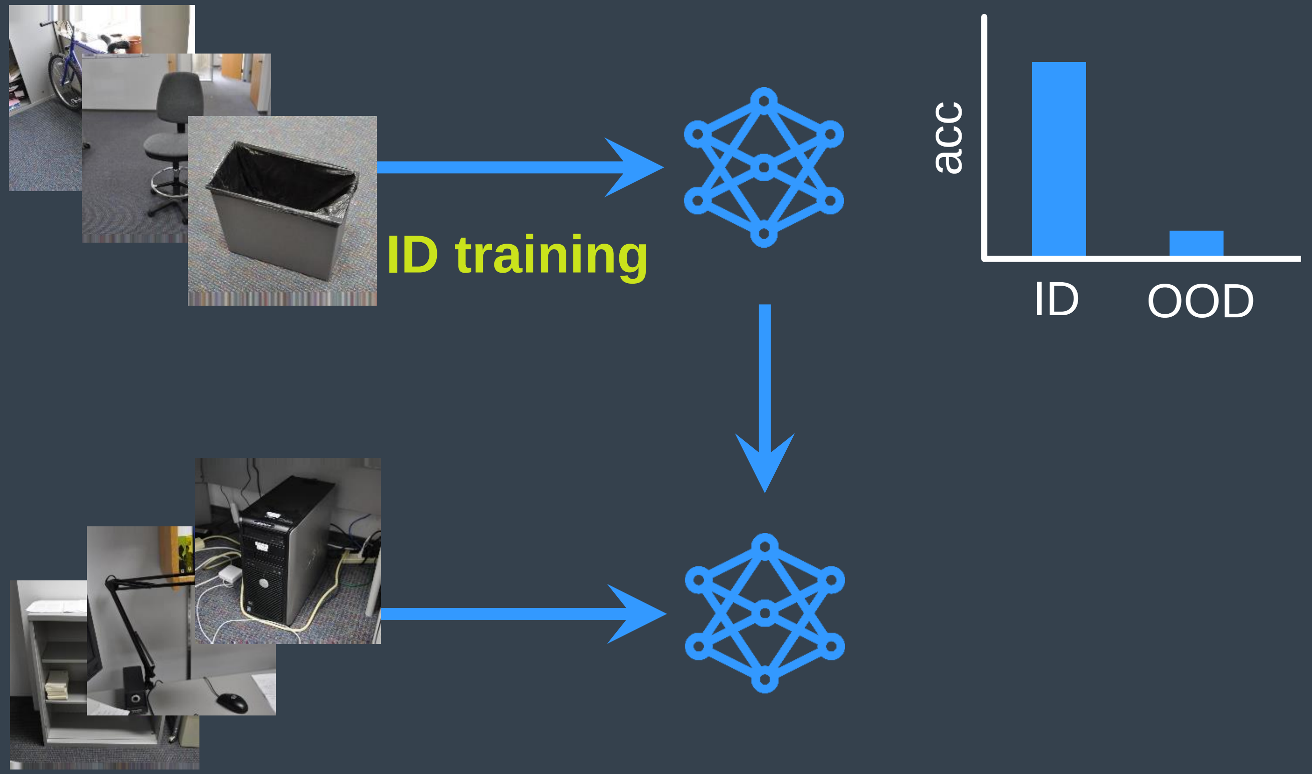
Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



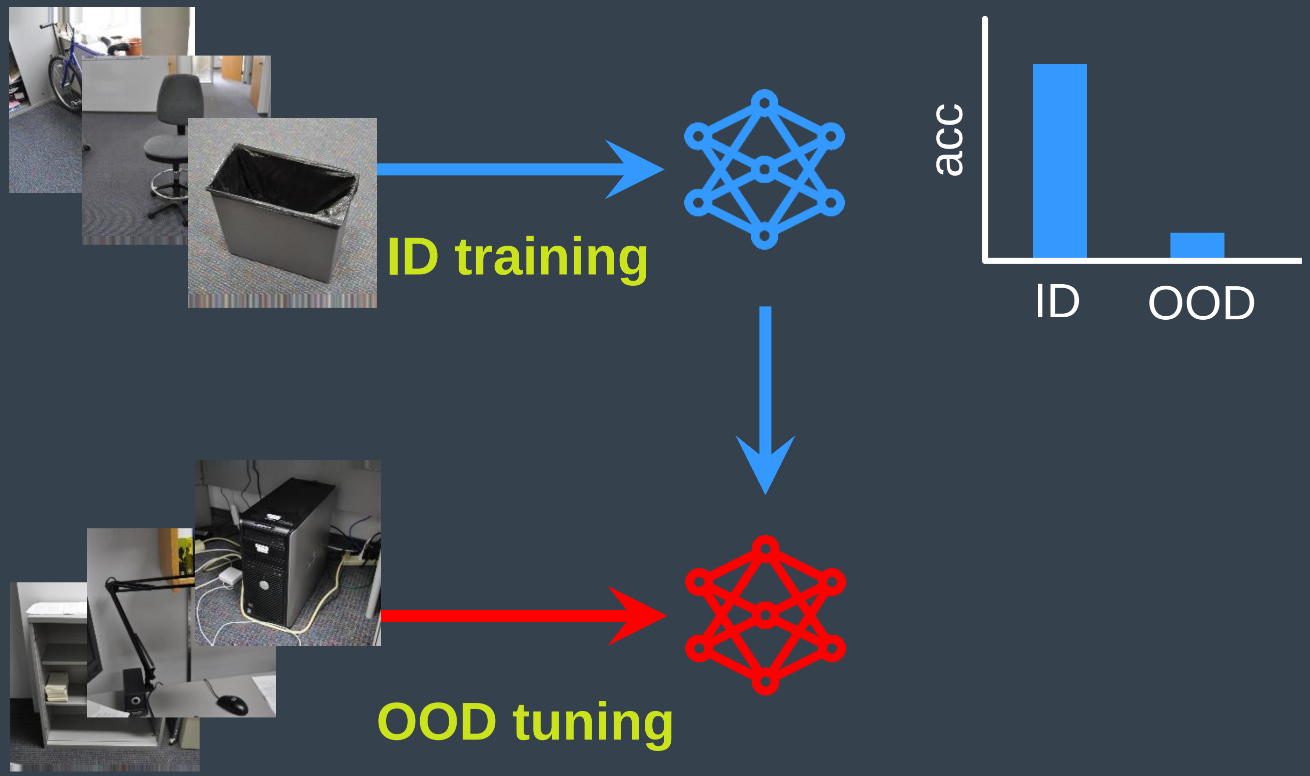
Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



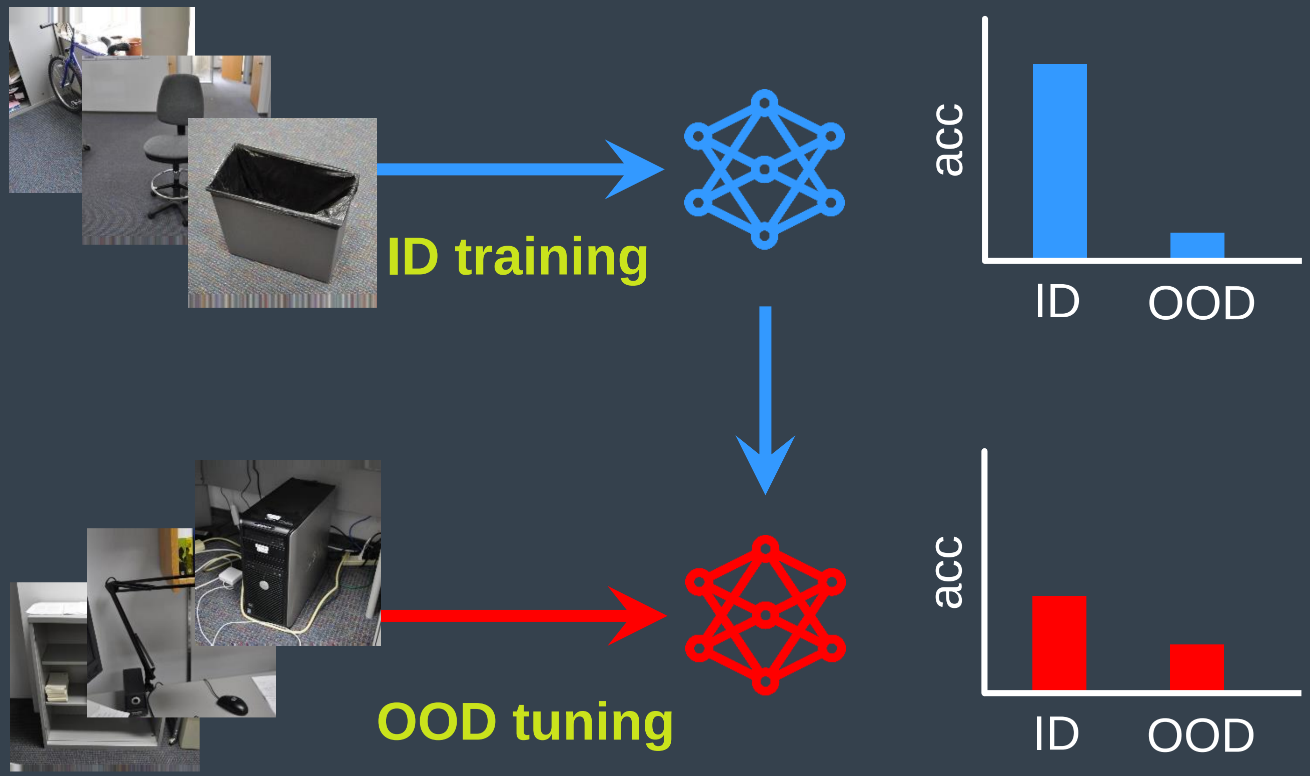
Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



Motivation

- When learning models meet Out-Of-Distribution (OOD) data, what happens?



Motivation

- Question:

Why different types of OOD data have different effects?

Motivation

- A causal perspective:

An (in-distribution) ID image

Motivation

- A causal perspective:

An (in-distribution) ID image

Motivation

- A causal perspective:

An (in-distribution) ID image

content

style

Motivation

- How to harness OOD data?

style \ content	known	unknown
known	ID data	
unknown		

Motivation

- How to harness OOD data?

style \ content	known	unknown
known	ID data	Malign data
unknown		Malign data

Motivation

- How to harness OOD data?

style \ content	known	unknown
known	ID data	Malign data
unknown	Benign data	Malign data

Motivation

• How to harness OOD data?

style \ content	known	unknown
known	ID data	Malign data
unknown	Benign data	Malign data

Malign data → harmful → OOD Detection

Motivation

- How to harness OOD data?

style \ content	known	unknown
known	ID data	Malign data
unknown	Benign data	Malign data

Malign data → harmful → OOD Detection

Benign data → helpful → OOD Generalization

Problems

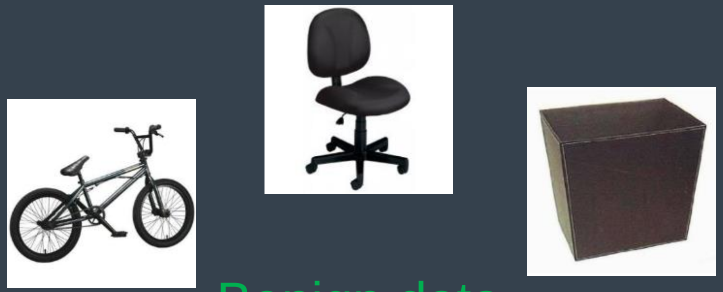
- The difficulty of distinguishing benign data and malign data:

Problems

- The difficulty of distinguishing benign data and malign data:

Problems

- The difficulty of distinguishing benign data and malign data:



Benign data

Problems

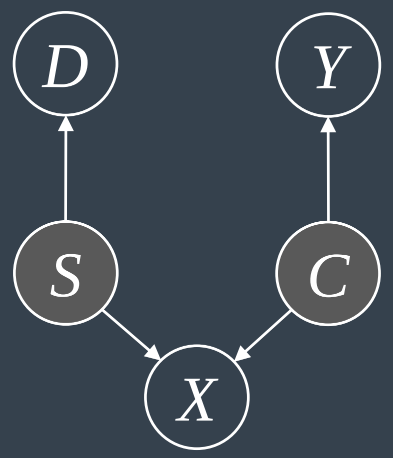
- The difficulty of distinguishing benign data and malign data:

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:

Problems

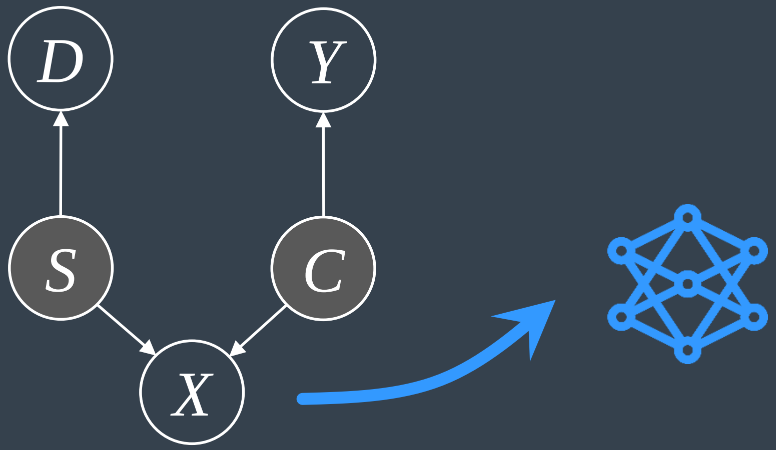
- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:



(1)

Problems

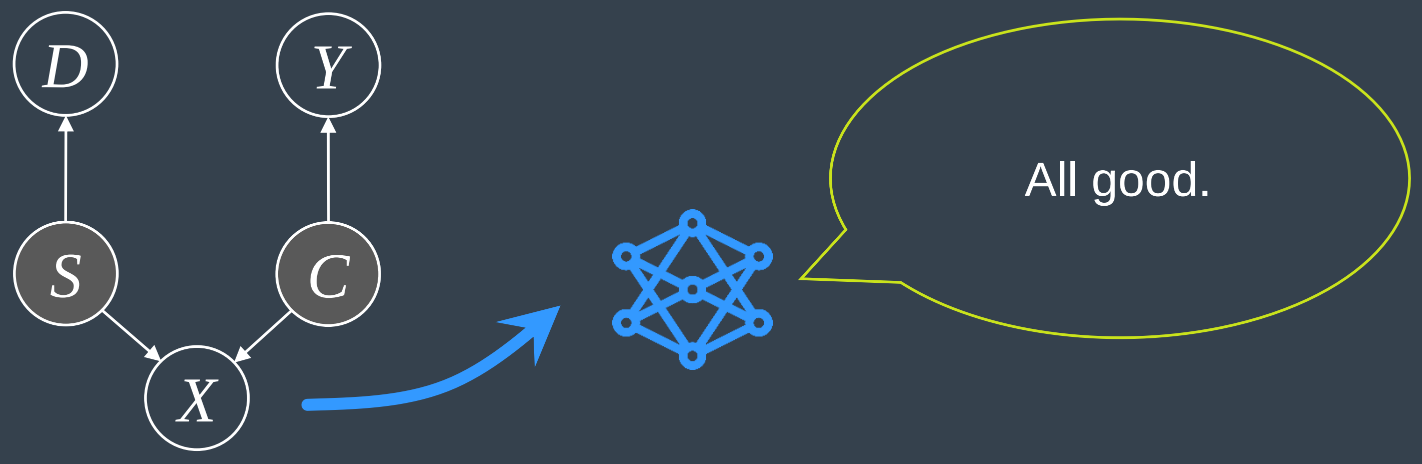
- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:



(1)

Problems

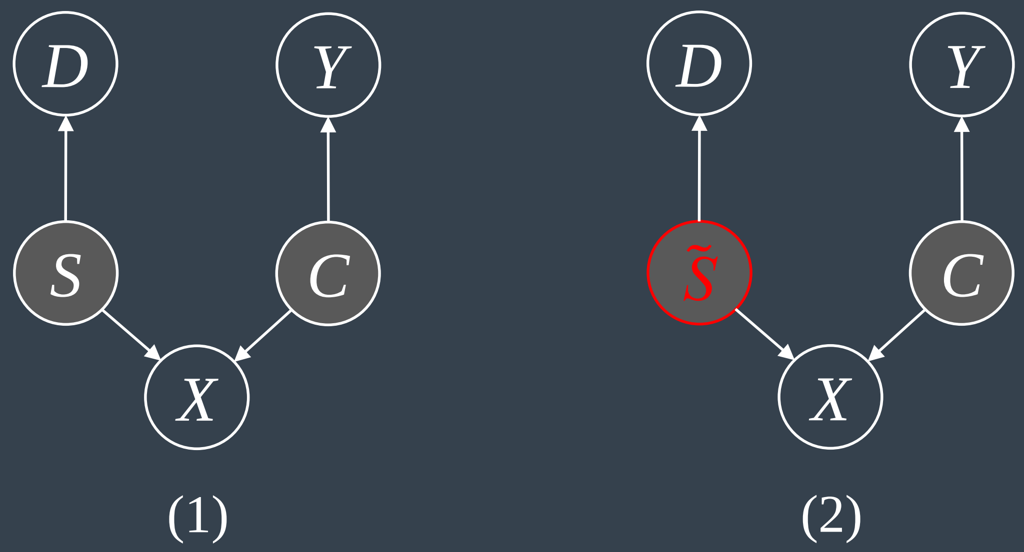
- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:



(1)

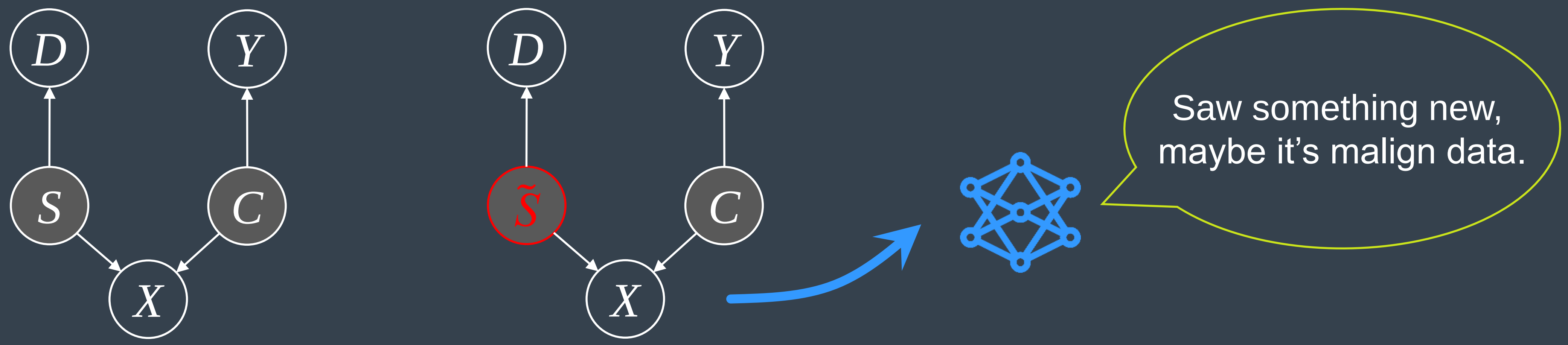
Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:



Problems

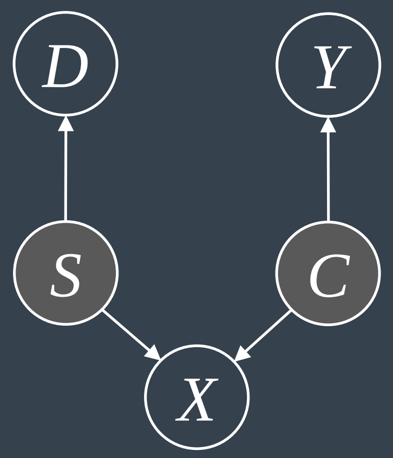
- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:



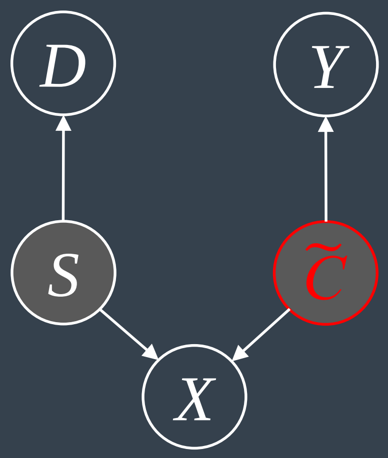
(1) (2)

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:



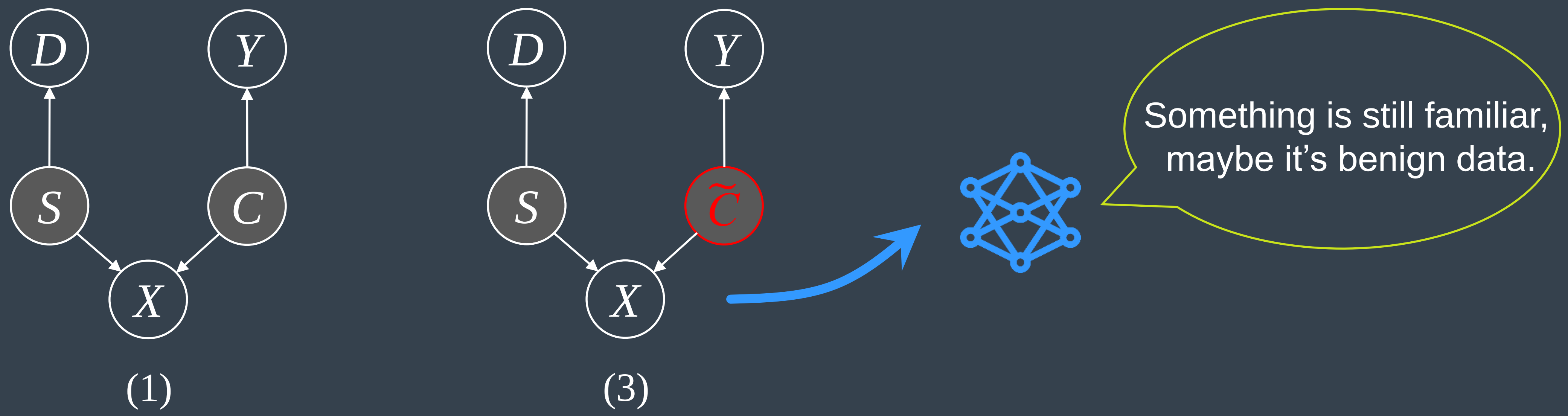
(1)



(3)

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style:



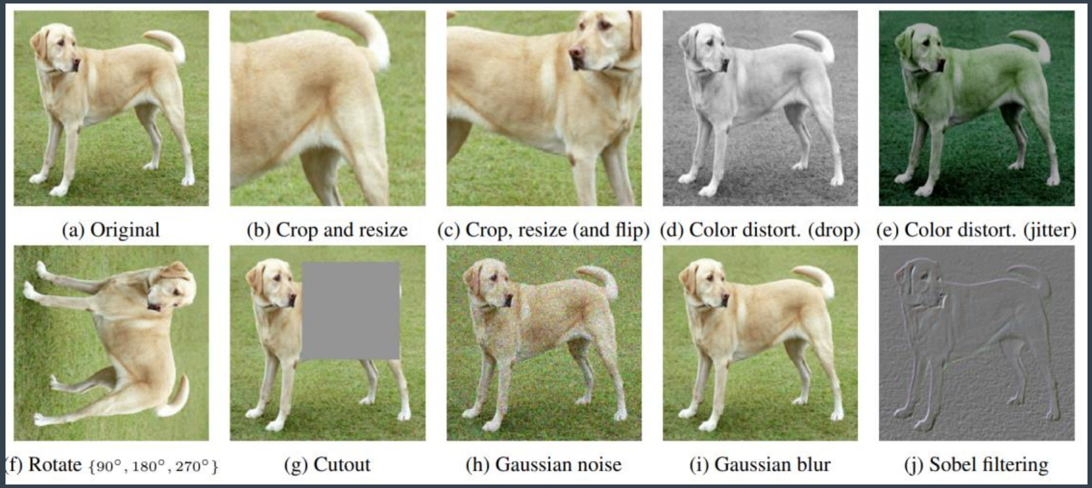
Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

Data augmentation improves generalization performance



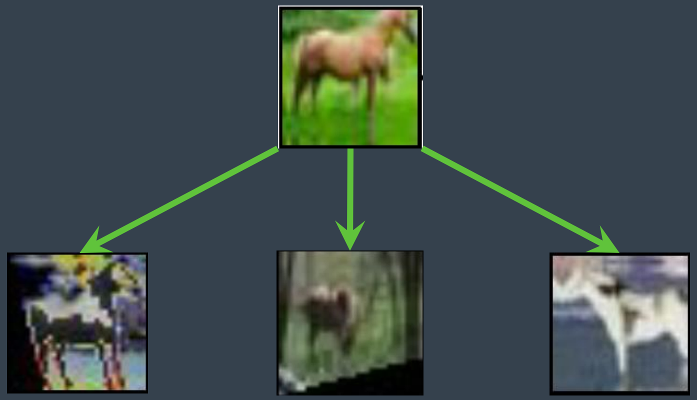
benign data

Chen et al., A Simple Framework for Contrastive Learning of Visual Representations

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

But, data augmentation cause content damage:

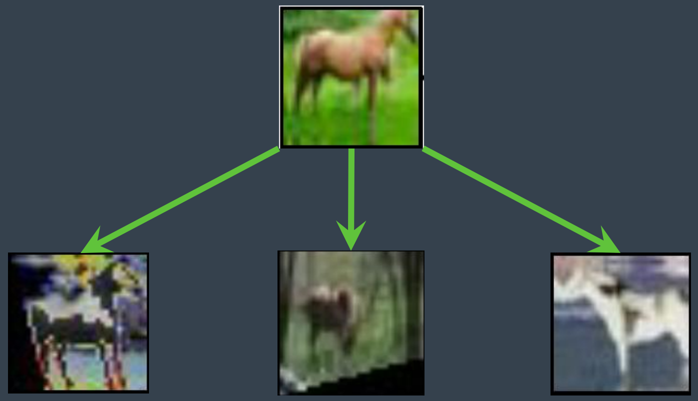


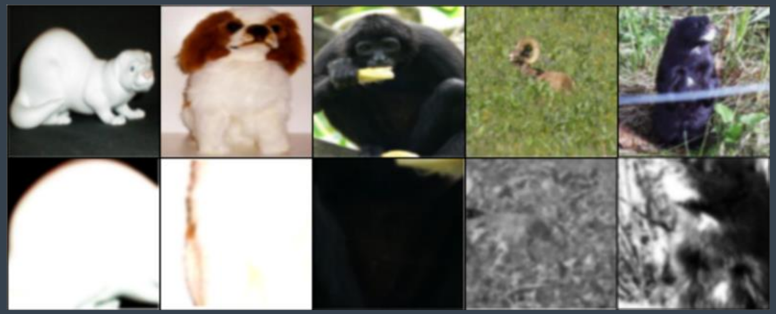
Sohn et al., FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  But, data augmentation cause content damage:



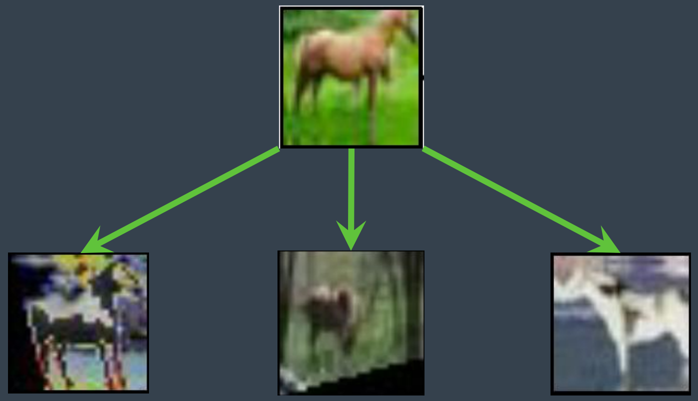


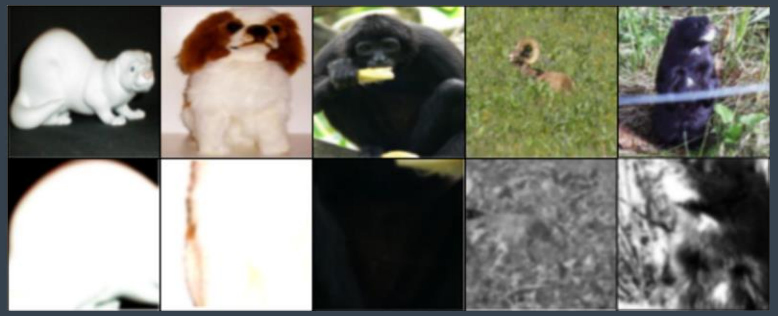
Wu et al., Unsupervised feature learning via nonparametric instance discrimination

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  So, benign data shall be more carefully produced.





Wu et al., Unsupervised feature learning via nonparametric instance discrimination

Problems

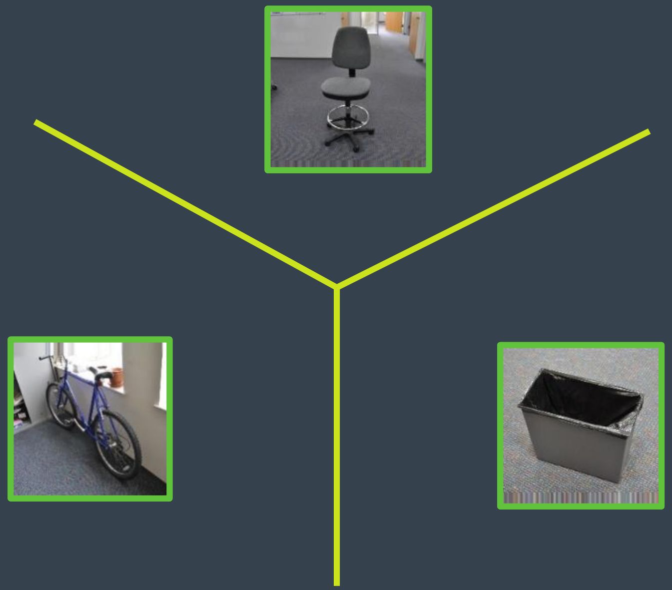
- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  Is malign data totally useless?

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  Is malign data totally useless?



Problems

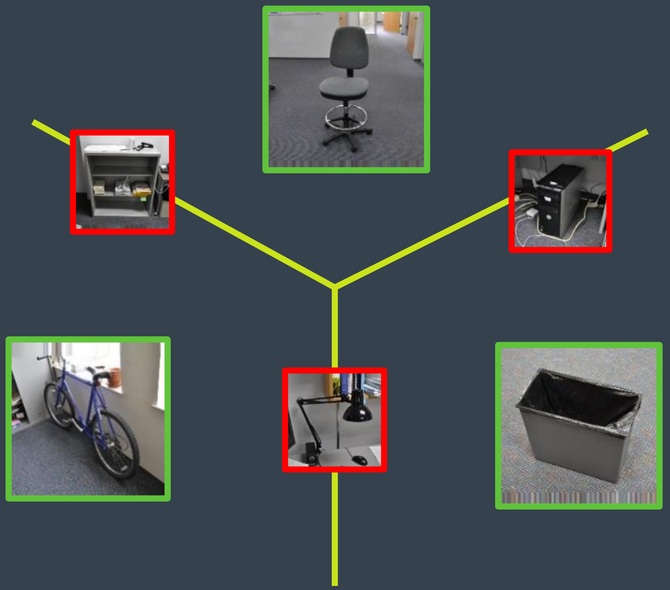
- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  Is malign data totally useless?

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  Is malign data totally useless?



Problems

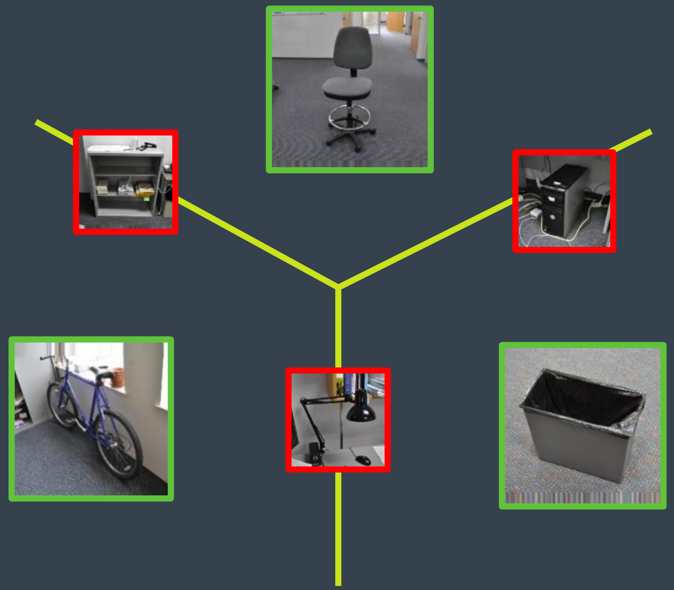
- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  Is malign data totally useless?

Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

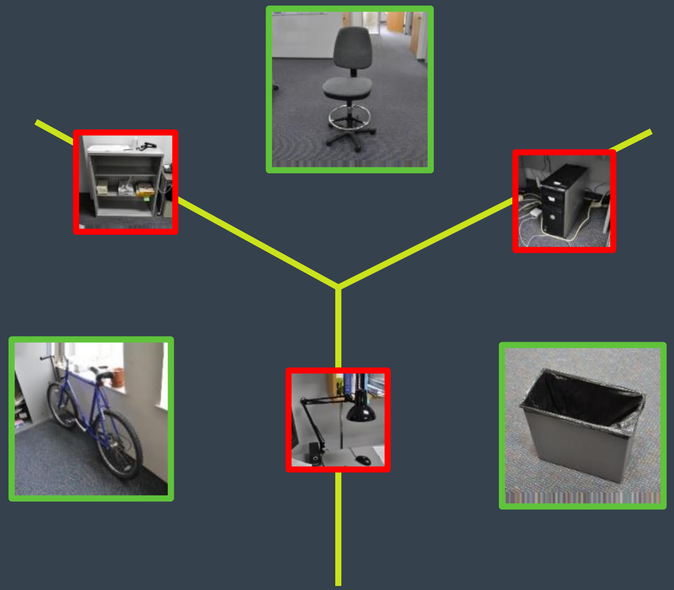
  Is malign data totally useless?



Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

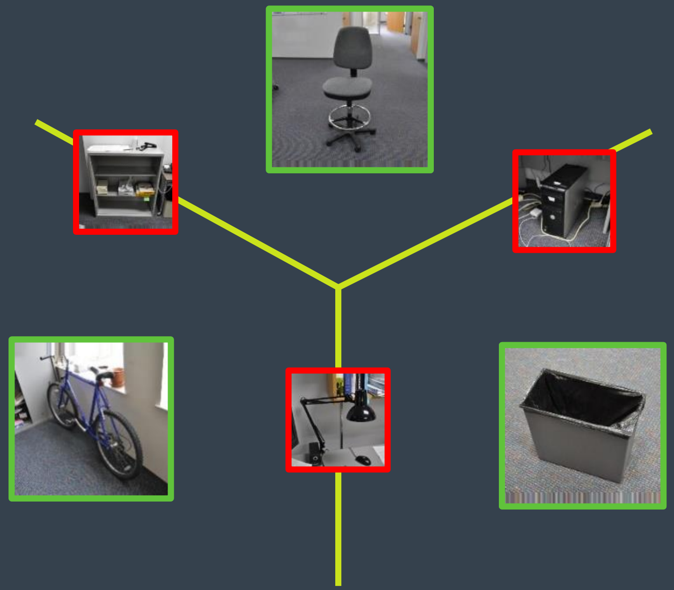
  Is malign data totally useless?

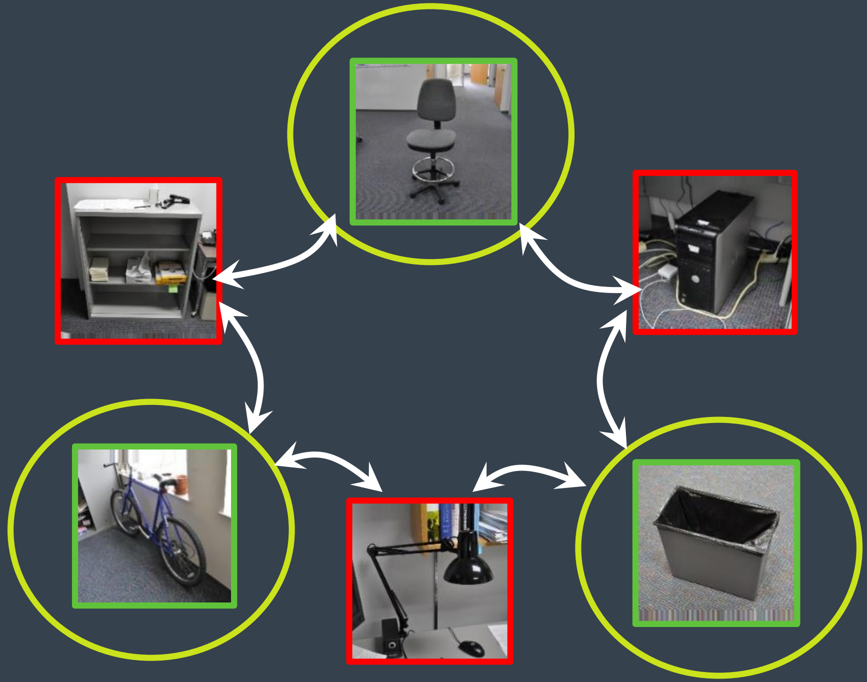


Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

  Is malign data totally useless?

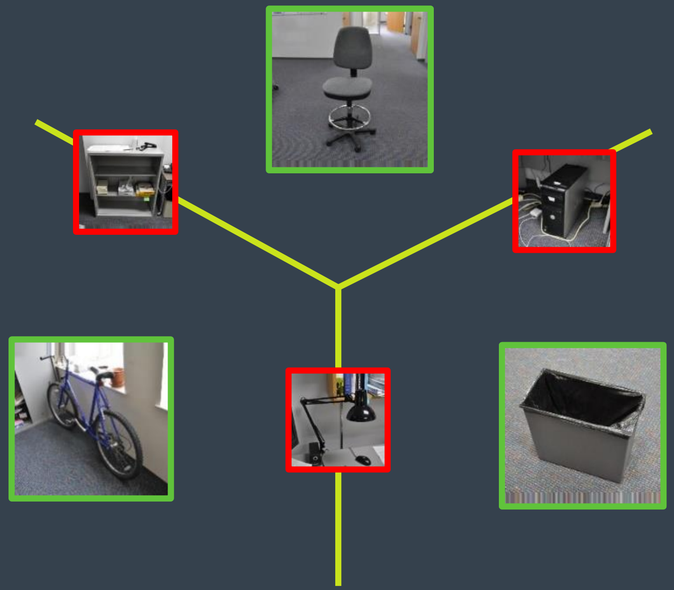


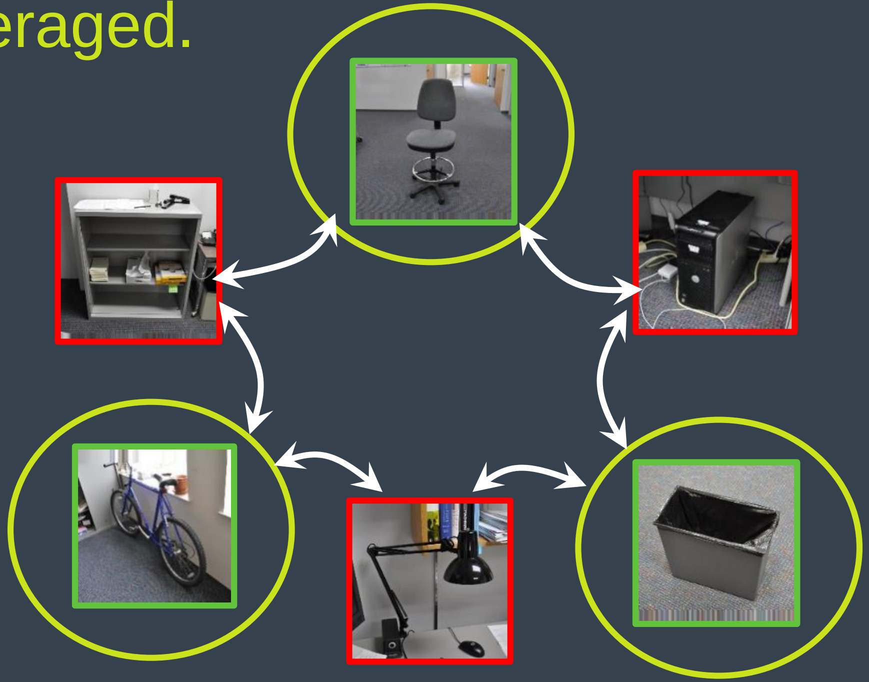


Problems

- The difficulty of distinguishing benign data and malign data.
- The entanglement of content and style.
- The utilization of benign data and malign data:

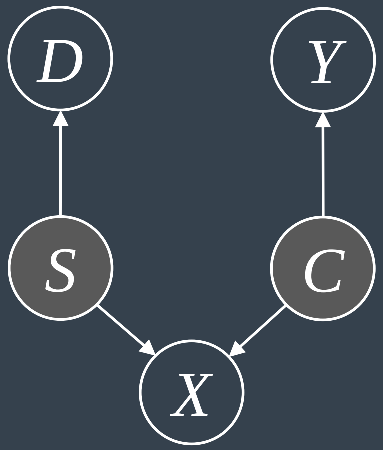
  So, malign data shall be properly leveraged.





Solution

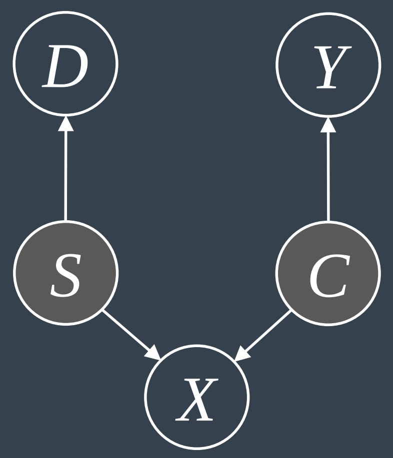
- Modeling the data generating process:



(1)

Solution

- Modeling the data generating process:

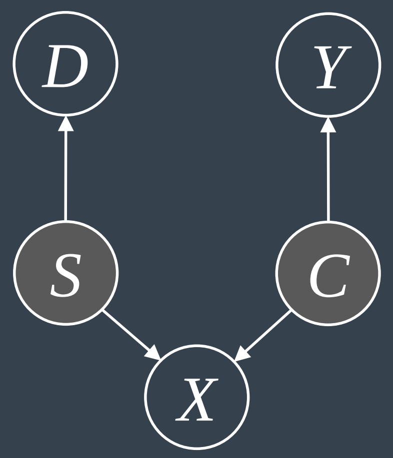


(1)

- ➢ Capture content feature using label information;
- ➢ Capture style feature using domain information;

Solution

- Modeling the data generating process:



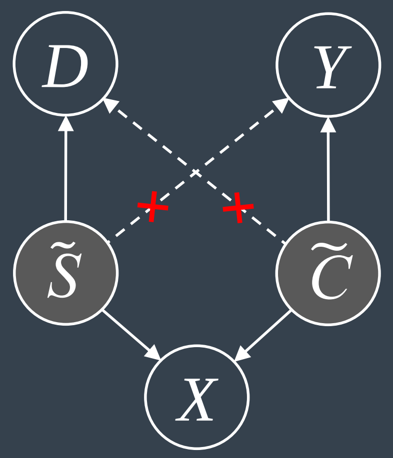
(1)

- ➢ Capture content feature using label information;
- ➢ Capture style feature using domain information;

$$P(X,Y,D,C,S) = P(C,S)P(Y,D \mid C,S)P(X \mid C,S)$$

Solution

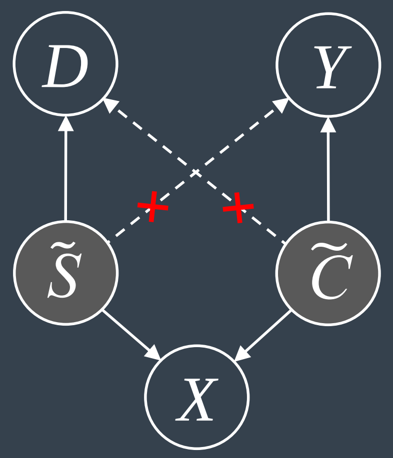
- Disentanglement of content and style:



(2)

Solution

- Disentanglement of content and style:

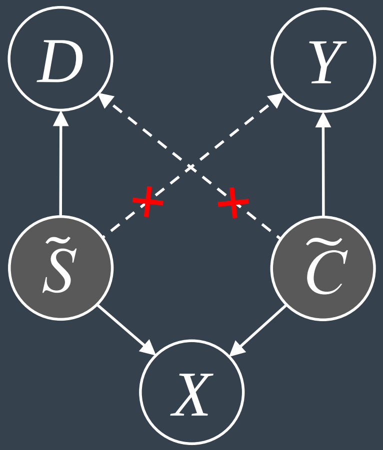


(2)

➢ Breaking unwanted paths: $\tilde{C} \to D, \tilde{S} \to Y$

Solution

- Disentanglement of content and style:



(2)

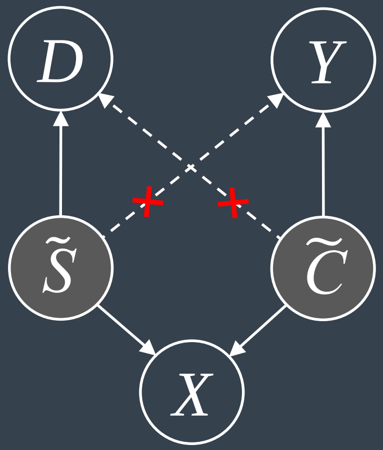
➢ Breaking unwanted paths: $\tilde{C} \to D, \tilde{S} \to Y$

$$\tilde{P}(X, Y, D, C, S) := \frac{P(C)P(S)P(Y \mid C)P(D \mid S)P(X \mid C, S)}{q_{\phi_s}(D \mid C)q_{\phi_c}(Y \mid S)}$$

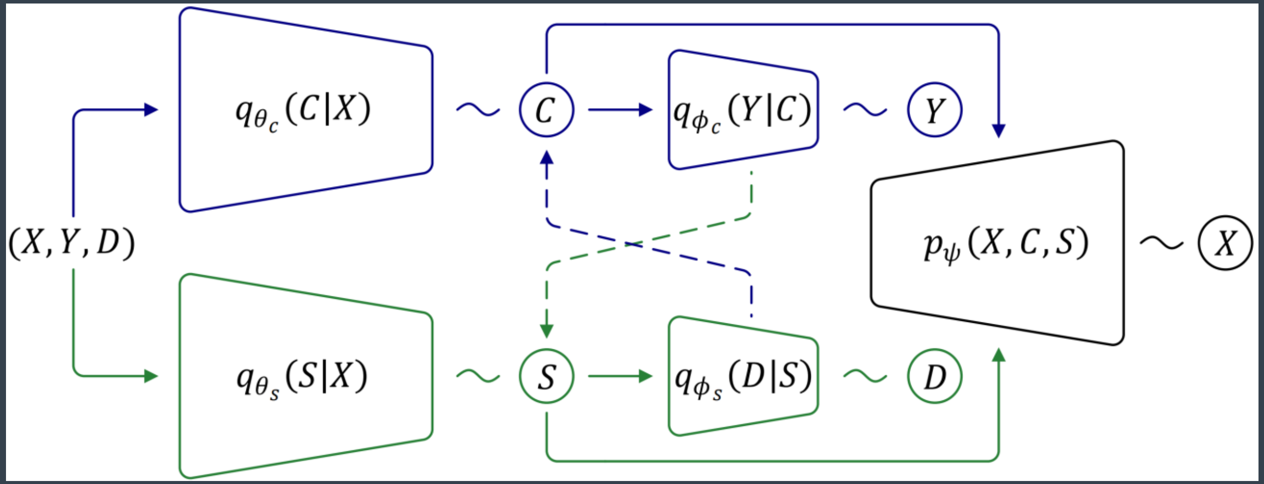
Solution

- Variational framework:

$$\tilde{P}(X, Y, D, C, S) := \frac{P(C)P(S)P(Y \mid C)P(D \mid S)P(X \mid C, S)}{q_{\phi_s}(D \mid C)q_{\phi_c}(Y \mid S)}$$

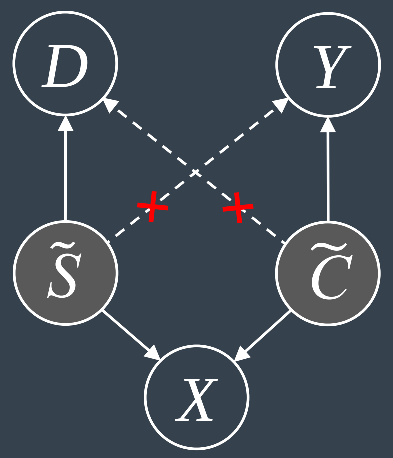


(2)



Solution

- Evidence Lower-Bound:

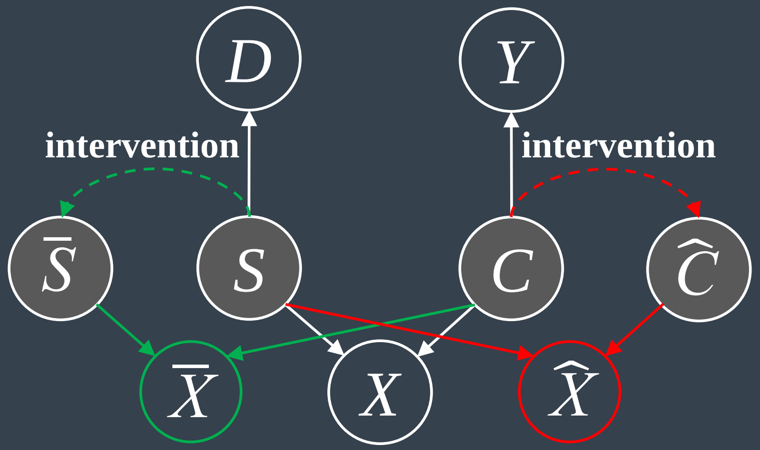


(2)

$$\widetilde{ELBO}(\mathbf{x}, y, d) =$$
$$-KL\left(q_{\theta_c}(c \mid \mathbf{x}) \| p(C)\right) - KL\left(q_{\theta_s}(s \mid \mathbf{x}) \| p(S)\right)$$
$$+\mathbb{E}_{c \sim q_{\theta_c}(c|\mathbf{x})}\left[\log q_{\phi_c}(y \mid c) - \log q_{\phi_s}(d \mid c)\right]$$
$$+\mathbb{E}_{s \sim q_{\theta_s}(s|\mathbf{x})}\left[\log q_{\phi_s}(d \mid s) - \log q_{\phi_c}(y \mid s)\right]$$
$$+\mathbb{E}_{(c,s) \sim q_{\theta}(c,s|\mathbf{x})}\left[\log q_{\psi}(\mathbf{x} \mid c, s)\right]$$

Solution

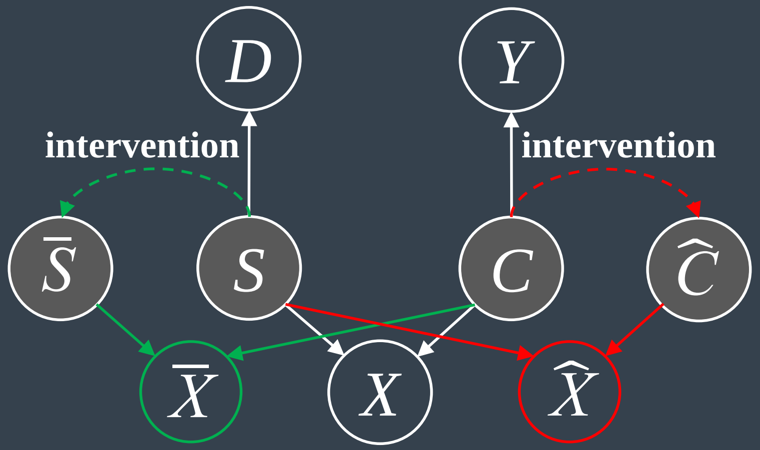
- Augmenting content and style:



(3)

Solution

- Augmenting content and style:



(3)

$q_{\phi_c}(\mathrm{Y} \mid \mathrm{C})$

$q_{\phi_s}(\mathrm{D} \mid \mathrm{S})$

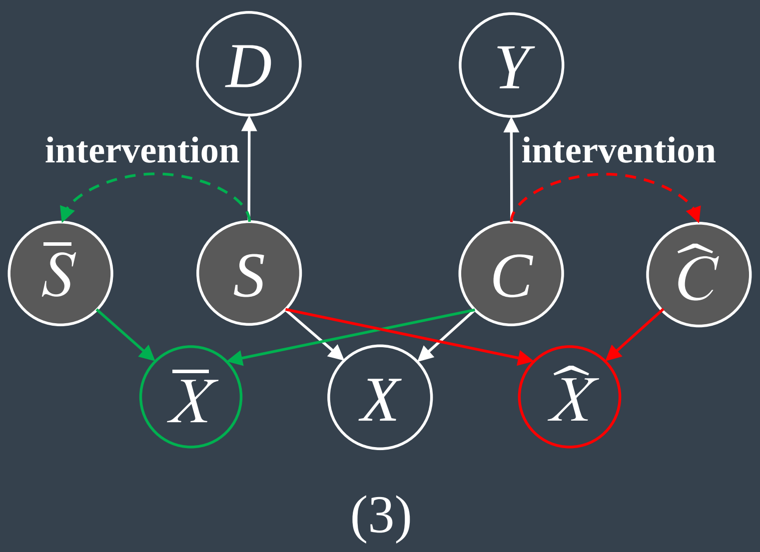
$q_{\theta_c}(\mathrm{C} \mid \mathrm{X})$

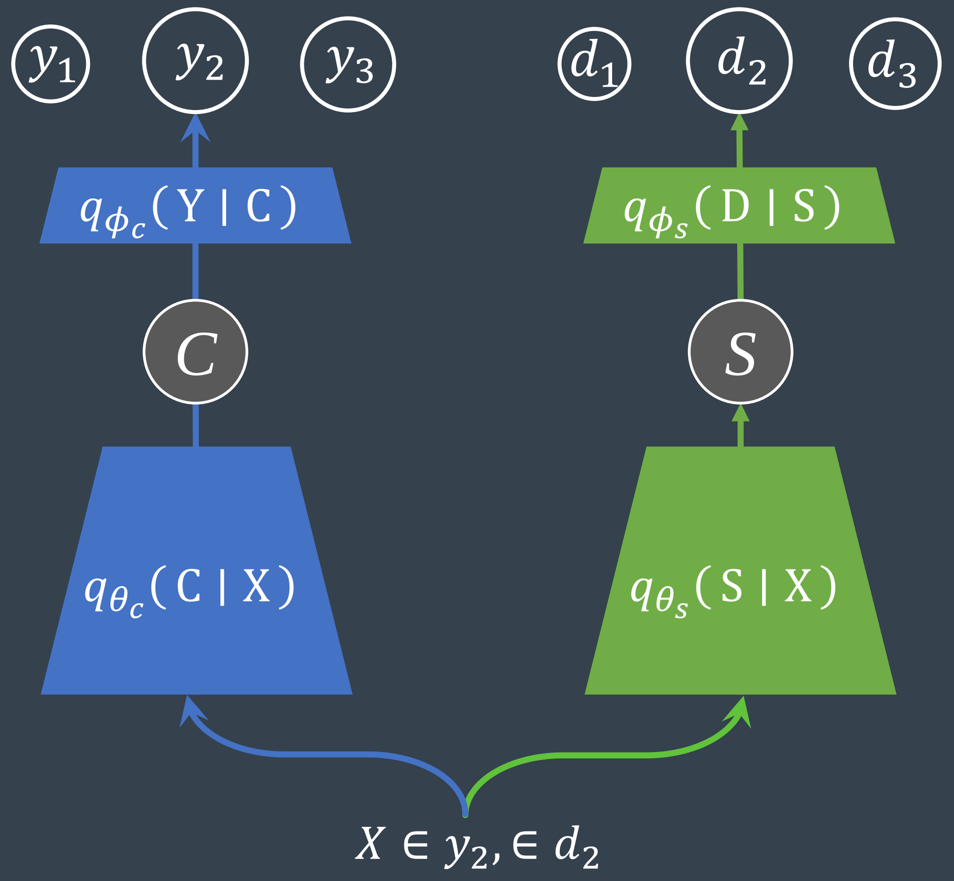
$q_{\theta_s}(\mathrm{S} \mid \mathrm{X})$

$X \in y_2, \in d_2$

Solution

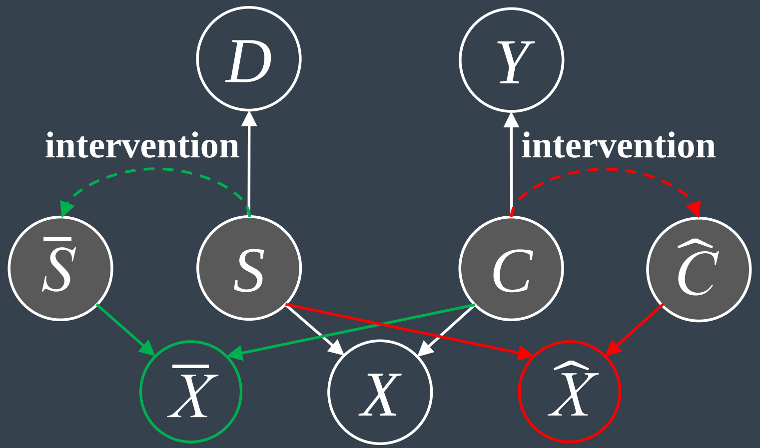
- Augmenting content and style:



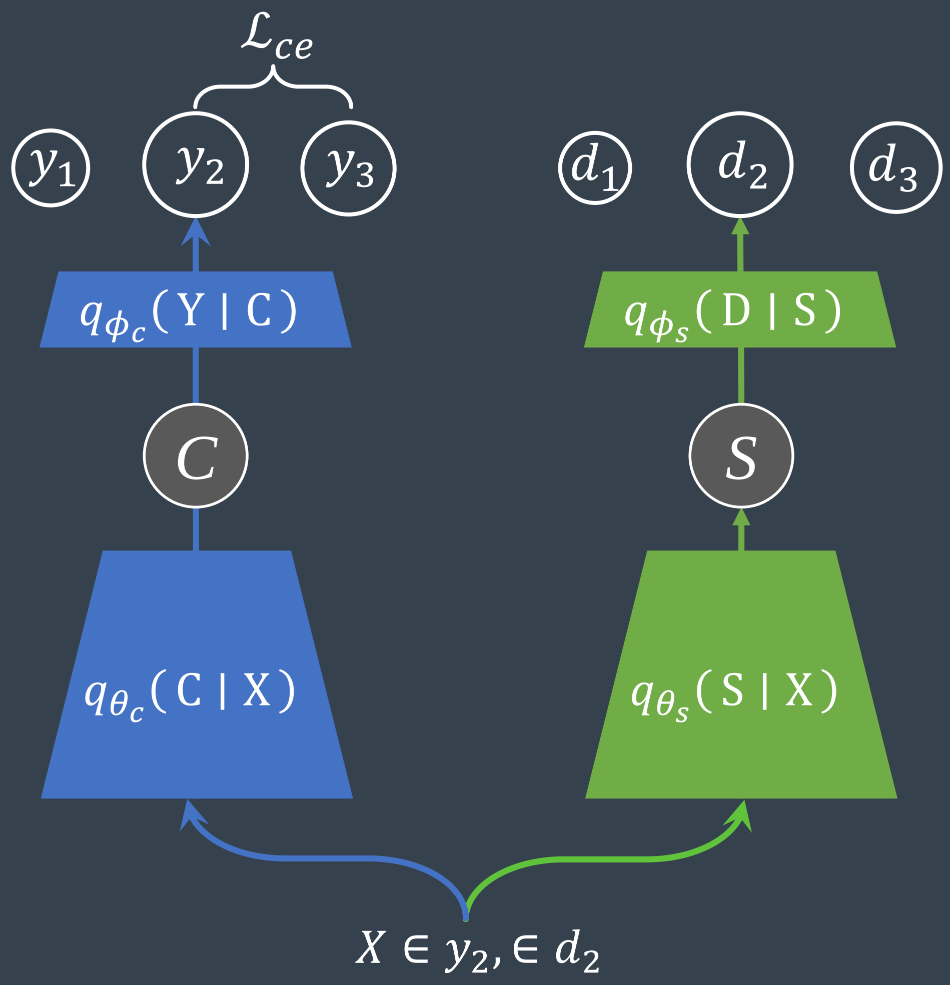


Solution

- Augmenting content and style:

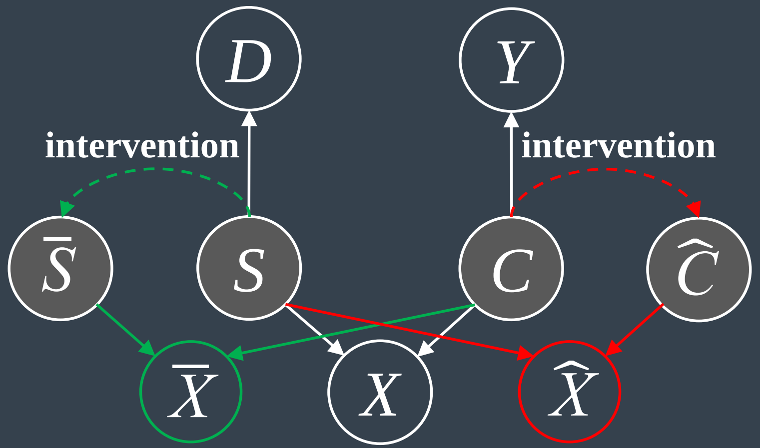


(3)

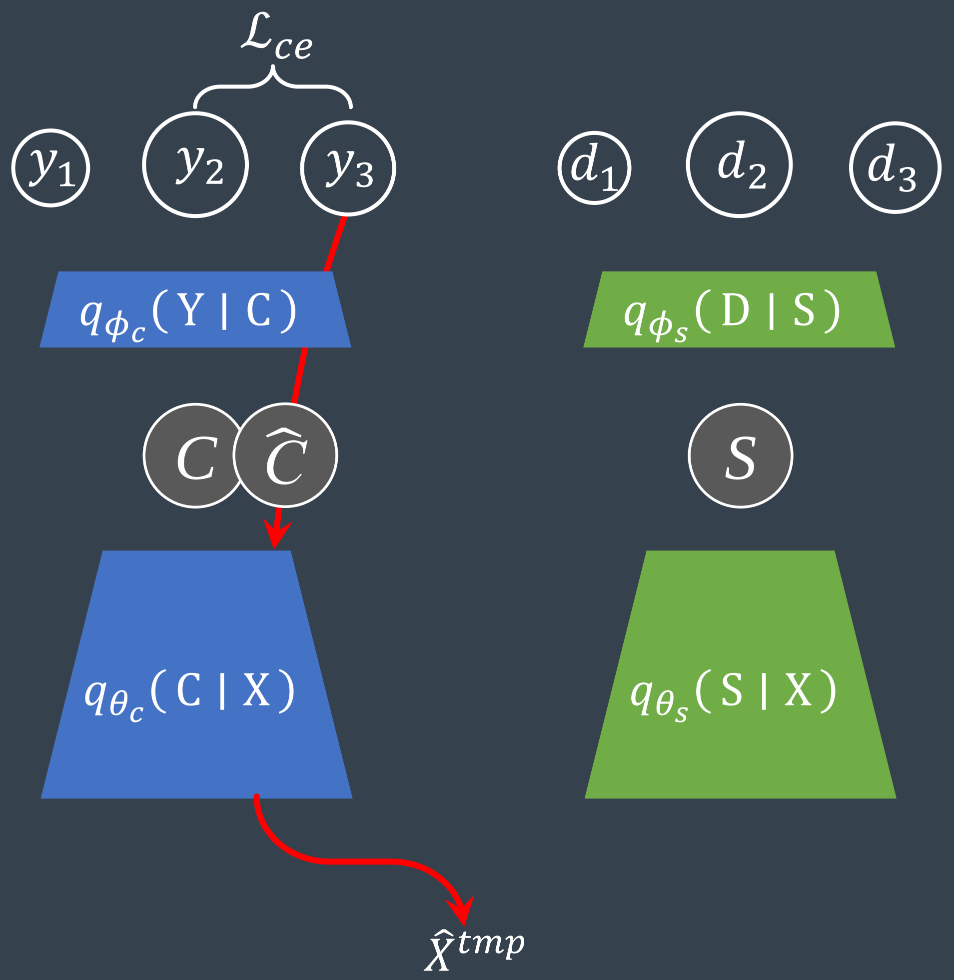


Solution

- Augmenting content and style:

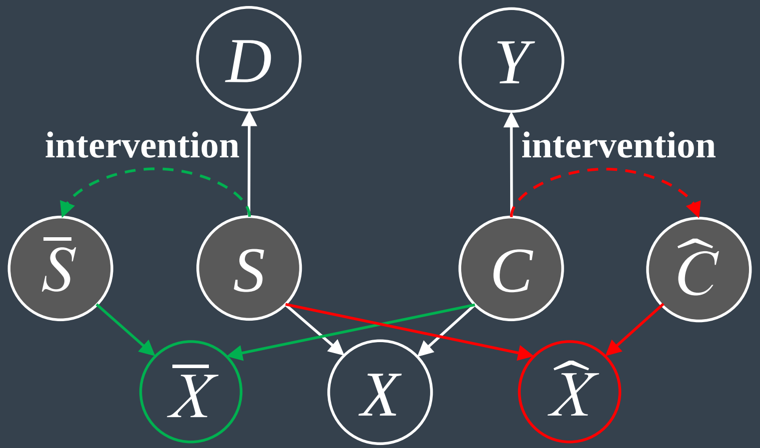


(3)

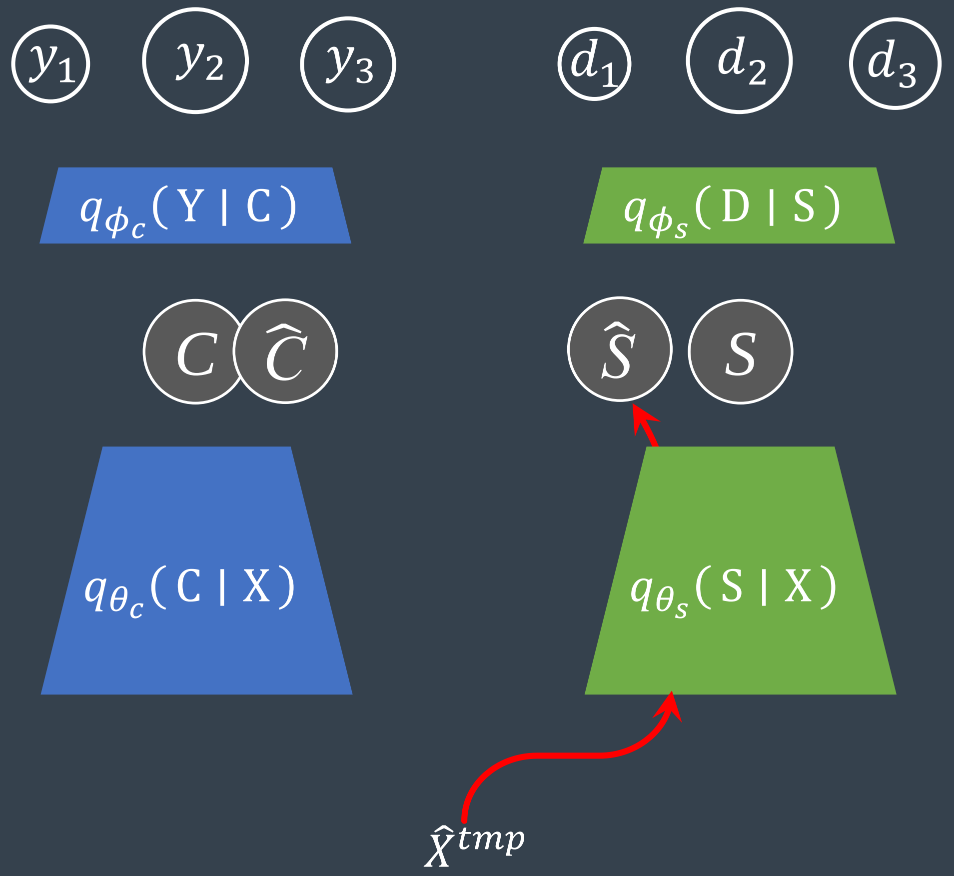


Solution

- Augmenting content and style:

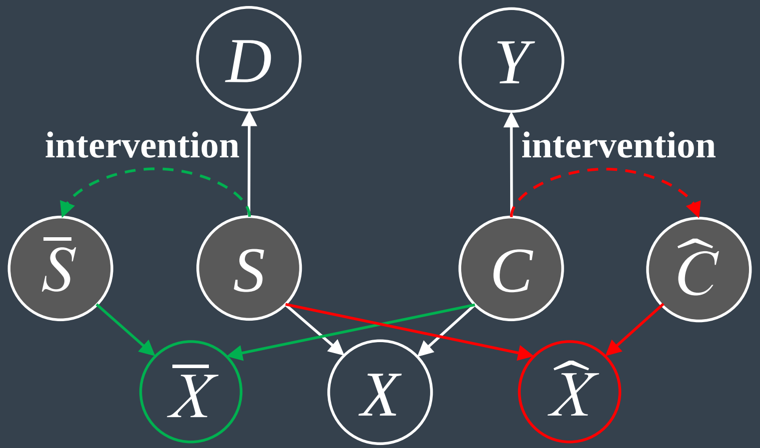


(3)

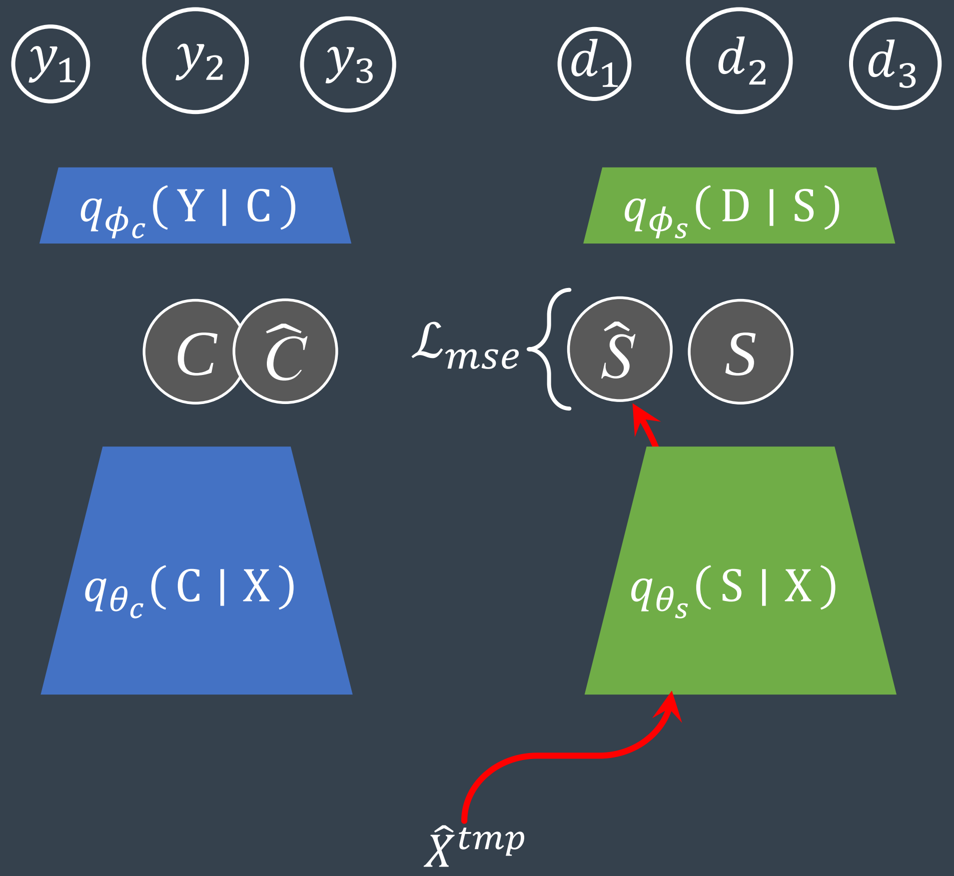


Solution

- Augmenting content and style:

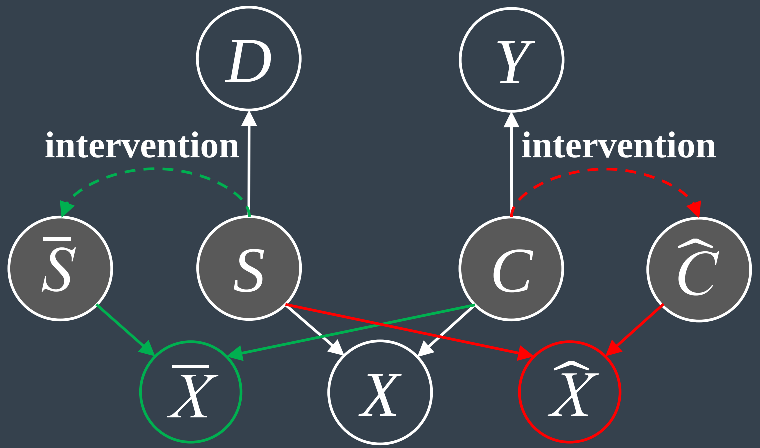


(3)

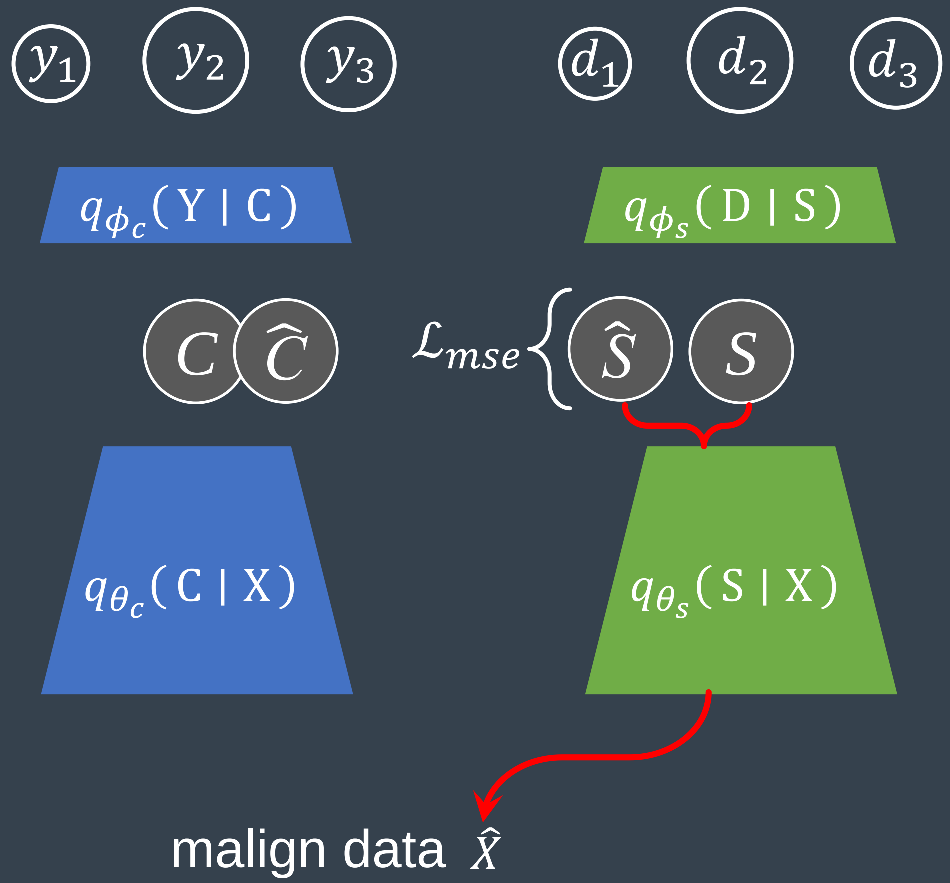


Solution

- Augmenting content and style:

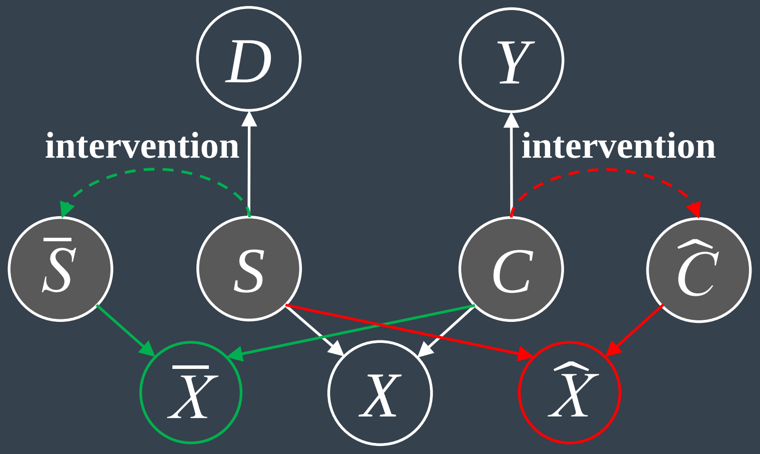


(3)

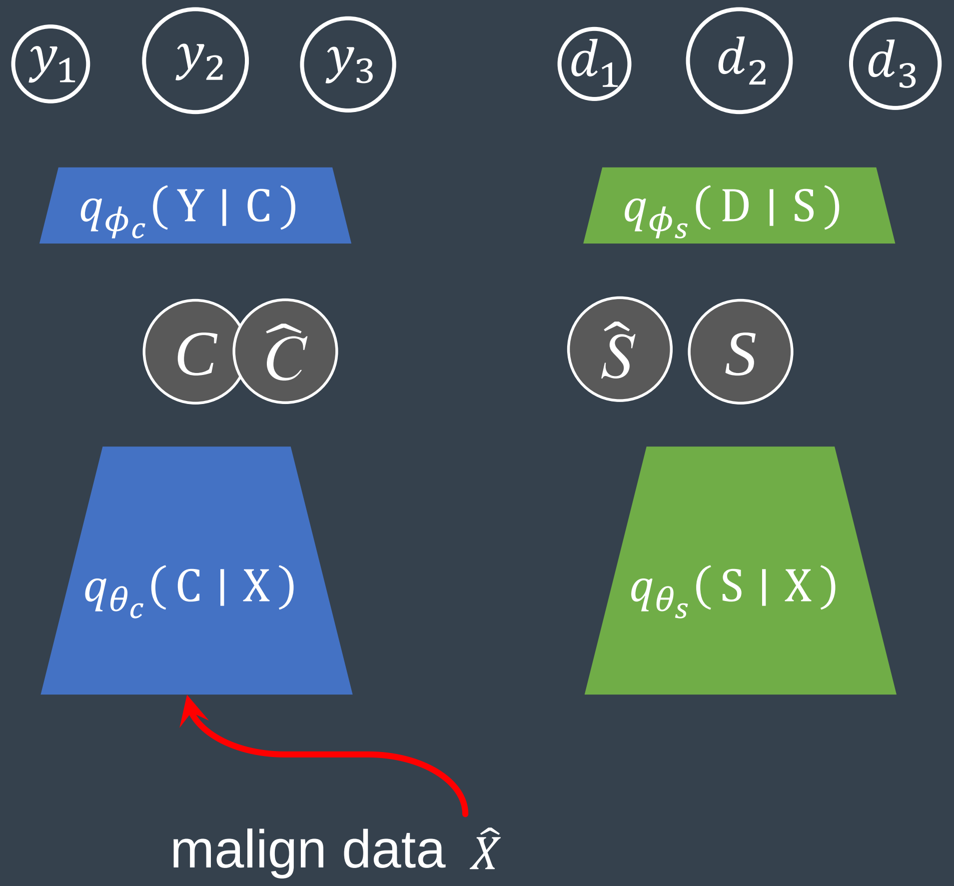


Solution

- Augmenting content and style:

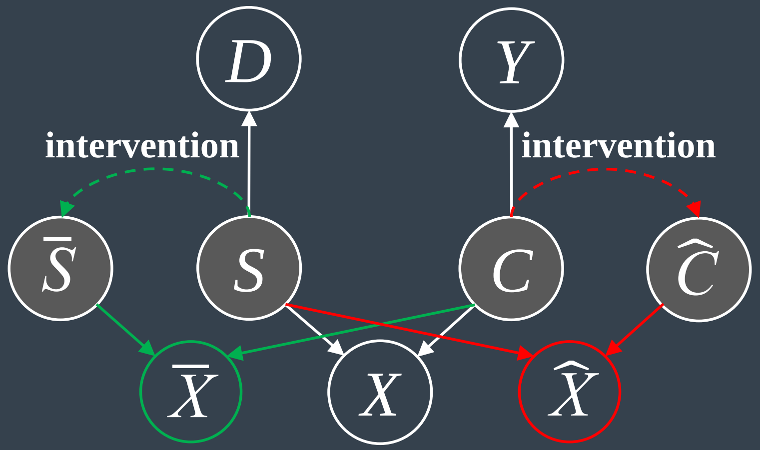


(3)

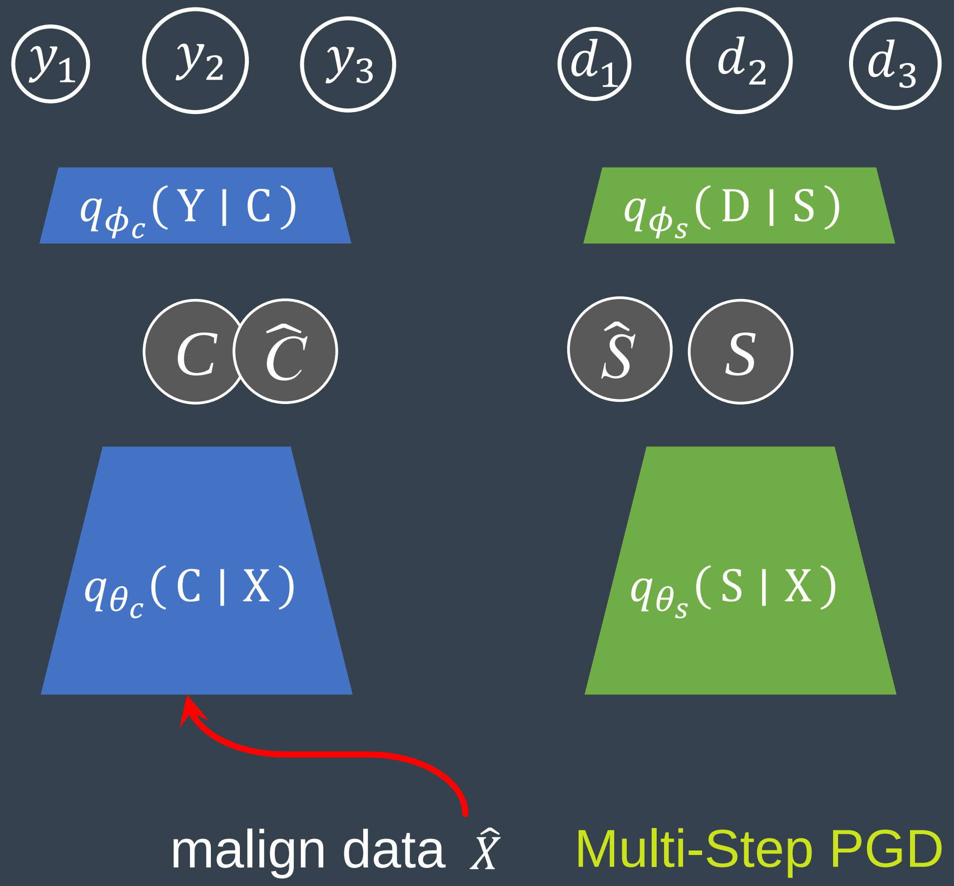


Solution

- Augmenting content and style:

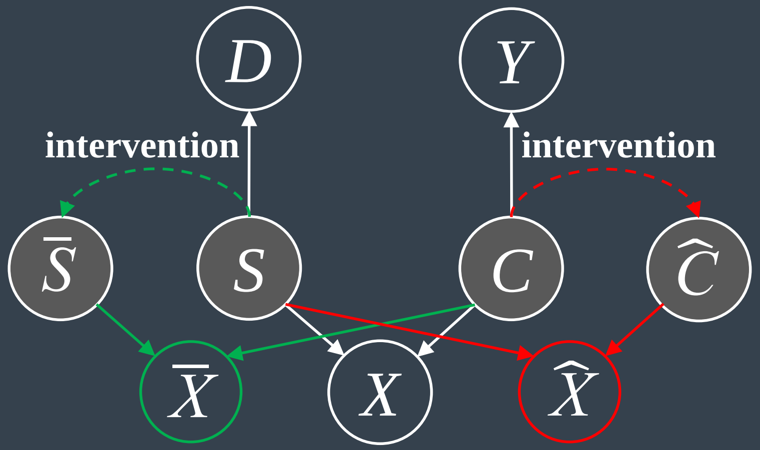


(3)

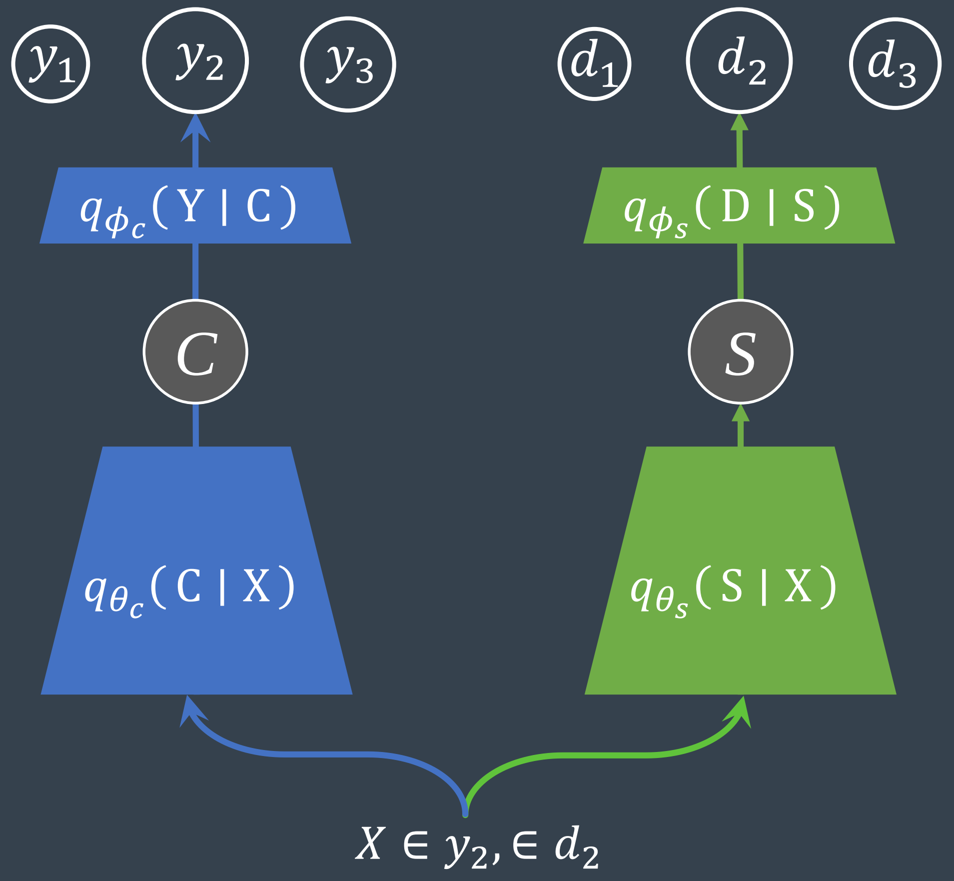


Solution

- Augmenting content and style:

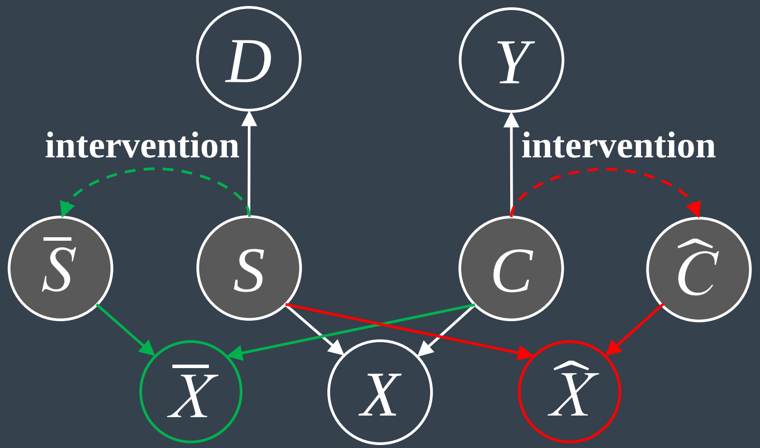


(3)

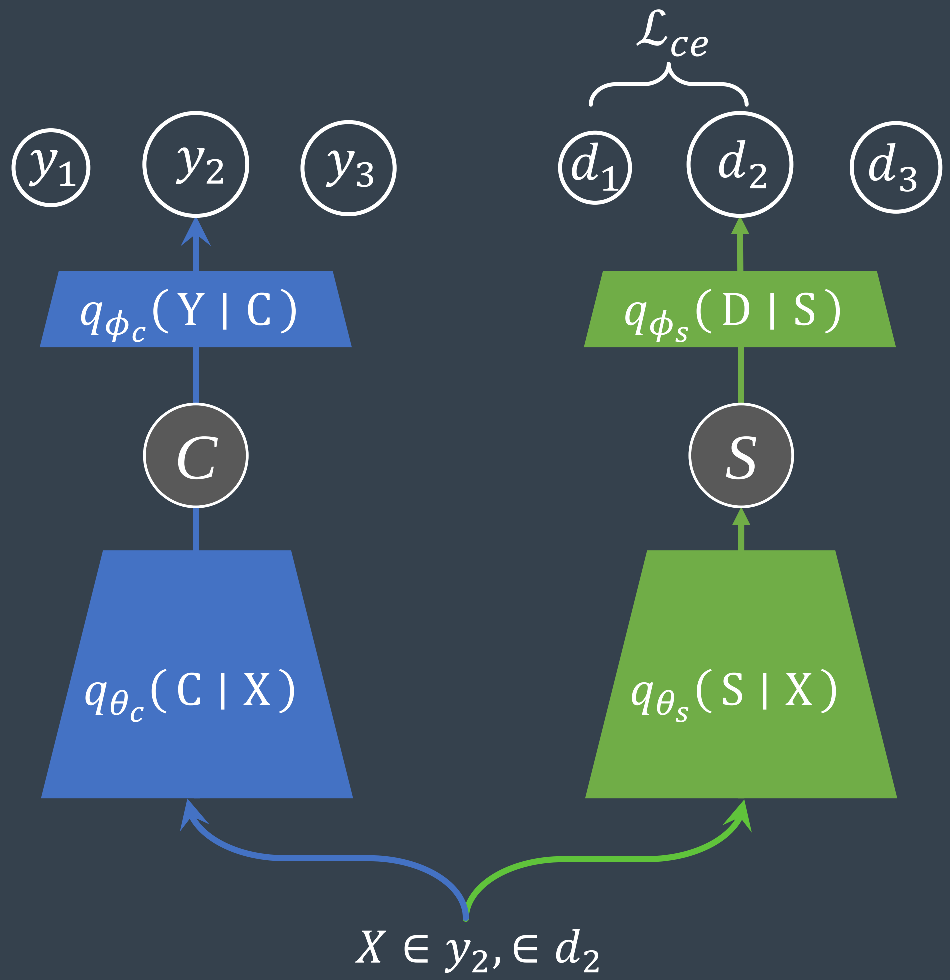


Solution

- Augmenting content and style:

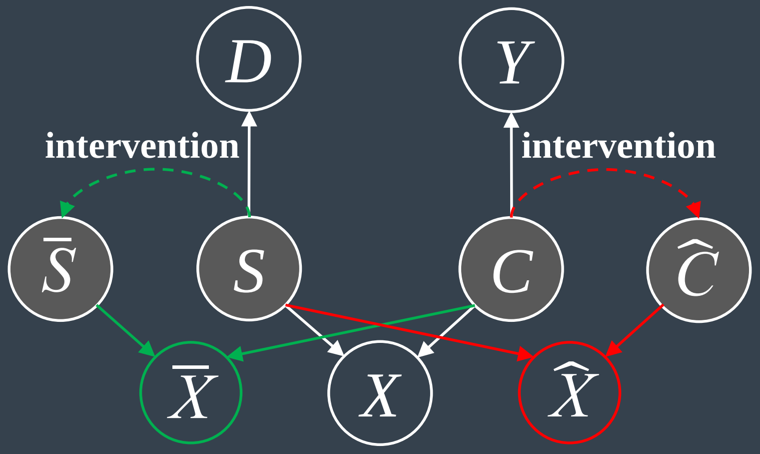


(3)

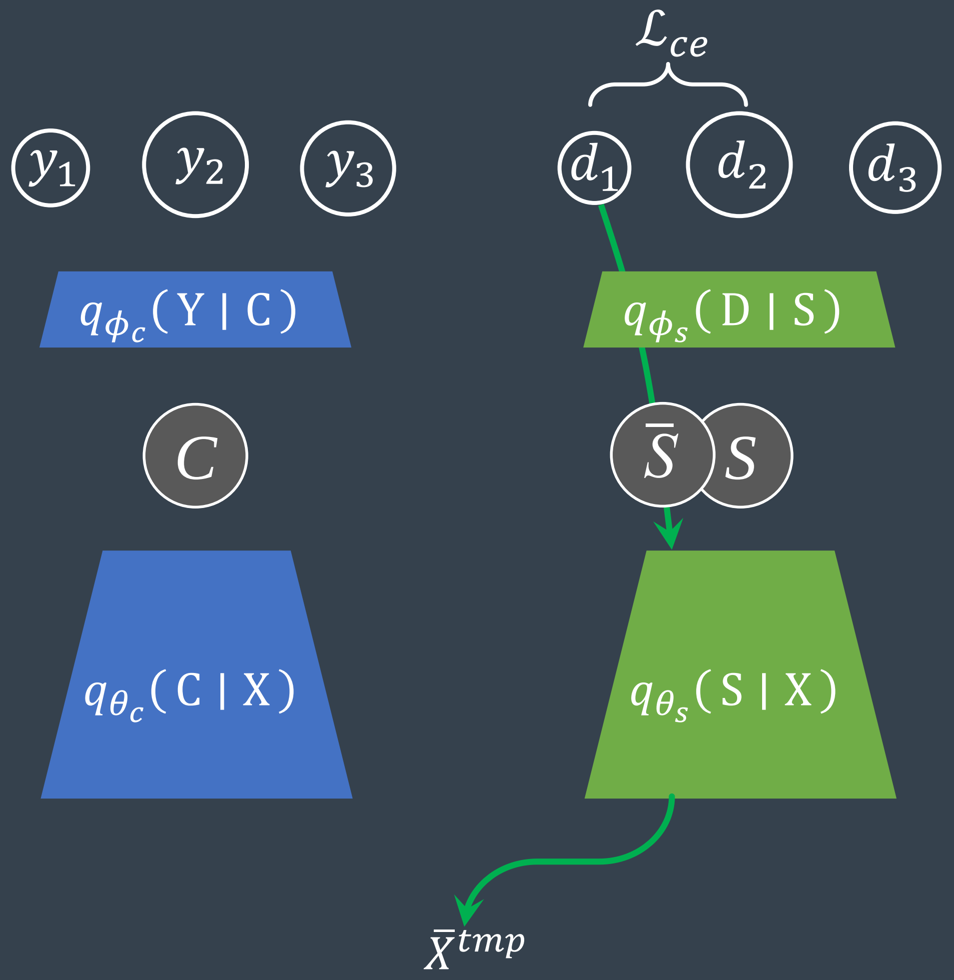


Solution

- Augmenting content and style:

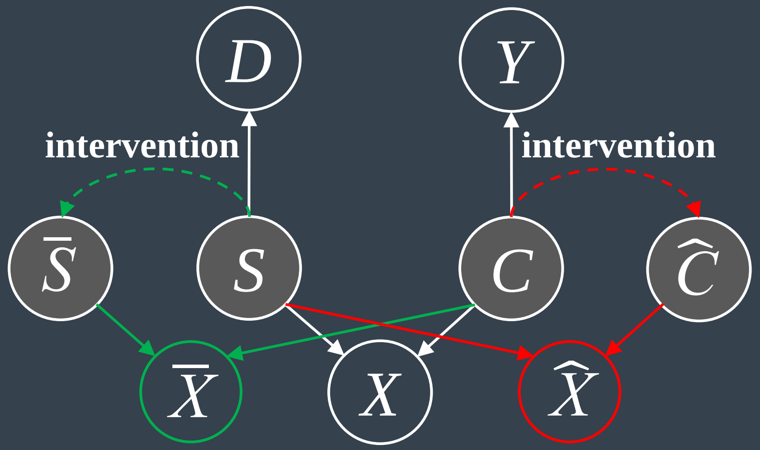


(3)

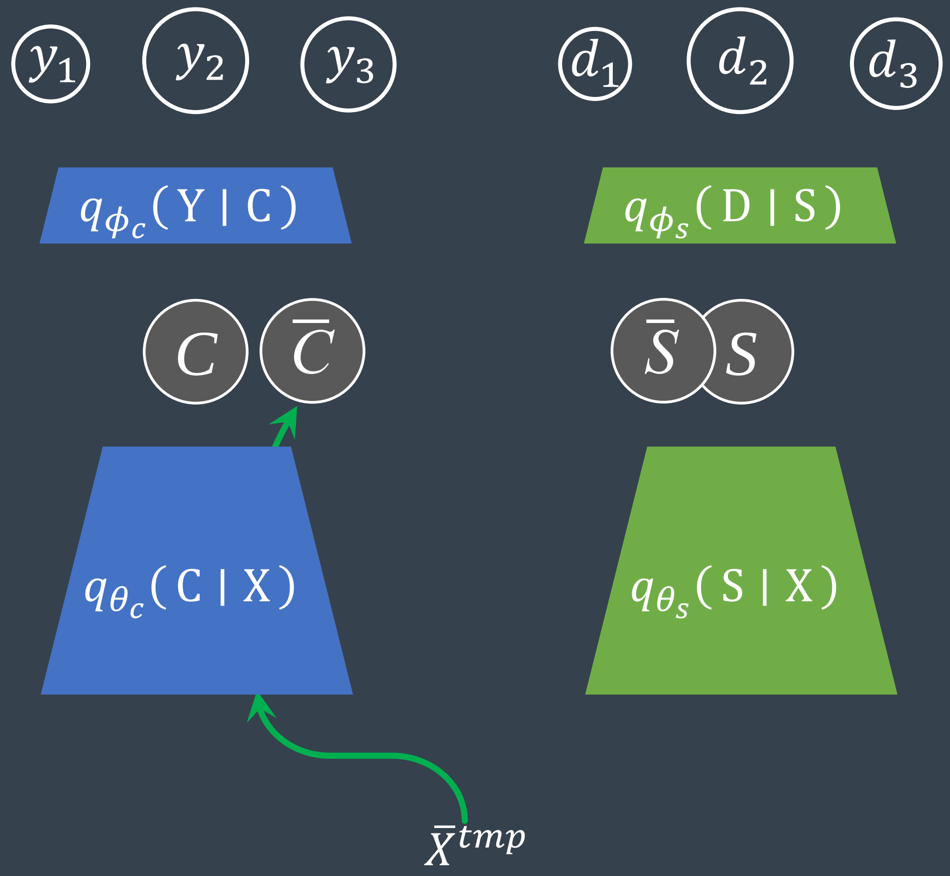


Solution

- Augmenting content and style:

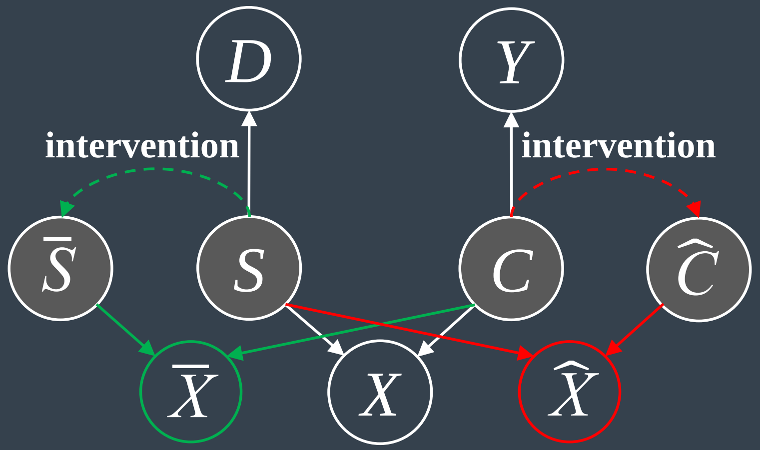


(3)

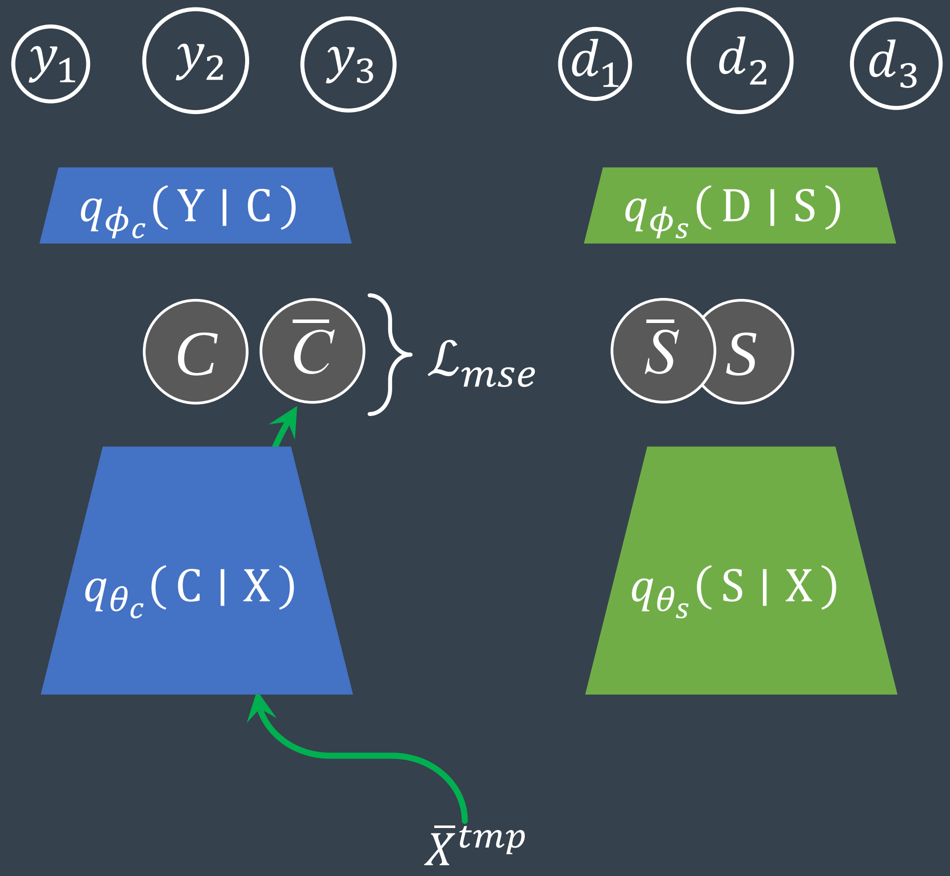


Solution

- Augmenting content and style:

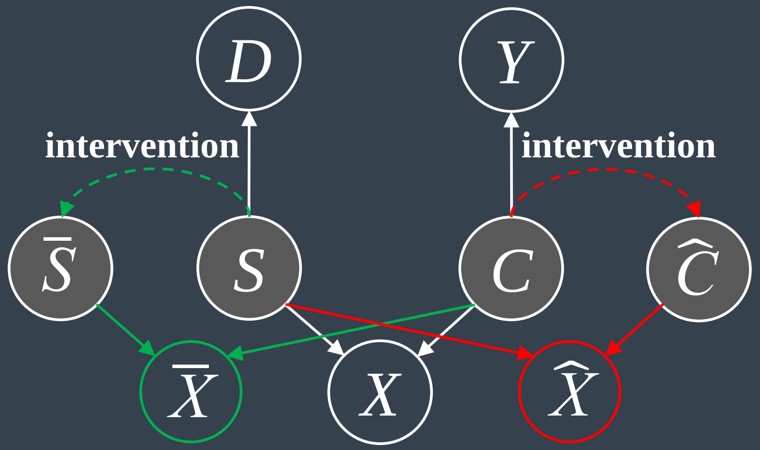


(3)

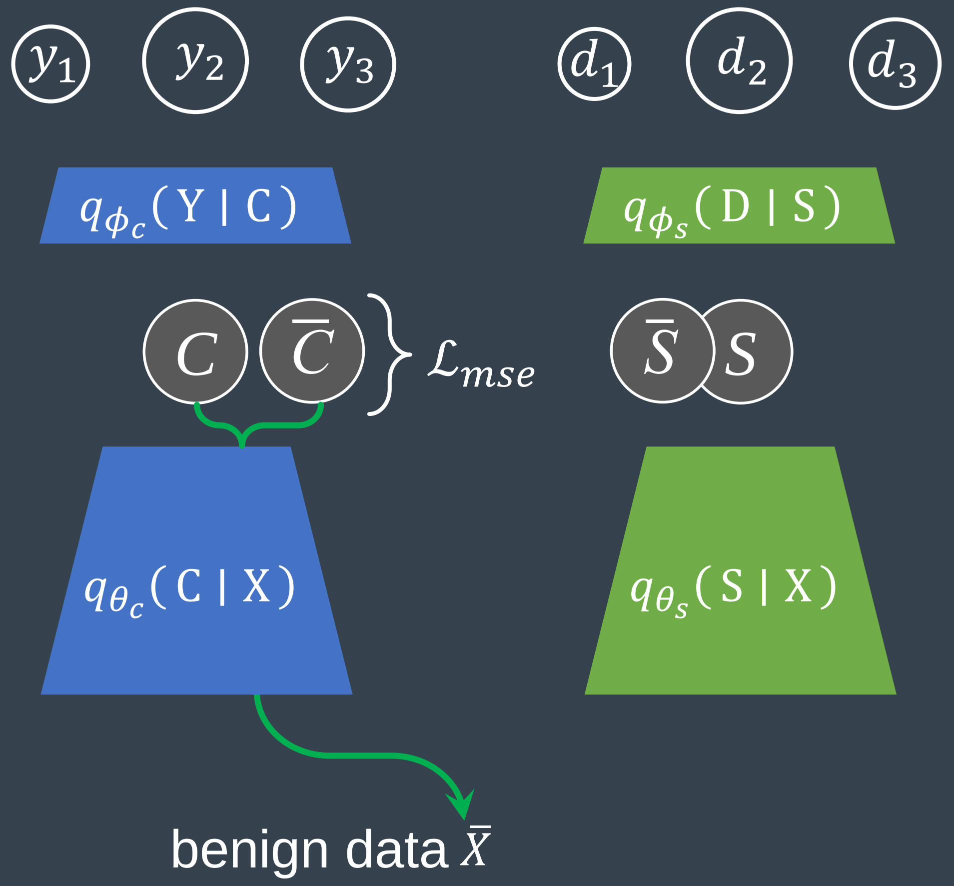


Solution

- Augmenting content and style:

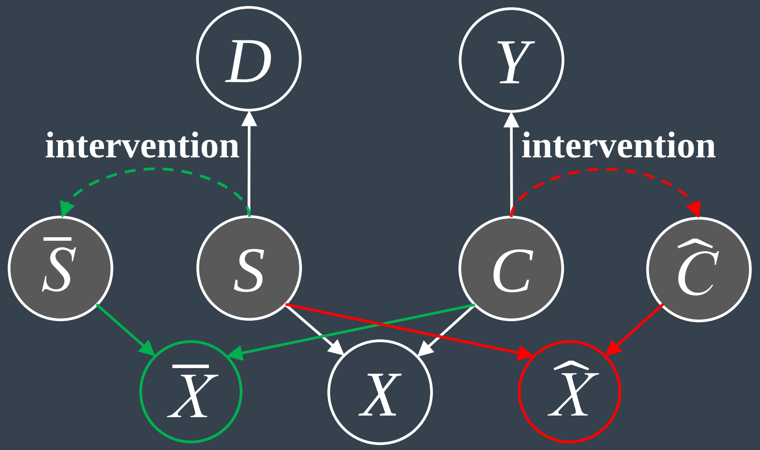


(3)

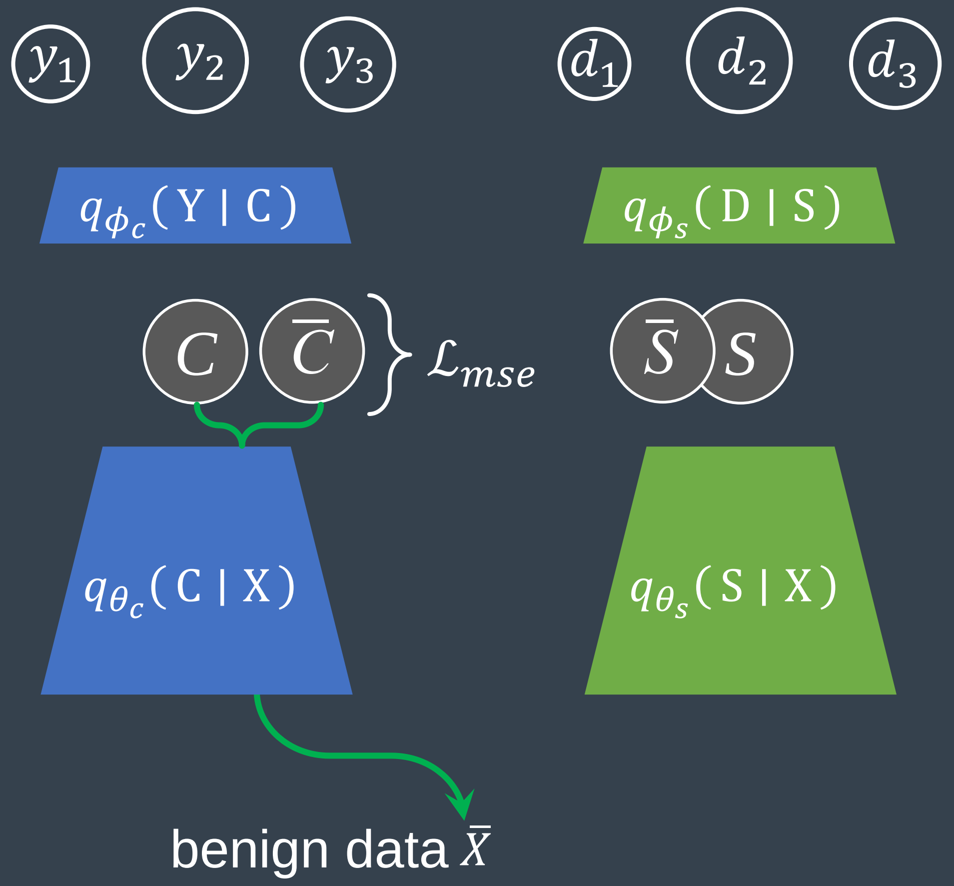


Solution

- Augmenting content and style:

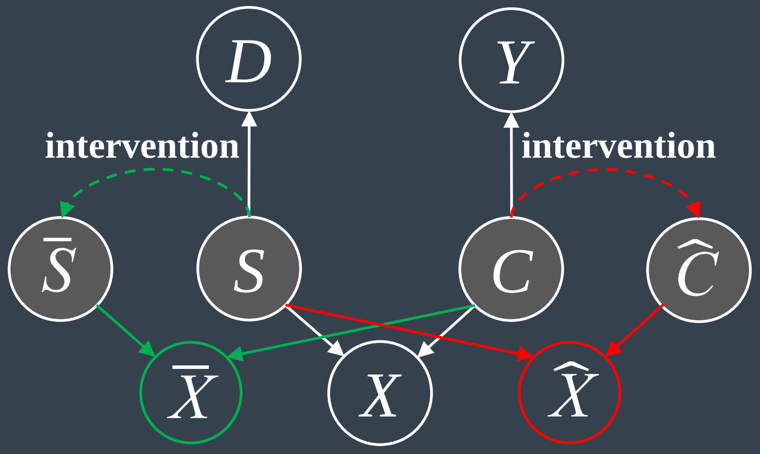


(3)

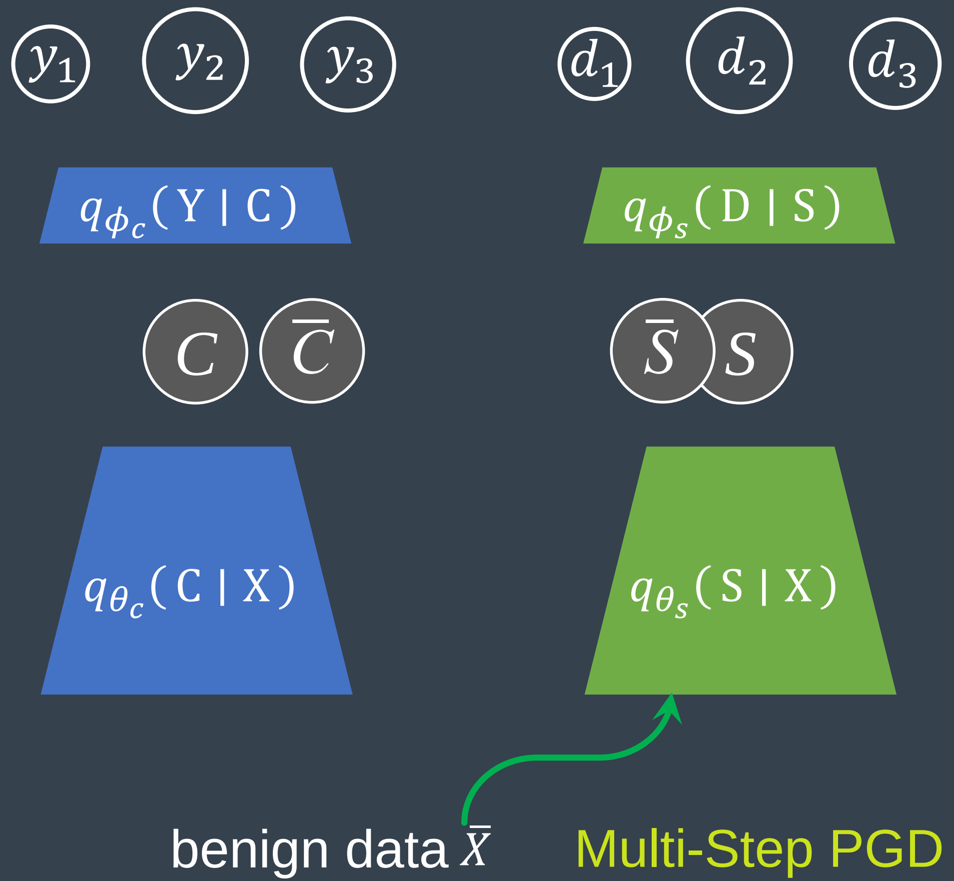


Solution

- Augmenting content and style:



(3)



Solution

- Harnessing benign data and malign data:

$$X \in y_2, \in d_2 \begin{cases} \text{➢ benign data: } \bar{X} \\ \\ \text{➢ malign data: } \hat{X} \end{cases}$$

Solution

- Harnessing benign data and malign data:

y_1

$$X \in y_2, \in d_2 \quad \left\{ \begin{array}{l} \text{➢ benign data: } \bar{X} \longrightarrow y_2 \\ \\ \text{➢ malign data: } \hat{X} \end{array} \right.$$

y_3

Solution

- Harnessing benign data and malign data:

y_1

$X \in y_2, \in d_2$

➢ benign data: $\bar{X} \longrightarrow y_2$

y_3

y_1

➢ malign data: $\hat{X}$ y_2

y_3

Solution

- Harnessing benign data and malign data:

y_1

➢ benign data: $\bar{X} \longrightarrow y_2$

y_3

$X \in y_2, \in d_2$

y_1

➢ malign data: $\hat{X} \not\longrightarrow y_2$

y_3

Solution

- Harnessing benign data and malign data:

y_1

$X \in y_2, \in d_2$

- benign data: $\bar{X} \longrightarrow y_2$

y_3

y_1

- malign data: $\hat{X}$

y_2

y_3

Solution

- Harnessing benign data and malign data:

$X \in y_2, \in d_2$

- benign data: $\bar{X}$ → y_2
- malign data: $\hat{X}$

y_1

y_2

y_3

negative	positive
y_1	y_1
y_2	y_2
y_3	y_3

Solution

- Harnessing benign data and malign data:

$X \in y_2, \in d_2$

- ➢ benign data: $\bar{X} \rightarrow y_2$
- ➢ malign data: $\hat{X} \rightarrow y_2$

y_1

y_2

y_3

negative	positive
y_1	y_1
y_2	y_2
y_3	y_3

Solution

- Harnessing benign data and malign data:

$X \in y_2, \in d_2$

- benign data: $\bar{X} \rightarrow y_2$ (of y_1, y_2, y_3)
- malign data: $\hat{X} \rightarrow y_2$ (of y_1, y_2, y_3)

y_1 y_1

y_2 y_2

y_3 y_3

negative positive

P($\hat{X} \in y_2$) = 0.8

Solution

- Harnessing benign data and malign data:

$X \in y_2, \in d_2$

- benign data: $\bar{X} \rightarrow y_2$

y_1

y_2

y_3

- malign data: $\hat{X} \rightarrow y_2$

negative	positive
y_1	y_1
y_2	y_2
y_3	y_3

P($\hat{X} \in y_2$) = 0.8

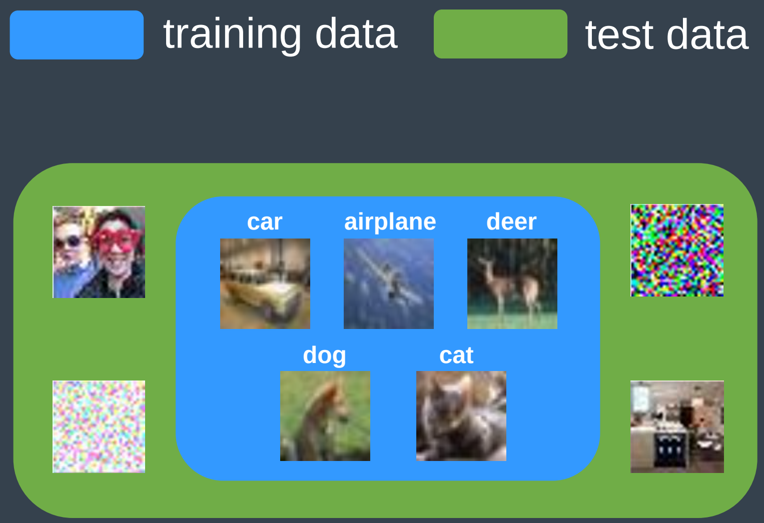
P($\hat{X} \in y_2$) = 1-0.8
= 0.2

Experiments

- OOD Detection.
- Open-Set Semi-Supervised Learning.
- Open-set Domain Adaptation.

Experiments

- OOD Detection:



Experiments

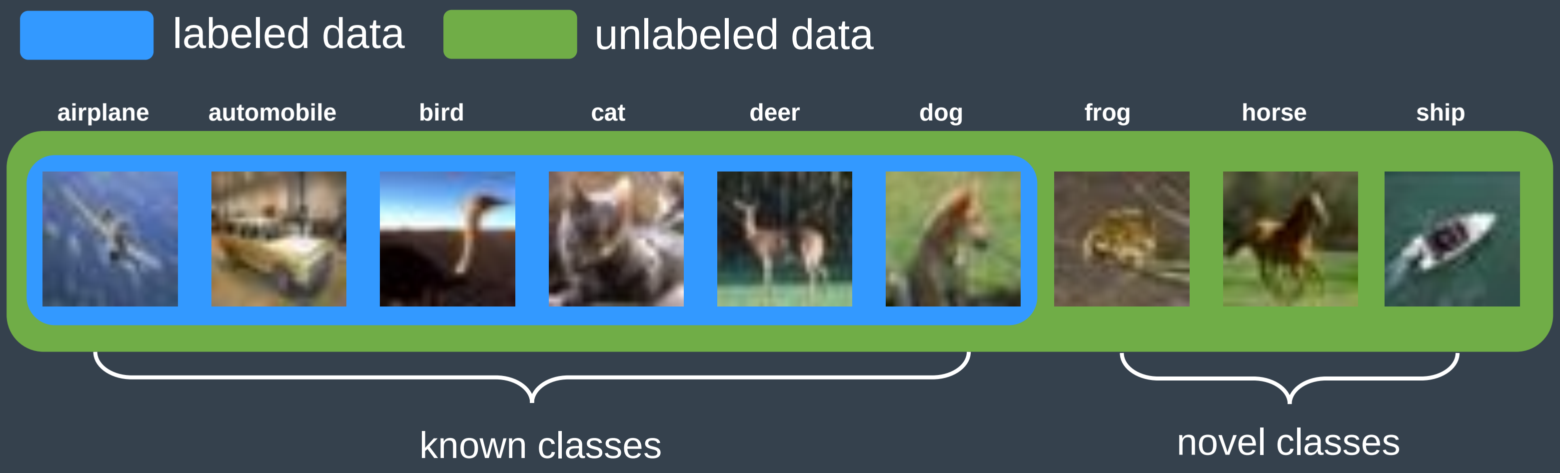
- OOD Detection.
  - ➢Results:

Table 1: Comparison with typical OOD detections methods. Averaged AUROC (%) with standard deviations are computed over three independent trails. The best results are highlighted in **bold**.

OOD dataset	LSUN	DTD	CUB	Flowers	Caltech	Dogs
ID dataset			SVHN			
Likelihood	52.25 ± 0.3	50.33 ± 0.7	48.76 ± 0.6	47.33 ± 0.2	51.54 ± 0.4	54.34 ± 0.4
ODIN	55.72 ± 0.2	53.32 ± 0.5	52.70 ± 0.4	50.47 ± 0.7	56.41 ± 0.4	61.16 ± 0.3
Likelihood Ratio	79.34 ± 0.5	78.42 ± 0.3	75.90 ± 0.7	74.53 ± 0.4	76.25 ± 0.3	83.55 ± 0.4
OpenGAN	83.77 ± 0.4	80.36 ± 0.5	77.49 ± 0.8	79.26 ± 0.5	$\mathbf{86.66 \pm 0.5}$	86.84 ± 0.5
HOOD	$\mathbf{84.10 \pm 0.6}$	$\mathbf{80.68 \pm 0.6}$	$\mathbf{79.24 \pm 0.5}$	$\mathbf{80.93 \pm 0.7}$	85.34 ± 0.7	$\mathbf{87.58 \pm 0.8}$
ID dataset			CIFAR10			
Likelihood	54.32 ± 0.5	52.16 ± 0.4	50.67 ± 0.4	49.26 ± 0.3	53.86 ± 0.4	56.92 ± 0.2
ODIN	58.60 ± 0.3	55.59 ± 0.6	58.48 ± 0.7	51.44 ± 0.9	59.36 ± 0.4	64.82 ± 0.5
Likelihood Ratio	81.41 ± 0.6	79.77 ± 0.5	79.35 ± 0.8	77.17 ± 0.7	80.67 ± 0.5	86.76 ± 0.3
OpenGAN	84.03 ± 0.4	81.29 ± 0.8	82.84 ± 1.0	$\mathbf{82.32 \pm 0.4}$	86.78 ± 0.3	90.14 ± 0.5
HOOD	$\mathbf{86.12 \pm 0.6}$	$\mathbf{83.64 \pm 0.5}$	$\mathbf{83.53 \pm 0.6}$	81.56 ± 0.8	$\mathbf{87.24 \pm 0.8}$	$\mathbf{90.86 \pm 0.6}$

Experiments

- OOD Detection.
- Open-Set Semi-Supervised Learning:



Experiments

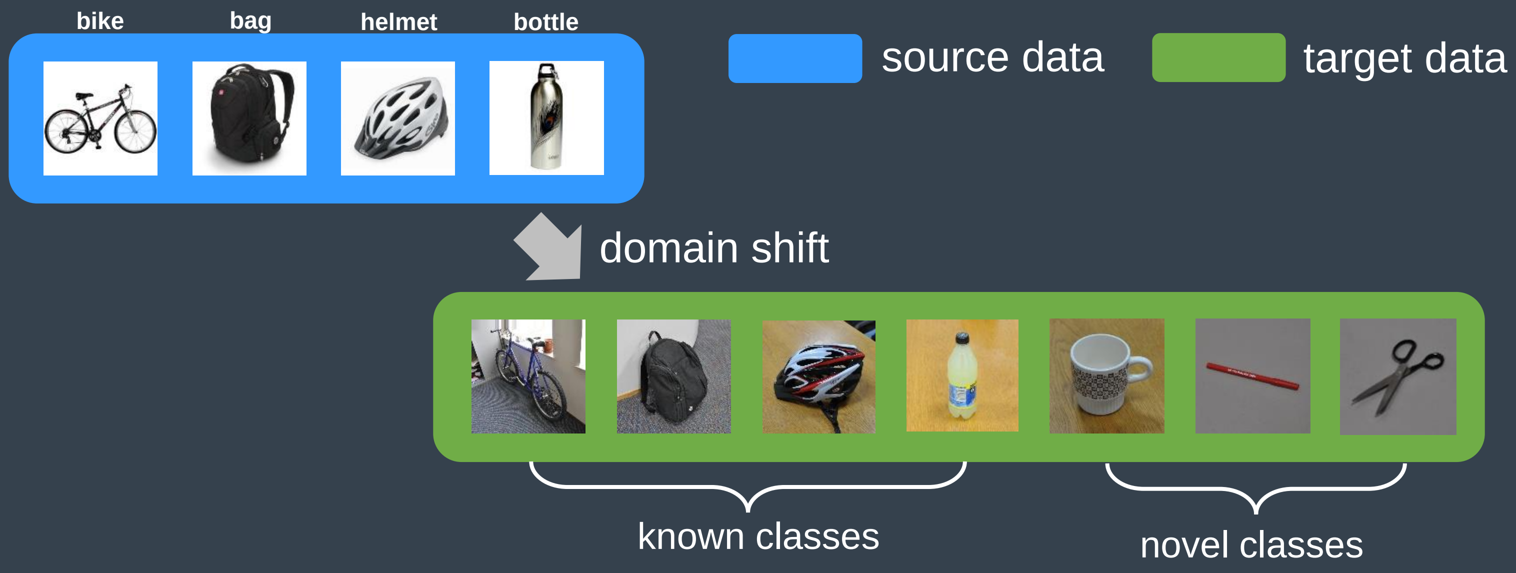
- OOD Detection.
- Open-Set Semi-Supervised Learning.
  - ➢Results:

Table 2: Comparison with typical Open-set SSL methods. Averaged test accuracies (%) with standard deviations are computed over three independent trails. The best results are highlighted in **bold**.

Training dataset		CIFAR10			CIFAR100		
No. of Labeled data		50	100	400	50	100	400
Clean Acc.	UASD	72.82 ± 0.9	75.53 ± 1.8	76.74 ± 1.7	58.87 ± 0.6	61.68 ± 1.2	65.97 ± 2.4
	DS3L	74.44 ± 1.3	76.89 ± 1.5	78.80 ± 0.6	60.40 ± 0.5	64.35 ± 1.5	67.65 ± 1.3
	MTCF	79.88 ± 1.3	81.41 ± 1.0	83.92 ± 0.8	62.78 ± 0.5	65.84 ± 2.1	69.46 ± 0.6
	OpenMatch	$\mathbf{84.10 \pm 1.1}$	$\mathbf{85.30 \pm 0.4}$	$\mathbf{87.92 \pm 1.0}$	65.76 ± 0.9	$\mathbf{68.46 \pm 0.5}$	72.87 ± 1.4
	T2T	82.74 ± 1.2	83.56 ± 1.4	85.97 ± 0.8	65.16 ± 1.2	67.58 ± 0.9	71.96 ± 1.1
	HOOD	83.55 ± 1.2	84.16 ± 1.5	86.22 ± 2.7	$\mathbf{66.39 \pm 1.7}$	68.03 ± 2.6	$\mathbf{73.32 \pm 0.6}$
Corrupted Acc.	UASD	39.36 ± 1.2	41.38 ± 0.7	42.66 ± 1.8	31.55 ± 2.0	33.39 ± 1.7	35.20 ± 0.8
	DS3L	39.97 ± 0.8	42.58 ± 0.8	44.39 ± 0.6	33.72 ± 0.8	34.67 ± 0.8	36.64 ± 0.6
	MTCF	40.16 ± 1.2	40.58 ± 1.1	43.33 ± 0.7	32.72 ± 0.8	34.33 ± 2.3	35.53 ± 0.6
	OpenMatch	41.38 ± 0.7	42.90 ± 0.6	45.79 ± 0.8	35.98 ± 1.3	36.47 ± 0.7	38.56 ± 0.6
	T2T	41.39 ± 1.6	45.56 ± 1.6	49.88 ± 1.5	$\mathbf{41.03 \pm 1.7}$	39.64 ± 0.7	41.38 ± 1.6
	HOOD	$\mathbf{44.42 \pm 1.7}$	$\mathbf{48.38 \pm 0.9}$	$\mathbf{50.74 \pm 0.6}$	40.82 ± 1.5	$\mathbf{41.65 \pm 0.9}$	$\mathbf{43.72 \pm 2.2}$

Experiments

- OOD Detection.
- Open-Set Semi-Supervised Learning.
- Open-set Domain Adaptation:



Experiments

- OOD Detection.
- Open-Set Semi-Supervised Learning.
- Open-set Domain Adaptation.
  - Results:

Table 3: Comparison with typical Open-set DA methods. Averaged test accuracies (%) with standard deviations are computed over three independent trails. The best results are highlighted in **bold**.

Dataset	Office						VisDA
Domain	A→W	A→D	D→W	W→D	D→A	W→A	Synthetic→Real
OSBP	86.5 ± 2.0	88.6 ± 1.4	97.0 ± 1.0	97.9 ± 0.9	88.9 ± 2.5	85.8 ± 2.5	62.9 ± 1.3
UAN	87.7 ± 1.2	87.0 ± 0.8	93.5 ± 1.3	97.2 ± 1.6	88.4 ± 0.7	87.8 ± 1.6	63.8 ± 2.4
STA	89.5 ± 0.6	93.7 ± 1.5	97.5 ± 0.2	$\mathbf{99.5 \pm 0.2}$	89.1 ± 0.5	87.9 ± 0.9	66.4 ± 1.3
HOOD	$\mathbf{90.1 \pm 1.5}$	$\mathbf{94.2 \pm 1.4}$	$\mathbf{99.6 \pm 0.6}$	98.3 ± 0.9	$\mathbf{89.8 \pm 0.8}$	$\mathbf{91.3 \pm 1.8}$	$\mathbf{72.4 \pm 1.6}$

Experiments

- Analysis:

Experiments

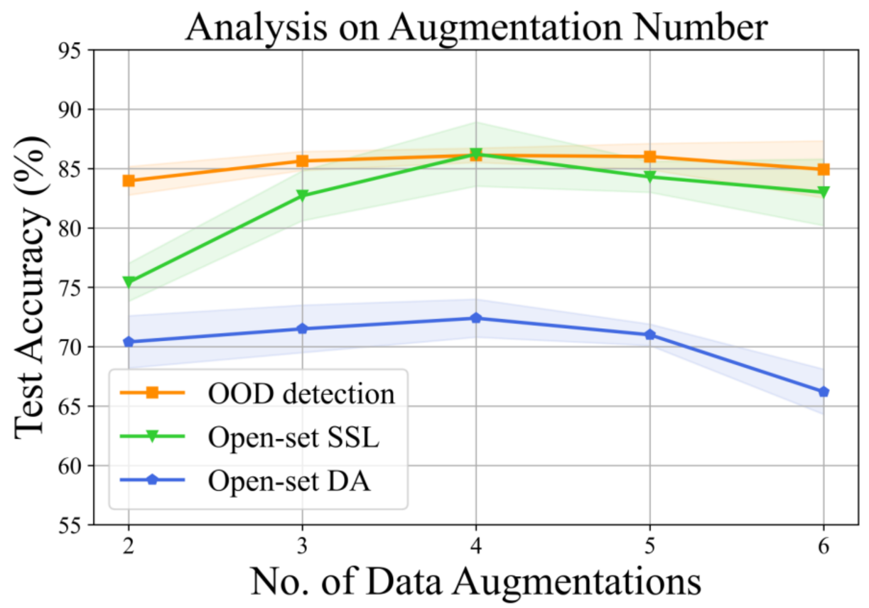
- Analysis:
  - Ablation Study:

Table 4: Ablation study on necessity of each module.

Application	OOD detection	Open-Set SSL	Open-Set DA
w/o disentanglement	84.94 ± 1.3	82.55 ± 1.1	64.6 ± 0.9
w/o benign OOD data	85.95 ± 1.8	83.32 ± 2.0	66.3 ± 2.5
w/o malign OOD data	82.50 ± 2.2	85.40 ± 0.8	71.8 ± 1.2
w/o both augmentations	80.83 ± 0.8	81.14 ± 1.2	65.4 ± 1.2
HOOD	$\mathbf{86.12 \pm 0.6}$	$\mathbf{86.22 \pm 2.7}$	$\mathbf{72.4 \pm 1.6}$

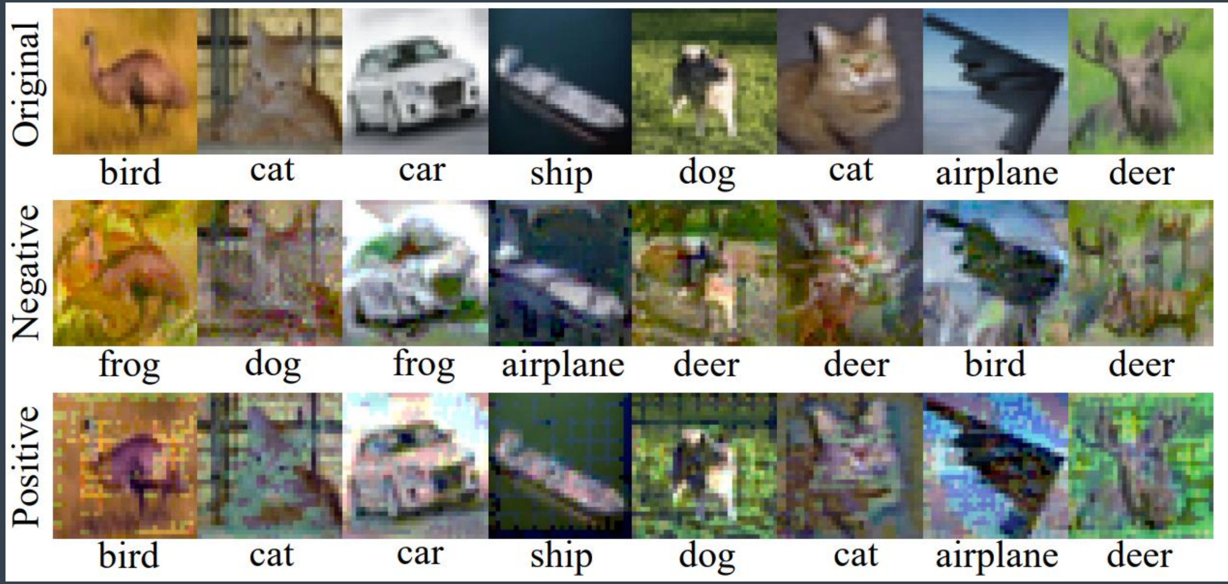
Experiments

- Analysis:
  - ➢Ablation Study.
  - ➢Parameter Analysis :



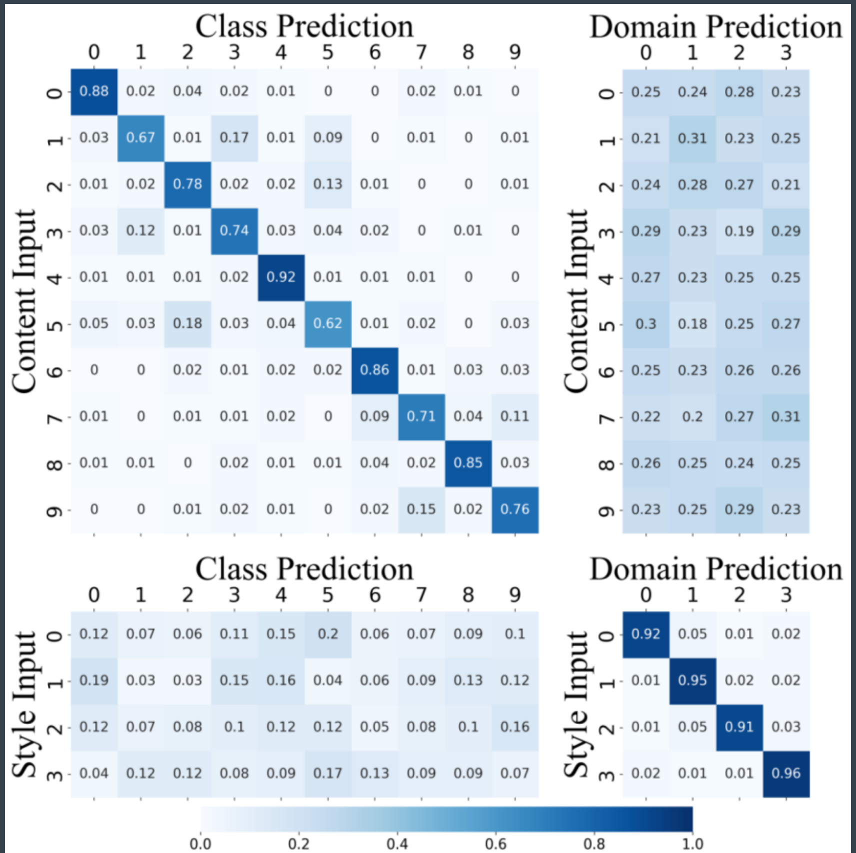
Experiments

- Analysis:
  - ➢Ablation Study.
  - ➢Parameter Analysis.
  - ➢Visualization:



Experiments

- Analysis:
  - ➢Ablation Study.
  - ➢Parameter Analysis.
  - ➢Visualization.
  - ➢Disentanglement Study:



Thank you!

Paper:

Code: