Style Transfer from Non-Parallel Text by Cross-Alignment

NIPS 2017

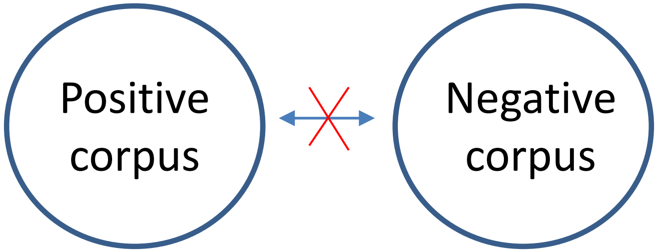
박정수

Data Mining & Information Systems Lab.
Department of Computer Science and Engineering,
College of Informatics, Korea University

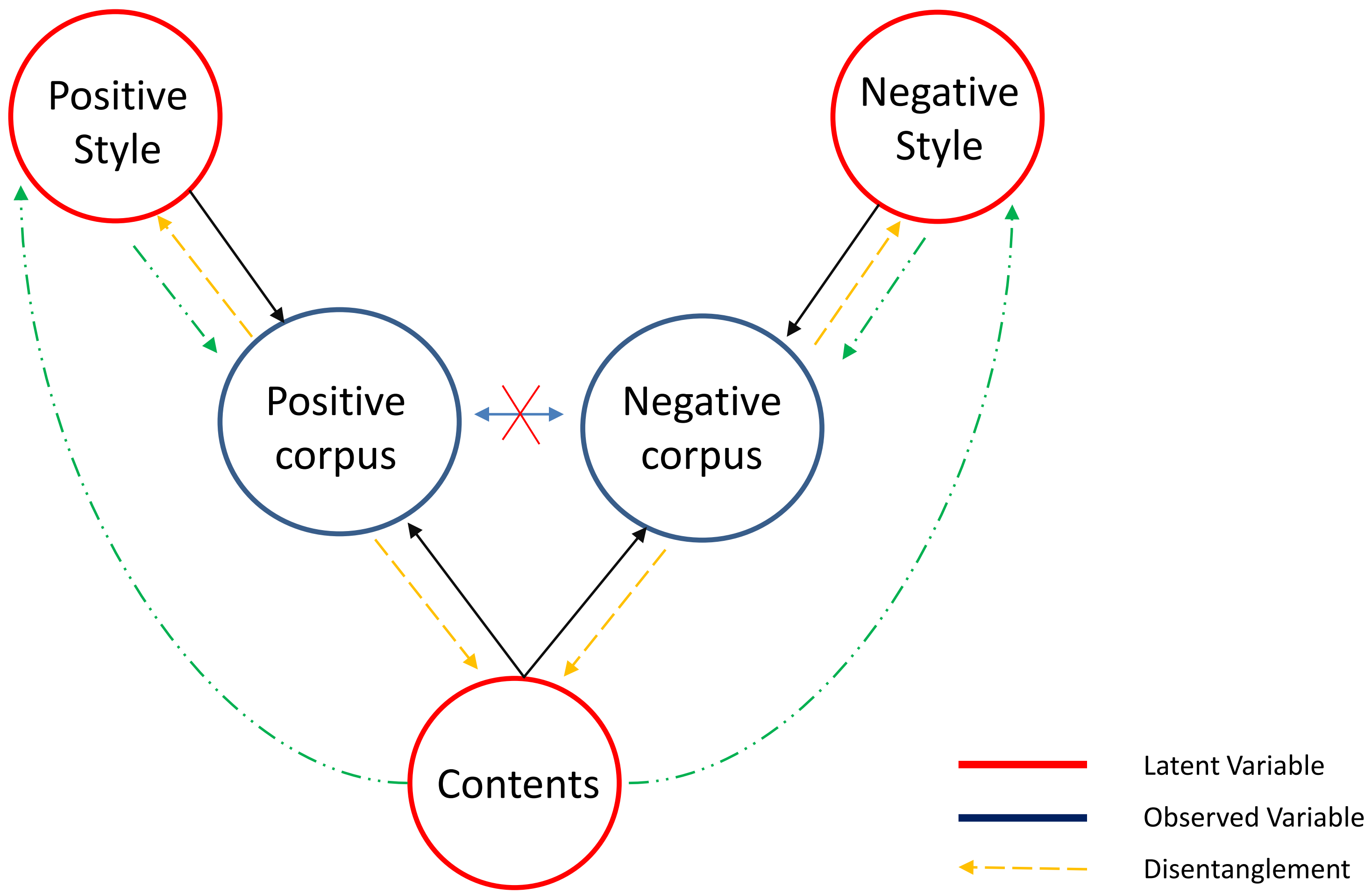
1 Introduction

How can we generate text
in a controlled way?

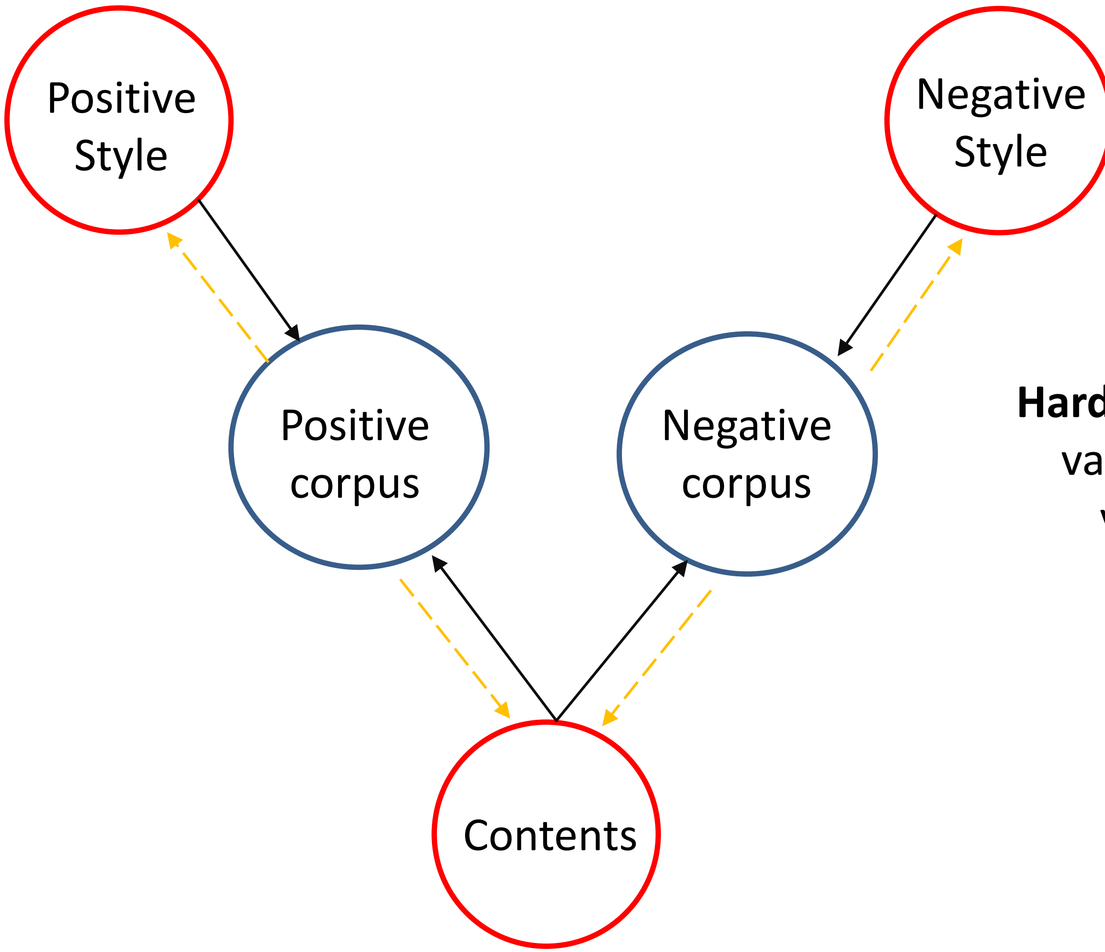
When non-parallel corpus are given



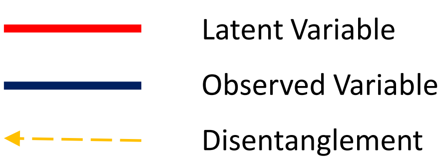
1 Introduction



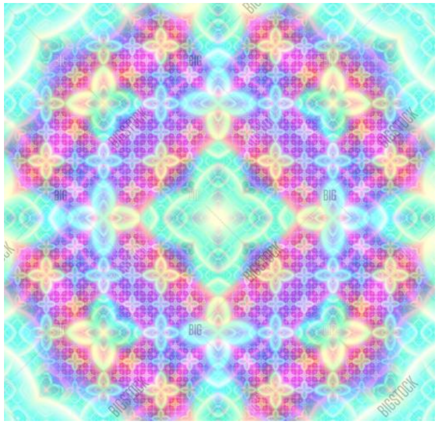
2 Problem



Hard to disentangle style latent variable and content latent variable independently



2 Problem



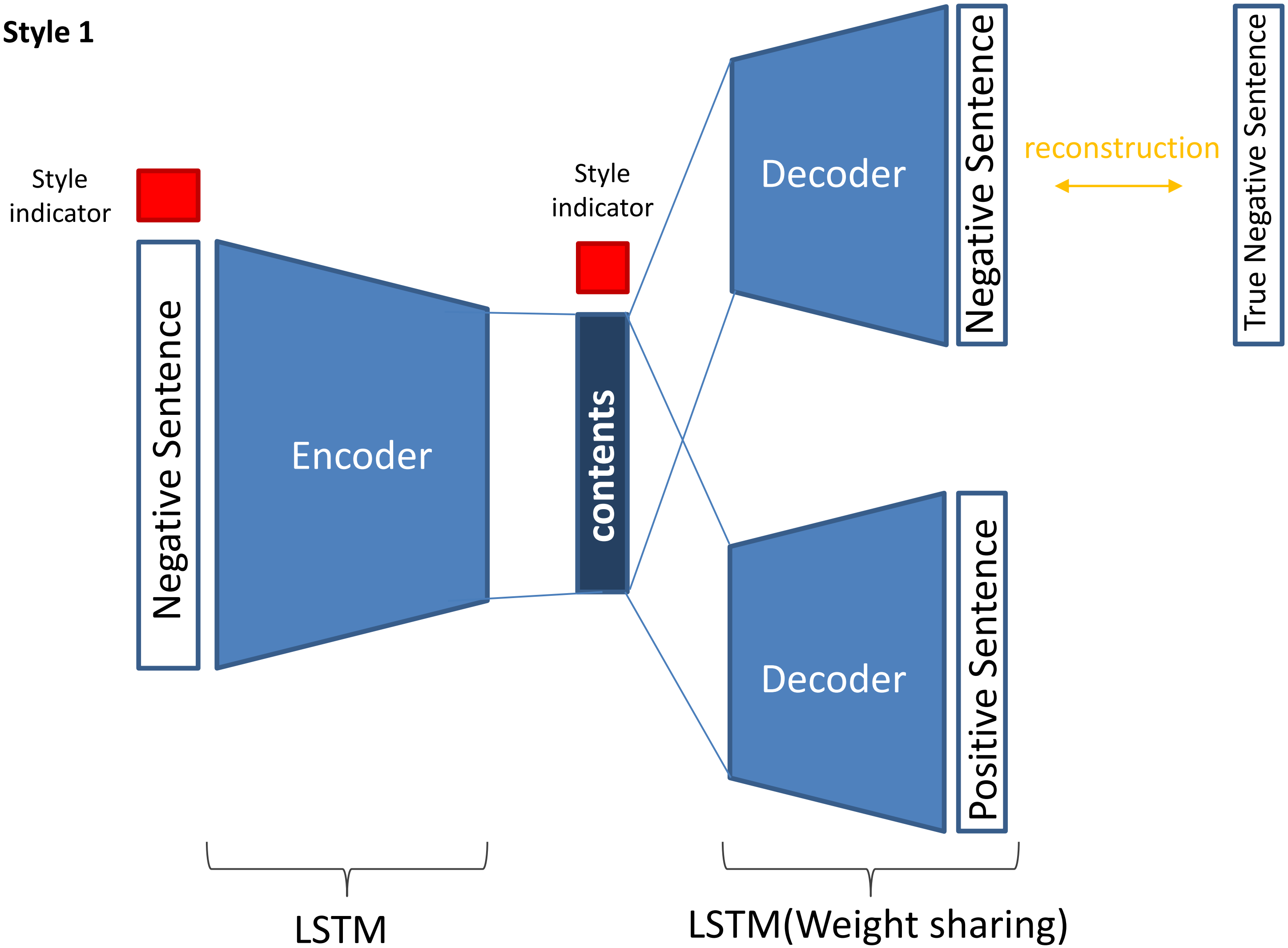
Images: pixel values are continuous

Text: Finite set of characters thus discrete

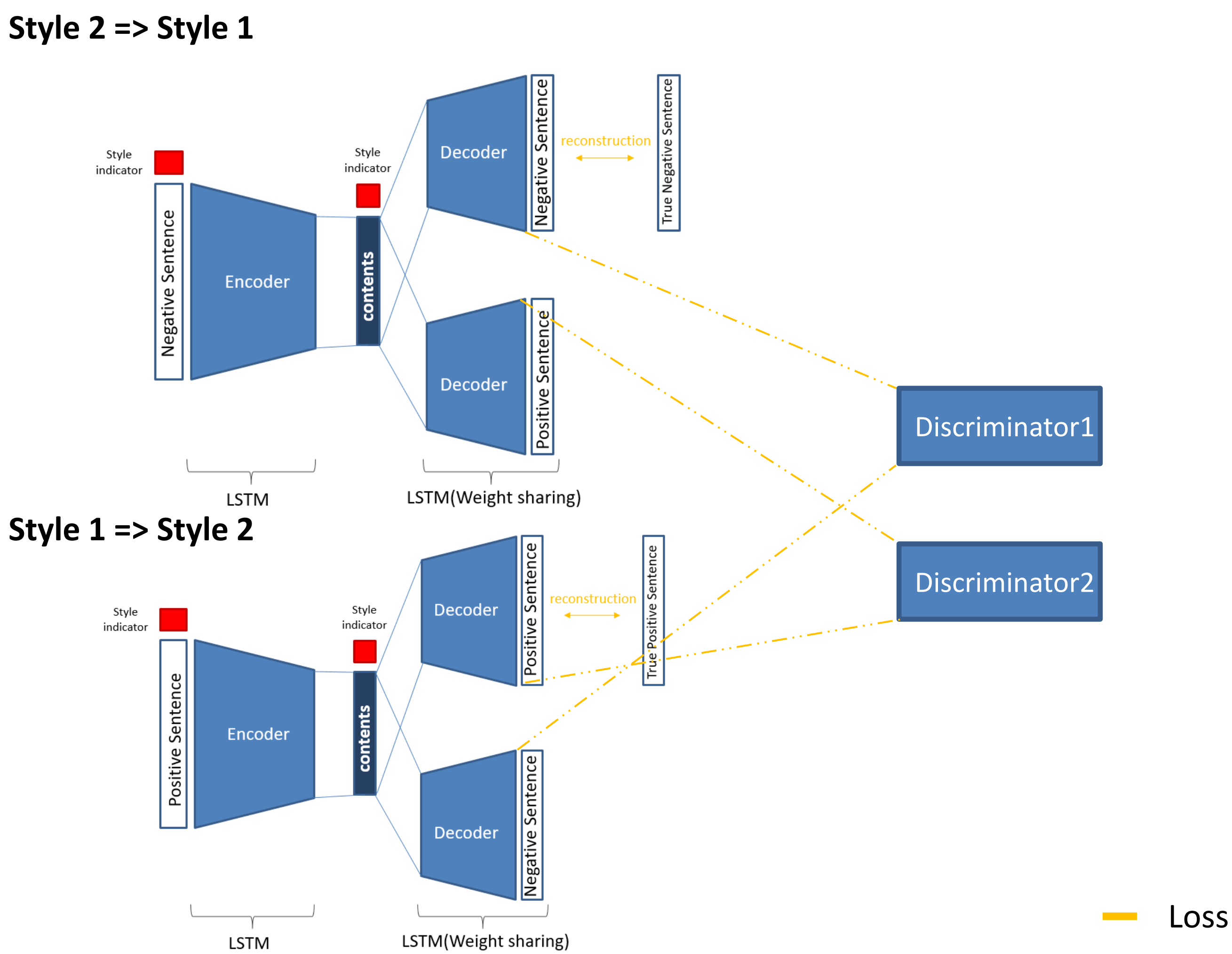
Text generated by machine can look weird **due to the sampling process** and discreteness hinders backpropagation

3 Method

Style 2 => Style 1

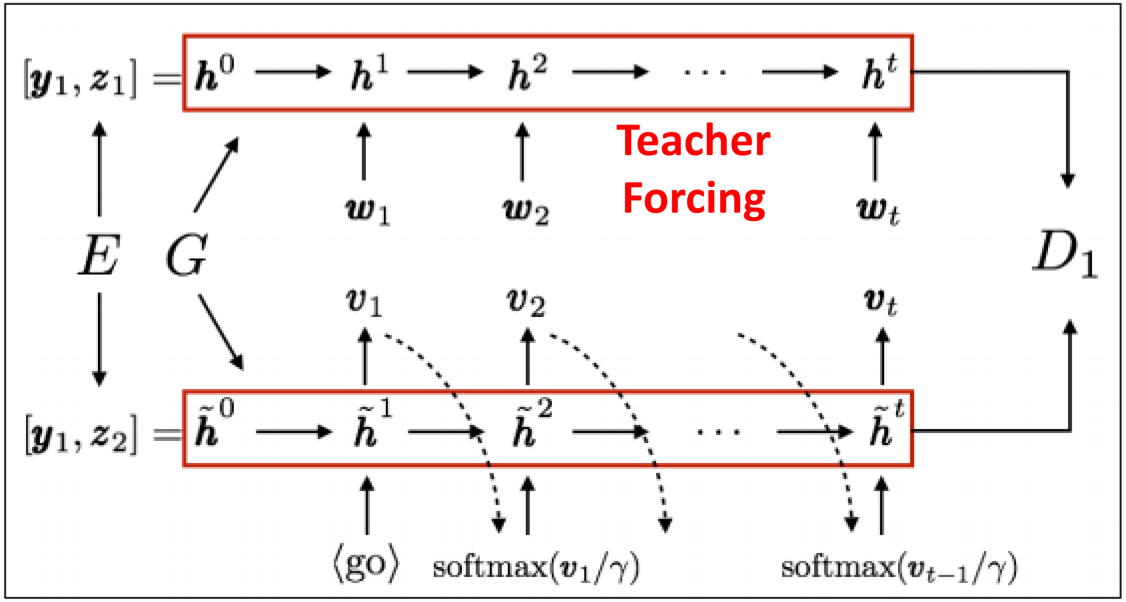


3 Method


Style 2 => Style 1
Style indicator
Negative Sentence
Encoder
Style indicator
contents
Decoder
Negative Sentence
reconstruction
True Negative Sentence
Decoder
Positive Sentence
LSTM
LSTM(Weight sharing)
Style 1 => Style 2
Style indicator
Positive Sentence
Encoder
Style indicator
contents
Decoder
Positive Sentence
reconstruction
True Positive Sentence
Decoder
Negative Sentence
LSTM
LSTM(Weight sharing)
Discriminator1
Discriminator2
Loss

3 Method

Generator



- Adopts Professor Forcing algorithm which makes generation process' hidden states similar to the behavior of answer fed's ones(Teacher Forcing)

- Gumbell-Softmax approximation is used for relaxation(with Tau parameter)

Sentiment Modification

From negative to positive
consistently slow . consistently good . consistently fast .
my goodness it was so gross . my husband 's steak was phenomenal . my goodness was so awesome .
it was super dry and had a weird taste to the entire slice . it was a great meal and the tacos were very kind of good . it was super flavorful and had a nice texture of the whole side .
From positive to negative
i love the ladies here ! i avoid all the time ! i hate the doctor here !
my appetizer was also very good and unique . my bf was n't too pleased with the beans . my appetizer was also very cold and not fresh whatsoever .
came here with my wife and her grandmother ! came here with my wife and hated her ! came here with my wife and her son .

Example

Method	accuracy
Hu et al. (2017)	83.5
Variational auto-encoder	23.2
Aligned auto-encoder	48.3
Cross-aligned auto-encoder	78.4

Accuracy by pre-trained classifier

Method	sentiment	fluency	overall transfer
Hu et al. (2017)	70.8	3.2	41.0
Cross-align	62.6	2.8	41.5

Human Evaluation

Appendix

Algorithm 1 Cross-aligned auto-encoder training. The hyper-parameters are set as $\lambda = 1, \gamma = 0.001$ and learning rate is 0.0001 for all experiments in this paper.

Input: Two corpora of different styles $\boldsymbol{X}_1, \boldsymbol{X}_2$. Lagrange multiplier λ, temperature γ.
Initialize $\boldsymbol{\theta}_E, \boldsymbol{\theta}_G, \boldsymbol{\theta}_{D_1}, \boldsymbol{\theta}_{D_2}$
repeat
for $p = 1, 2; q = 2, 1$ **do**
Sample a mini-batch of k examples $\{\boldsymbol{x}_p^{(i)}\}_{i=1}^k$ from $\boldsymbol{X}_p$
Get the latent content representations $\boldsymbol{z}_p^{(i)} = E(\boldsymbol{x}_p^{(i)}, \boldsymbol{y}_p)$
Unroll G from initial state $(\boldsymbol{y}_p, \boldsymbol{z}_p^{(i)})$ by feeding $\boldsymbol{x}_p^{(i)}$, and get the hidden states sequence $\boldsymbol{h}_p^{(i)}$
Unroll G from initial state $(\boldsymbol{y}_q, \boldsymbol{z}_p^{(i)})$ by feeding previous soft output distribution with temperature γ, and get the transferred hidden states sequence $\tilde{\boldsymbol{h}}_p^{(i)}$
end for
Compute the reconstruction $\mathcal{L}_{\text{rec}}$ by Eq. (3)
Compute D_1's (and symmetrically D_2's) loss:

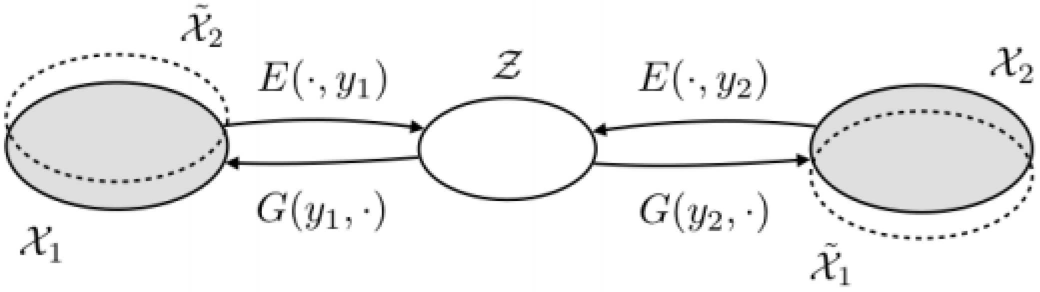
$$\mathcal{L}_{\text{adv}_1} = -\frac{1}{k}\sum_{i=1}^{k} \log D_1(\boldsymbol{h}_1^{(i)}) - \frac{1}{k}\sum_{i=1}^{k} \log(1 - D_1(\tilde{\boldsymbol{h}}_2^{(i)})) \tag{8}$$

Update $\{\boldsymbol{\theta}_E, \boldsymbol{\theta}_G\}$ by gradient descent on loss

$$\mathcal{L}_{\text{rec}} - \lambda(\mathcal{L}_{\text{adv}_1} + \mathcal{L}_{\text{adv}_2}) \tag{9}$$

Update $\boldsymbol{\theta}_{D_1}$ and $\boldsymbol{\theta}_{D_2}$ by gradient descent on loss $\mathcal{L}_{\text{adv}_1}$ and $\mathcal{L}_{\text{adv}_2}$ respectively
until convergence
Output: Style transfer functions $G(\boldsymbol{y}_2, E(\cdot, \boldsymbol{y}_1)) : \mathcal{X}_1 \to \mathcal{X}_2$ and $G(\boldsymbol{y}_1, E(\cdot, \boldsymbol{y}_2)) : \mathcal{X}_2 \to \mathcal{X}_1$

Appendix



$\tilde{\mathcal{X}}_2$
$E(\cdot, y_1)$
$\mathcal{Z}$
$E(\cdot, y_2)$
$\mathcal{X}_2$
$G(y_1, \cdot)$
$G(y_2, \cdot)$
$\mathcal{X}_1$
$\tilde{\mathcal{X}}_1$