

Style Transfer from Non-Parallel Text by Cross-Alignment

NIPS 2017

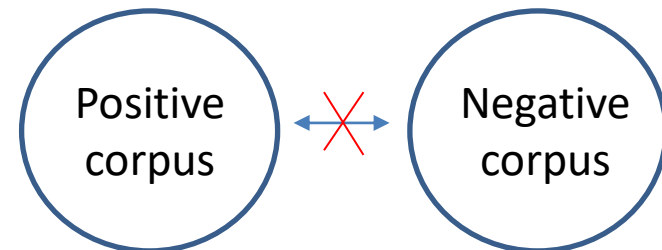
박정수

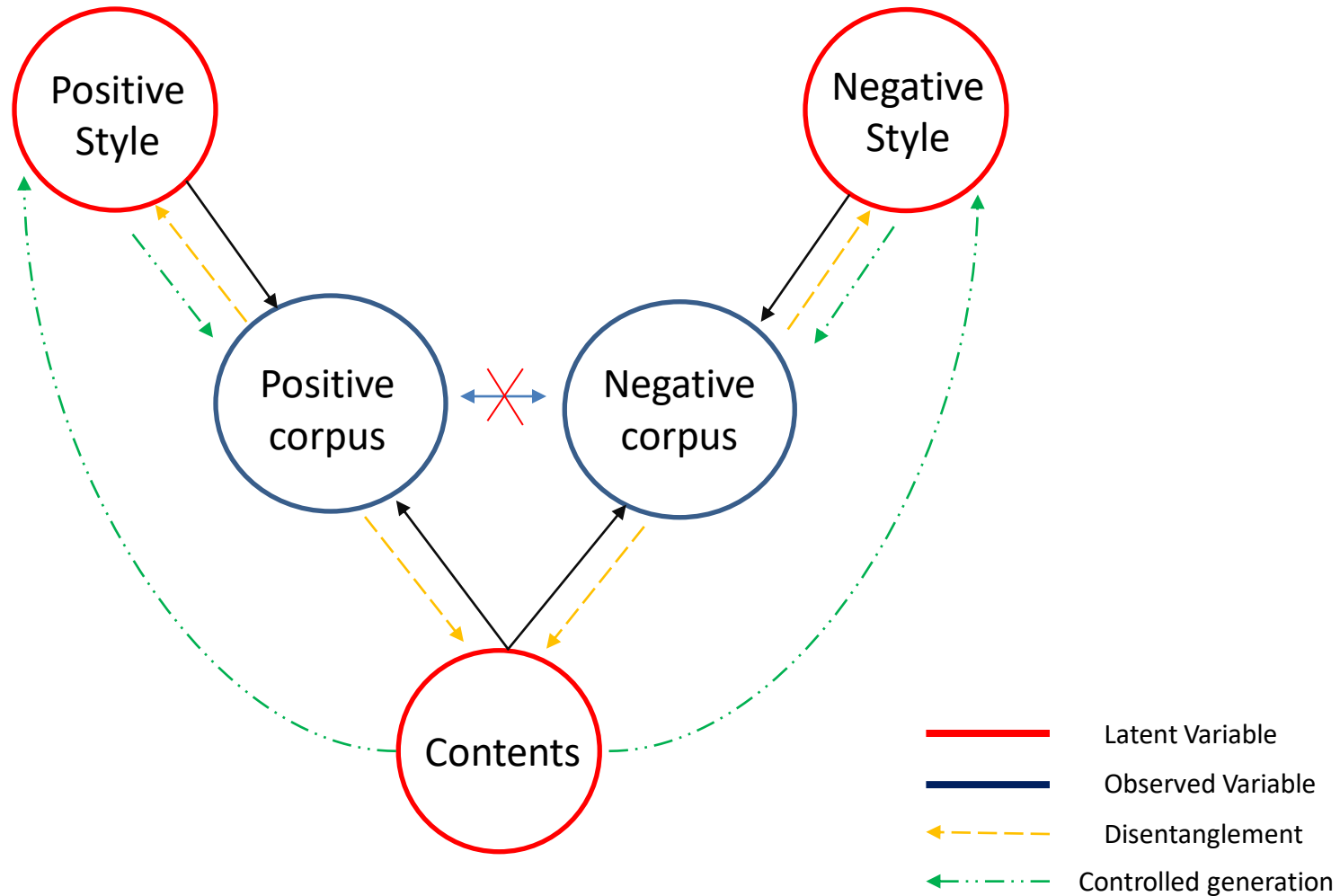
Data Mining & Information Systems Lab.
Department of Computer Science and Engineering,
College of Informatics, Korea University

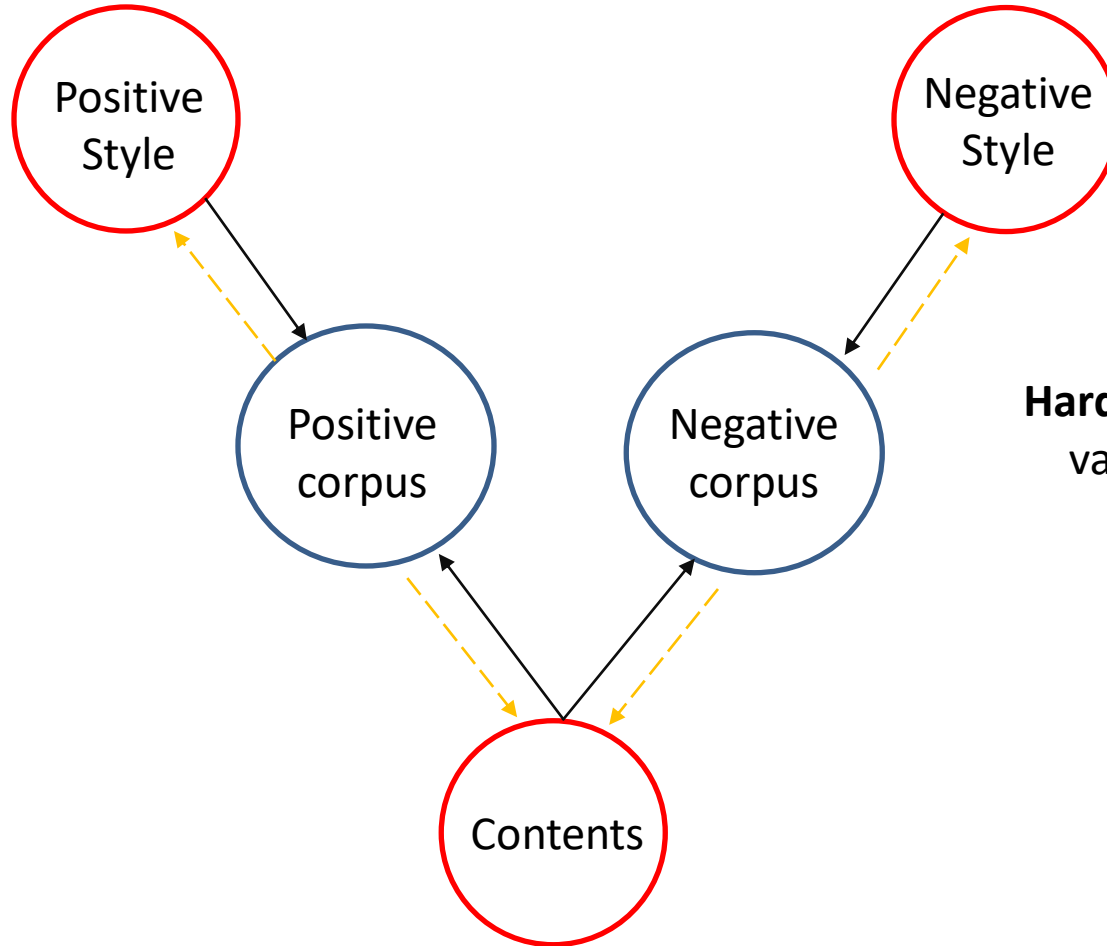
How can we generate text
in a controlled way?



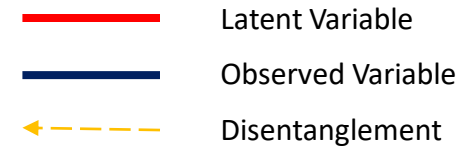
When non-parallel corpus are given

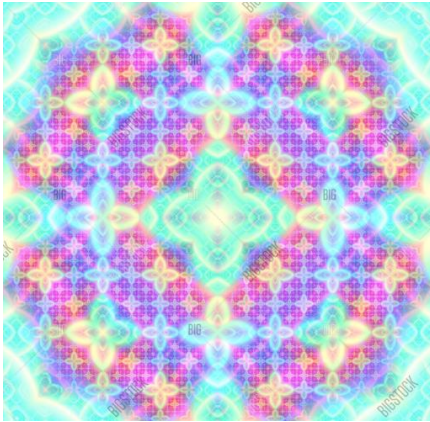






Hard to disentangle style latent variable and content latent variable independently





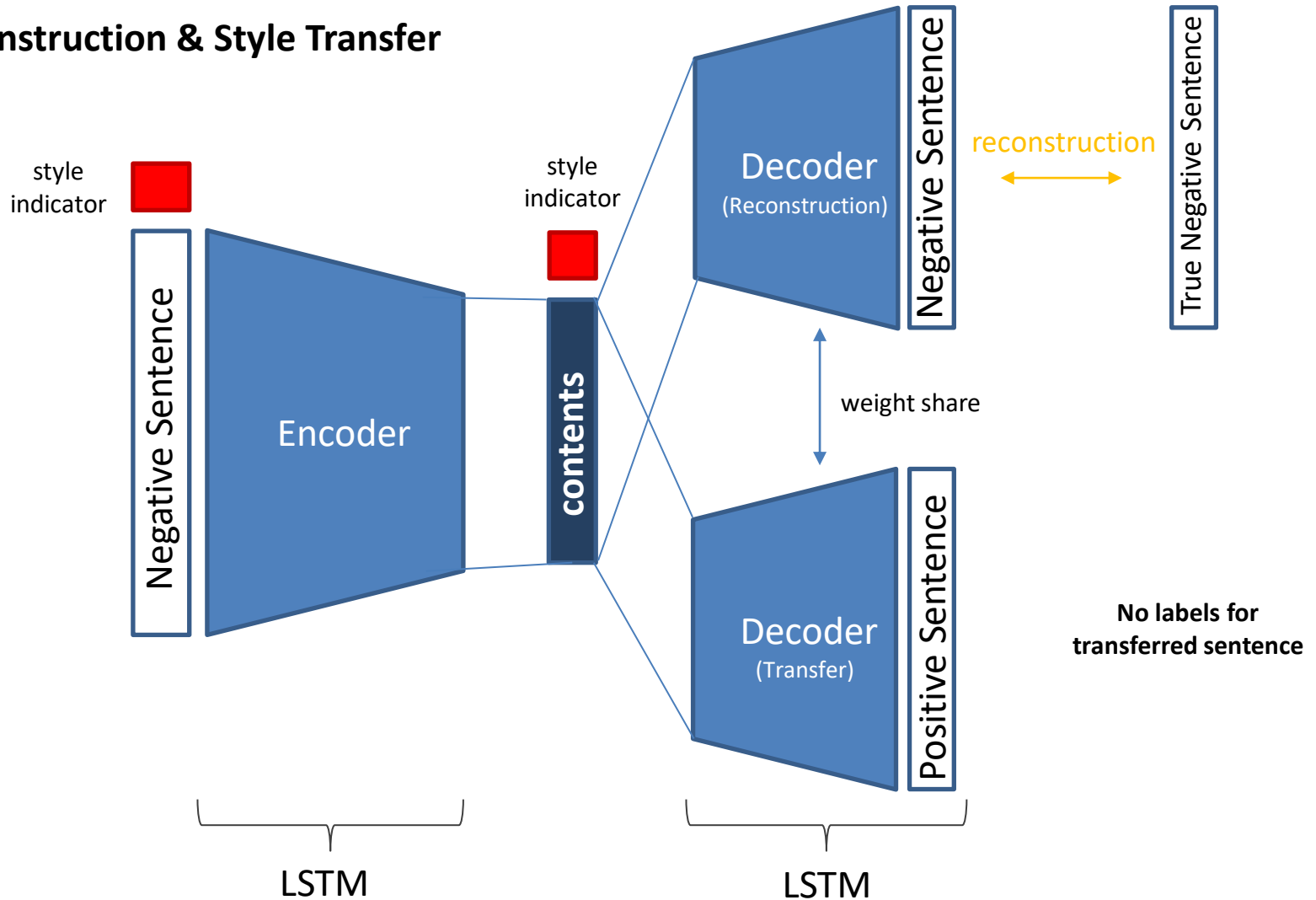
Images: pixel values are continuous



Text: Finite set of characters thus discrete

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

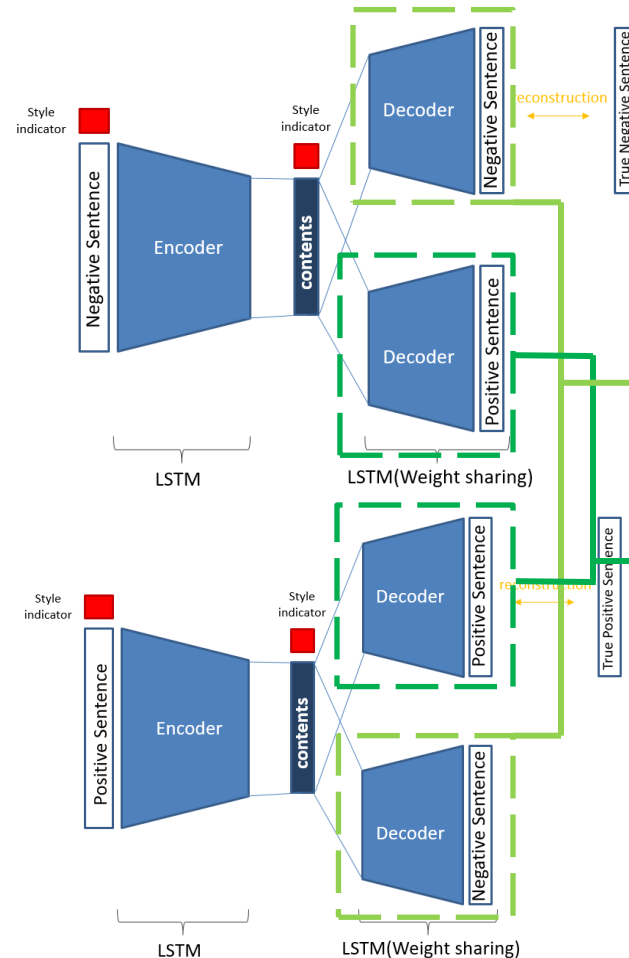
Reconstruction & Style Transfer



General Pipeline – Cross Alignment

Style 1 -> Style 2

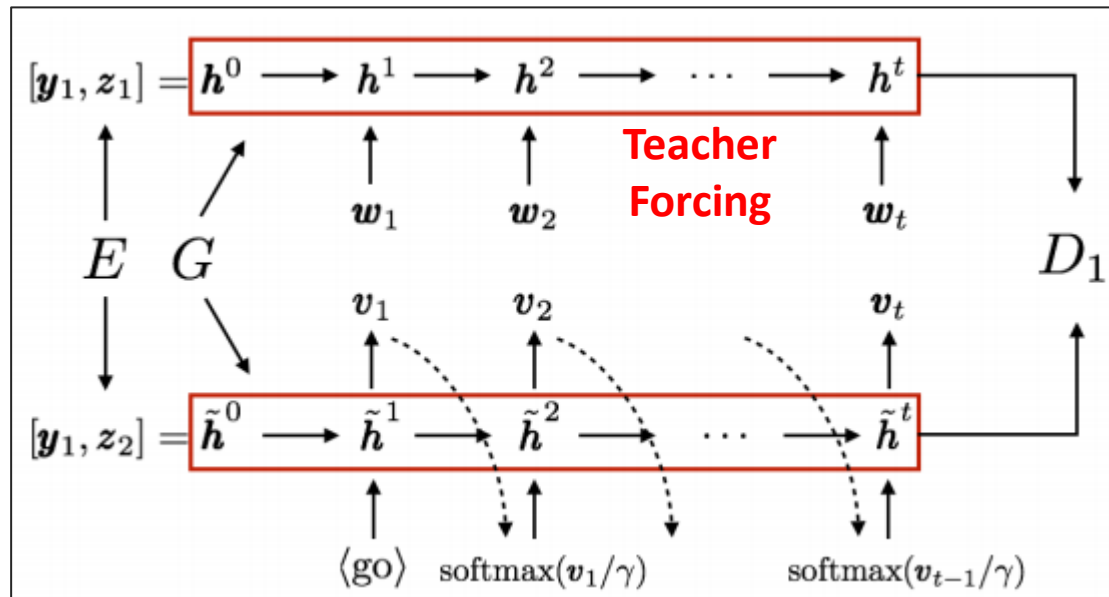
Style 2 -> Style 1



$$\mathcal{L}_{\text{rec}}(\theta_E, \theta_G) = \mathbb{E}_{\mathbf{x}_1 \sim X_1} [-\log p_G(\mathbf{x}_1 | \mathbf{y}_1, E(\mathbf{x}_1, \mathbf{y}_1))] + \mathbb{E}_{\mathbf{x}_2 \sim X_2} [-\log p_G(\mathbf{x}_2 | \mathbf{y}_2, E(\mathbf{x}_2, \mathbf{y}_2))]$$

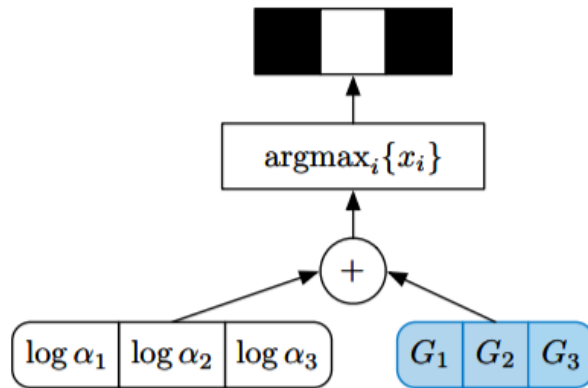
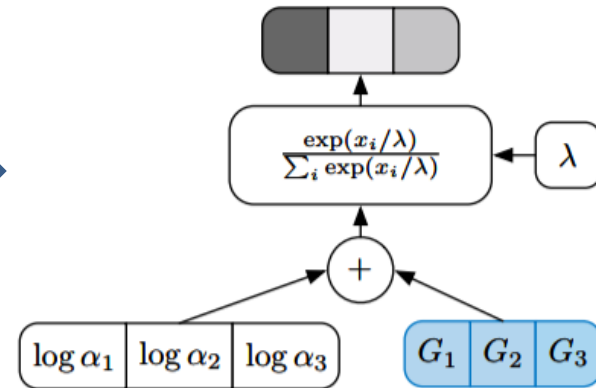
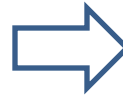
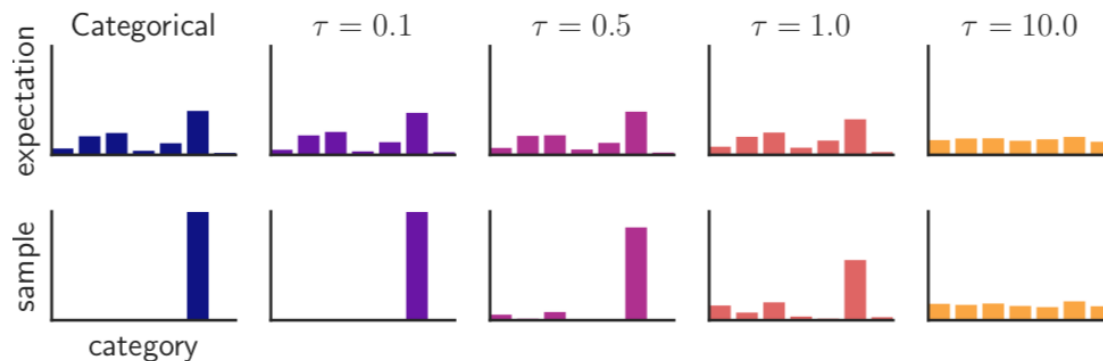
$$\mathcal{L}_{\text{adv}}(\theta_E, \theta_D) = \mathbb{E}_{\mathbf{x}_1 \sim X_1} [-\log D(E(\mathbf{x}_1, \mathbf{y}_1))] + \mathbb{E}_{\mathbf{x}_2 \sim X_2} [-\log(1 - D(E(\mathbf{x}_2, \mathbf{y}_2)))]$$

Generator



- Adopts **Professor Forcing algorithm** which makes generation process' hidden states similar to the behavior of answer fed one's (Teacher Forcing)
- Gumbell-Softmax** approximation is used for relaxation (with Tau parameter)

Gumbel-Softmax

(a) Discrete(α)(b) Concrete(α, λ)

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 $\mathbf{X}_1, \mathbf{X}_2$. Lagrange multiplier λ , temperature γ .

Initialize $\theta_E, \theta_G, \theta_{D_1}, \theta_{D_2}$

repeat

for $p = 1, 2; q = 2, 1$ **do**

 Sample a mini-batch of k examples $\{\mathbf{x}_p^{(i)}\}_{i=1}^k$ from \mathbf{X}_p

 Get the latent content representations $\mathbf{z}_p^{(i)} = E(\mathbf{x}_p^{(i)}, \mathbf{y}_p)$

 Unroll G from initial state $(\mathbf{y}_p, \mathbf{z}_p^{(i)})$ by feeding $\mathbf{x}_p^{(i)}$, and get the hidden states sequence $\mathbf{h}_p^{(i)}$

 Unroll G from initial state $(\mathbf{y}_q, \mathbf{z}_p^{(i)})$ by feeding previous soft output distribution with temperature γ , and get the transferred hidden states sequence $\tilde{\mathbf{h}}_p^{(i)}$

end for

 Compute the reconstruction \mathcal{L}_{rec} by Eq. (3)

 Compute D_1 's (and symmetrically D_2 's) loss:

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

 Update $\{\theta_E, \theta_G\}$ by gradient descent on loss

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

 Update θ_{D_1} and θ_{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 $G(\mathbf{y}_2, E(\cdot, \mathbf{y}_1)) : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ and $G(\mathbf{y}_1, E(\cdot, \mathbf{y}_2)) : \mathcal{X}_2 \rightarrow \mathcal{X}_1$

Appendix

