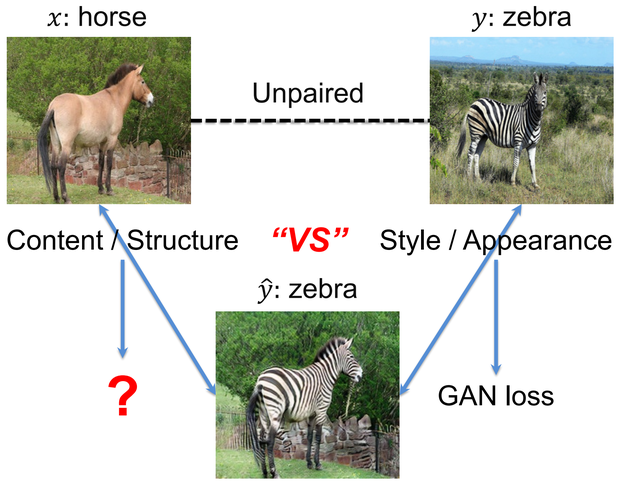
The Spatially-Correlative Loss for Various Image Translation Tasks

Chuanxia Zheng[1], Tat-Jen Cham[1], Jianfei Cai[2]
[1]School of Computer Science and Engineering, Nanyang Technological University [2]Department of Data Science & AI, Monash University

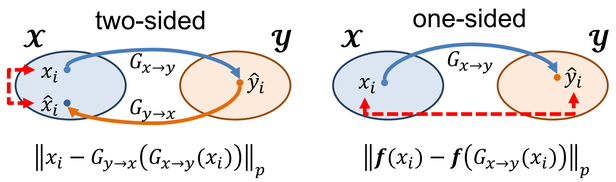
CVPR VIRTUAL JUNE 19-25

Motivation

Goal: Image-to-Image Translation



How to explicitly model the structure?



Issues:

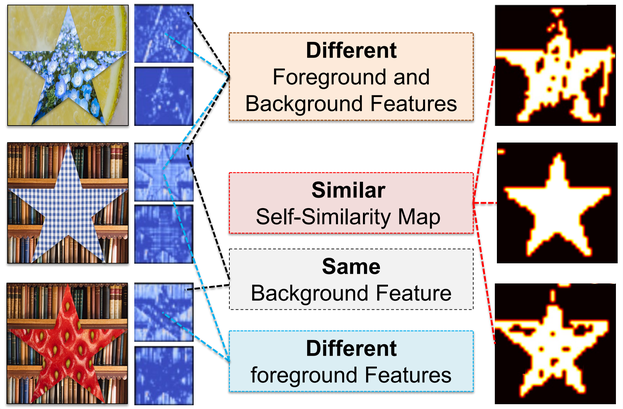
1. **Cycle** Loss
   - Lack explicit structure constraint in the target domain, unwanted content
   - Auxiliary generator and discriminator
2. **Pixel**-level or **Feature**-level Loss
   - *Entangled* structure and appearance
   - Unsuitable for large domain translation

Approach: *Disentangle* structure and appearance

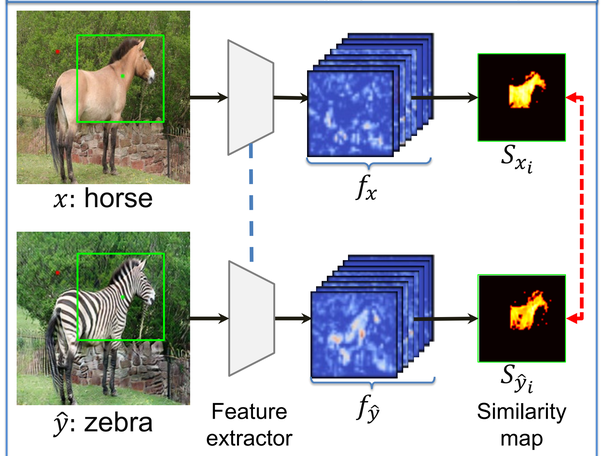
More Source: Project and Code

Key Insights

1. Scene structure is expressed as **repeated patterns** of self-similarity
2. Compare spatial patterns of ***co-occurring*** signals, regardless of ***absolute response***



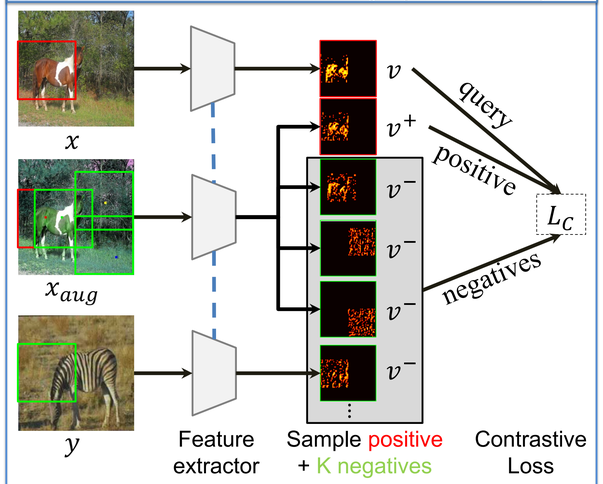
Fixed Self-Similarity (FLSeSim)



Self-Similarity Map: $S_{x_i} = (f_{x_i})^T (f_{x_i})$

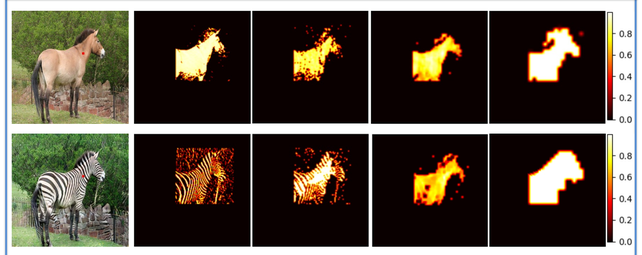
Spatially Correlative Loss: $L_s = d(S_x, S_y)$

Learned Self-Similarity (LSeSim)

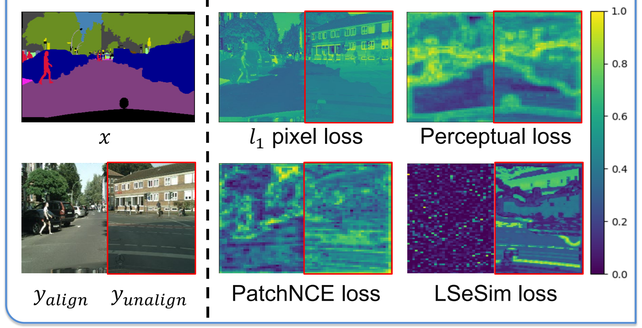


$$L_C = -\log \frac{e^{sim(v,v^+)/\tau}}{e^{sim(v,v^+)/\tau} + \sum e^{sim(v,v^-)/\tau}}$$

Analysis: FSeSim in Layers

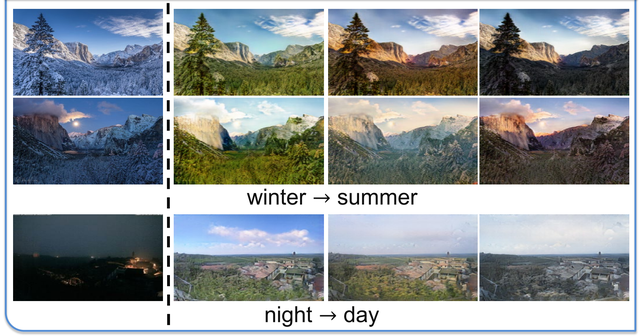


Analysis: Content Loss Visualization



Single-Modal Image-to-Image Translation

Multi-Modal I2I Translation



Single Image Translation