

One-shot unsupervised Cross Domain Translation (2018)

Sagie Benaim, Lior Wolf. advances in neural information processing systems conference on computer vision and pattern recognition. 2018.

IIE8557-01 동적계획법과 강화학습

시스템 인텔리전스 연구실 김민식

Table of Contents

- Background
- Introduction
- One-shot translation
- Experiment
- Conclusion
- Appendix

Few-shot learning

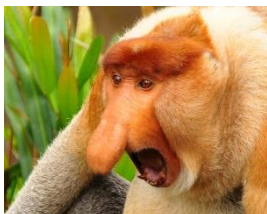
- 딥러닝 학습 vs 사람의 인지과정 차이점
 - 딥러닝: 많은 양의 데이터와 라벨이 필요, 대부분의 지도 학습 방식 채택(비용, 시간)
 - 인지과정: 데이터의 특징을 보고 구분하는 방법을 배움
- Meta learning: 구분 하는 방법[학습하는 방법]을 학습하는 것(Learning to learn)
- Few-shot learning: Meta learning의 일종, 적은 양의 데이터를 가지고 좋은 성능을 얻기 위한 연구분야

Support Set

긴코원숭이



안경원숭이



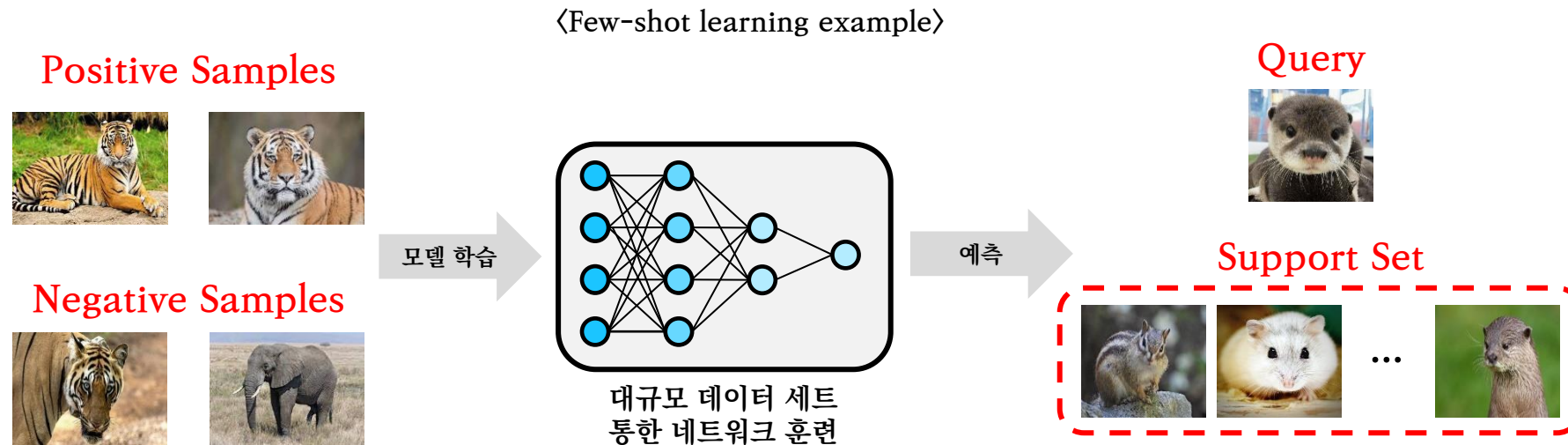
Query



Which one is it?

Few-shot learning

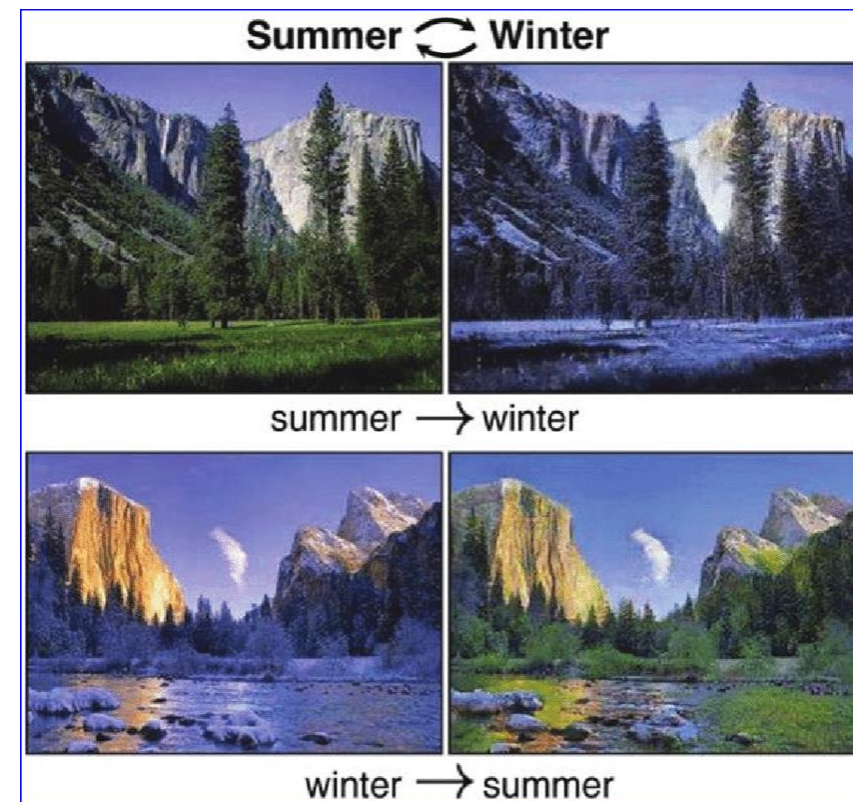
- Few-shot 목적
 - 매우 적은 데이터(Support set)로 평가 데이터(Query set)를 올바르게 예측하는 것
 - 기본 학습 방법: 유사성을 학습하는 것(Support 데이터와 Query 데이터 사이의 유사한지)
- K-way N-shot
 - 클래스 개수와 샘플 수를 기준으로 Support set 구성(K: 클래스 개수, N: 각 클래스별 샘플 수)
 - **One-shot learning:** 하나의 샘플(K = 1)만 사용하여 새로운 클래스를 학습하는 경우



Domain translation

- Source에서 Target Domain으로 데이터나 모델을 변환(번역)하는 과정
 - 예: 영어 → 스페인어 문장 번역, 여름 → 겨울 기풍의 사진 변화
- 도메인 간 샘플에 대해 **Mapping function** 찾는 것이 핵심
- Domain Adaptation의 한 종류로 볼 수 있음
 - Translation은 Source에서 Target Domain으로 데이터를 변형(이동)시키는 것이 중점
 - Adaptation은 Domain 간의 분포차이를 조정하여, 새로운 환경(Target)에 적응하도록 하는 것이 중점

〈Domain translation example〉



One-shot unsupervised cross domain transition

- Accumulating knowledge
 - 지식(Knowledge)은 이전에 본 샘플과 새로운 샘플을 생성할 수 있는 모델을 통해 획득
 - 새로운 샘플이 관찰되면
 - 1) 관찰된 샘플과 유사한 가상의 샘플을 생성
 - 2) 관찰된 샘플과 생성된 샘플의 객체를 비교하여 모델 업데이트 진행
- One-shot unsupervised domain transition
 - “**알고 있지 않는 것을 어떻게 배울 수 있을까?**”에 대한 문제점에서 시작
 - Accumulating knowledge 과정을 One-shot unsupervised cross domain translation에 적용
 - “**새롭게 관찰된 샘플은 이전 샘플과 매우 다를 수 있는데, 어떻게 유사한 이미지를 구성할까?**”
 - 알려지지 않는 Domain A의 하나의 샘플 x 가 있을 때, 많은 샘플이 있는 Domain B를 통해 x 와 유사한 Domain B에 맞는 샘플 y 를 생성

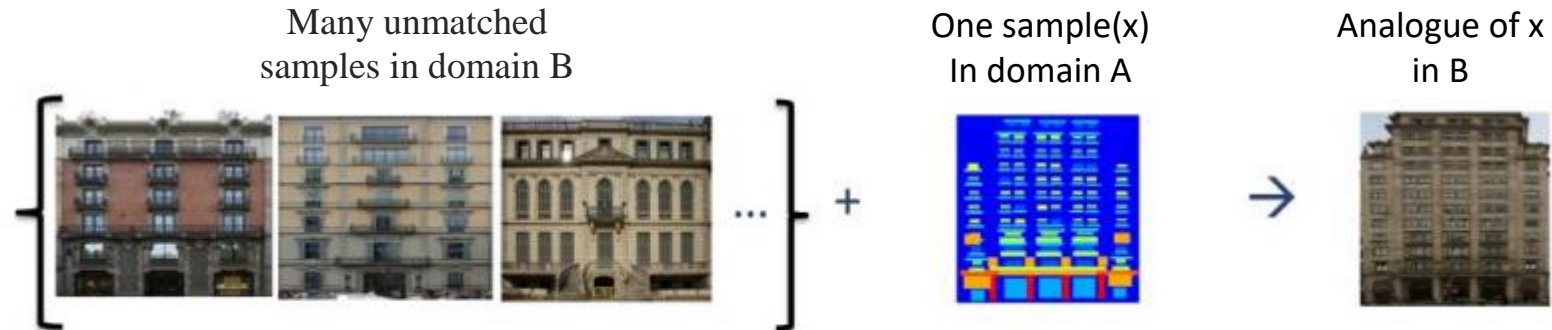
본 논문의 Key Idea

- **Key Idea**

- 학습 과정에 다른 Domain 이미지가 주어지지 않더라도 샘플을 어떻게 Mapping할까?
- 샘플이 많은 Domain을 통해서 샘플이 적은(**One-shot**) Domain와 유사한 샘플을 생성
 - **Selective backpropagation(shared, unshared layer 구분)**

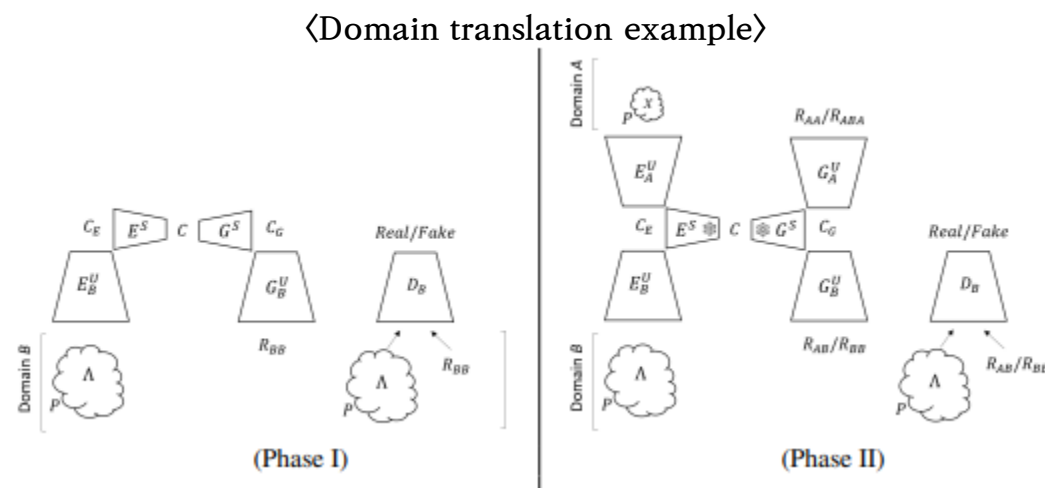
- **Contribution**

- Unsupervised domain translation mechanism에 대한 새로운 관점
- Low shot 상황(샘플이 적은 상황)에서 기존 알고리즘 보다 높은 성능을 보임



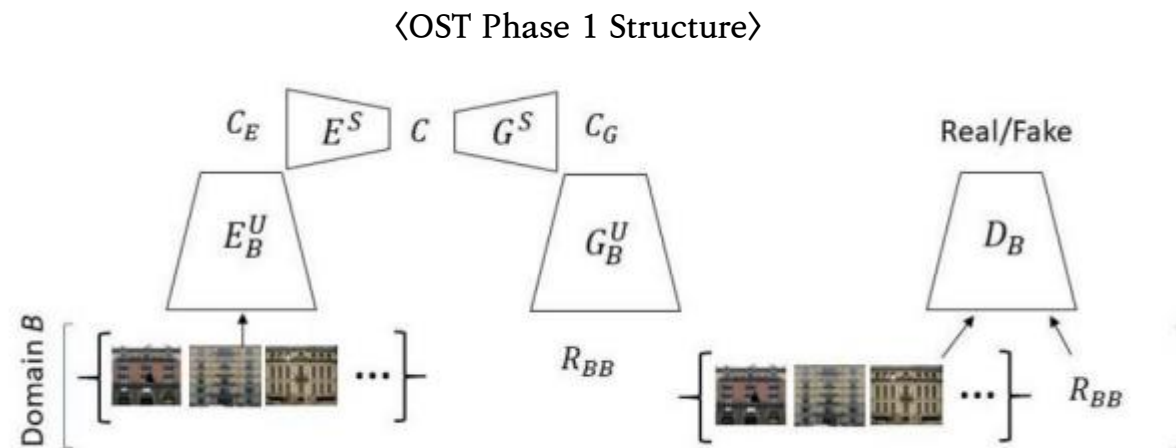
One-shot translation(OST)

- Unsupervised cross-domain translation 문제(두 Domain에 대한 Unlabeled dataset)
- (오른쪽 그림과 같이) 2단계를 통해서 Domain translation 진행
- Encoder의 최상단 부분(E^S), Decoder의 최하단 부분(G^S)을 공유 레이어로 사용
 - Encoder 하단 부분, Decoder 상단 부분은 Domain-specific 특성이 있어 공유하지 않음
 - 공유 레이어가 **Transition** 과정에서 강한 규제 역할(Domain A가 Domain B의 transition 후 style, content 가질 수 있도록 함)
- 추가적으로 Domain B의 모델링 위해 Discriminator 학습 진행
 - Domain B Distribution에 속하는 이미지가 생성되도록 함



Phase 1: Domain B 통한 VAE 모델 구축

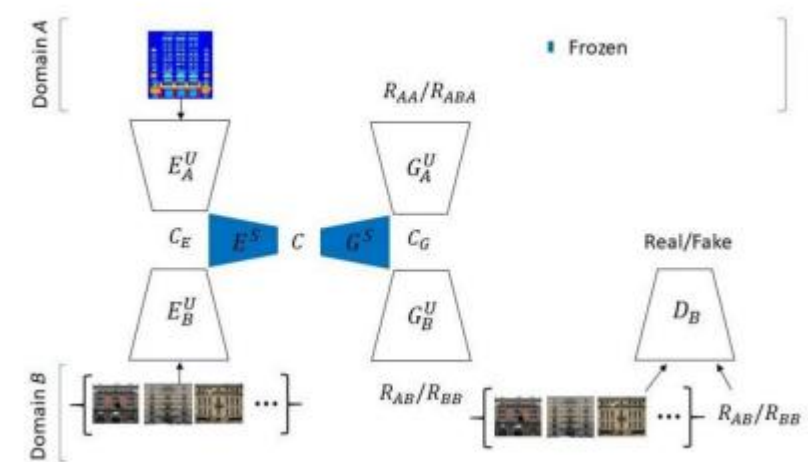
- Domain B에 대한 샘플을 효과적으로 Encoding 진행,
Latent space vector에서 새로운 샘플 생성 가능
- 일반화 성능을 위해 데이터 증강(Data Augmentation) 적용:
수평 이동 + 약간의 회전



Phase 2: 공유 레이어를 통한 Domain transition

- Encoder 상단 레이어와 Decoder 하단 레이어의 Weight를 공유하는 VAE 복사본 생성
 - Phase 1에서 학습된 Encoder/Decoder를 Domain A VAE의 초기 Weight
 - Domain B의 샘플의 경우: 전체 레이어(E_B^U, G_B^U, E^S, G^S)의 Gradient 업데이트
 - Domain A의 샘플의 경우: 비공유 레이어(E_A^U, G_A^U)에 대해서 Gradient 업데이트
- 공유 Layer는 Domain B의 샘플과 Domain의 샘플(x) 동일한 구조 적용
 - Domain B만으로 모델을 훈련시키는 것이 Domain A 샘플 feature 특성을 포함한다고 가정
 - 공유 Layer를 통해서 Domain transition 간의 보정이 이루어짐
- Selective backpropagation의 역할
 - Domain B 샘플을 기반으로 얻은 Representation을 Domain A의 적용시키는 것
 - Selective backpropagation 끝 경우, Domain A의 representation이 Overfitting됨
 - Selective backpropagation freezing할 경우, Mapping이 제대로 진행되지 않음

〈OST Phase 2 Structure〉



Experiment (1) MNIST-SVHN Translation

- [Domain A: MNIST, Domain B: SVHN] Target Image에 대한 분류 정확도 및 Ablation Study 진행
- 제안된 OST이 **Baseline**을 능가하는 성능을 보임
 - Domain Sample의 수가 증가함에 따라서 정확도 증가하는 경향
- Ablation study 통해서 가장 큰 영향을 준 요소는 “**Selective Propagation**”
 - 추가 실험(Shared Encoder & Decoder freezing 할 경우) Target domain distribution 이미지 생성할 수 없었음([Domain transition의 핵심 요소](#))

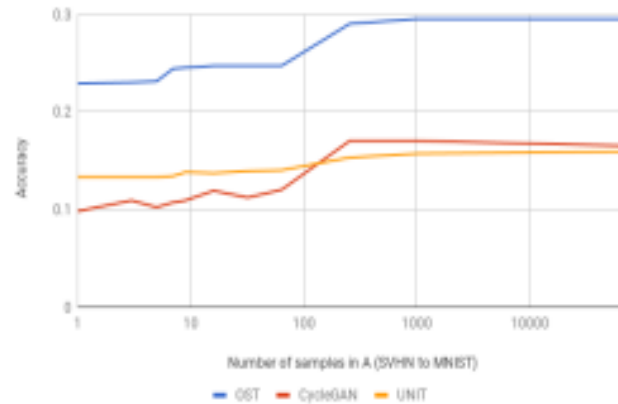
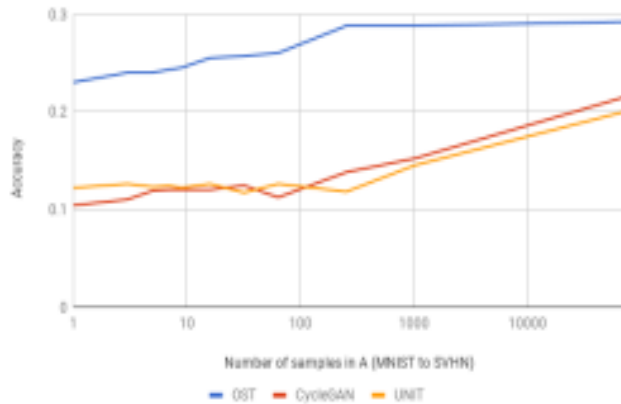


Table 1: Ablation study for the MNIST to SVHN translation (and vice versa). We consider the contribution of various parts of our method on the accuracy. Translation is done for one sample.

Augment- ation	One-way cycle	Selective backprop	Accuracy (MNIST to SVHN)	Accuracy (SVHN to MNIST)
False	False	False	0.07	0.10
True	False	False	0.11	0.11
False	True	False	0.13	0.13
True	True	False	0.14	0.14
False	False	True	0.19	0.20
True	False	True	0.20	0.20
False	True	True	0.22	0.23
True	True	No Phase II update of E^S and G^S	0.16	0.15
True	Two-way cycle	True	0.20	0.13
True	Two-way cycle	False	0.11	0.12
True	True	True	0.23	0.23

〈MNIST-SVHN Transition 실험 결과: (왼) sample 수에 따른 정확도 (오) Ablation 실험〉

Experiment (2) Style transfer task

- Perceptual distance: Input과 대응되는 다른 transfer style의 이미지 Output 간의 거리 기반
 - 낮은 값을 가질수록 Style transfer task에서 “내용이 많이 보존된다는 것”을 의미
- 제안된 OST Model의 경우, Cycle GAN과 UNIT와 유사한 성능을 보임
 - 단일 샘플로 학습한 Cycle GAN의 경우 **Target domain의 style은 학습하지만 내용은 보존하지 못하는 것을 확인**

Component	Dataset Samples in A	OST 1	UNIT [7] 1	CycleGAN [2] 1	UNIT [7] All	CycleGAN [2] All
(i) Content	Summer2Winter	0.64	3.20	3.53	1.41	0.41
	Winter2Summer	0.73	3.10	3.48	1.38	0.40
	Monet2Photo	3.75	6.82	5.80	1.46	1.41
	Photo2Monet	1.47	2.92	2.98	2.01	1.46
(ii) Style	Summer2Winter	1.64	6.51	1.62	1.69	1.69
	Winter2Summer	1.58	6.80	1.31	1.69	1.66
	Monet2Photo	1.20	6.83	0.90	1.21	1.18
	Photo2Monet	1.95	7.53	1.91	2.12	1.88



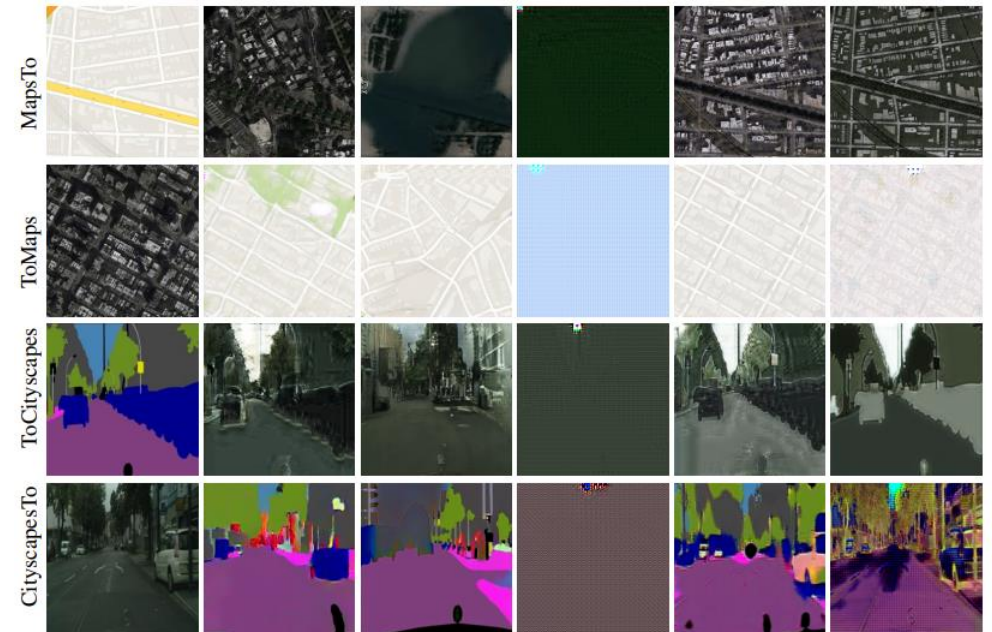
〈Style transfer task 실험 결과: (왼) 실험 결과 표 (오) 실제 이미지 사진〉

Experiment (3) Drawing task

- 단일 샘플로 학습된 OST은 전체 샘플로 학습된 Cycle GAN, UNIT와 유사한 **Perceptual distance**와 **Style difference** 얻음
 - 단일 샘플로 학습된 Cycle GAN, UNIT의 경우는 **Target distribution의 샘플을 생성하지 못함**
- 추가적 검증을 위해 **Source** 이미지와 **Target** 이미지에 대한 일치 여부에 대한 사용자 평가를 진행(%)

Method	Images to Facades	Facades to Images	Images To Maps	Maps to Images	Labels to Cityscapes	Cityscapes to Labels
(i) OST 1	4.76	5.05	2.49	2.36	3.34	2.39
UNIT [7] All	3.85	4.80	2.42	2.30	2.61	2.18
CycleGAN [2] All	3.79	4.49	2.49	2.11	2.73	2.28
(ii) OST 1	3.57	7.88	2.24	1.50	0.67	1.13
UNIT [7] All	3.92	7.42	2.56	1.59	0.69	1.21
CycleGAN [2] All	3.81	7.03	2.33	1.30	0.77	1.22
(iii) OST 1	91%	90%	83%	67%	66%	56%
UNIT [7] ALL	86%	83%	81%	75%	63%	37%
CycleGAN [2] ALL	93%	84%	97%	81%	72%	45%

(i) Perceptual distance (ii) Style difference (iii) User evaluation



〈Drawing task 실험 결과: (원) 실험 결과 표 (오) 실제 이미지 사진〉

Conclusion

- 본 논문의 방법은 이전에 본 적 없는 Domain에서 샘플이 많은 Domain으로 Mapping 할 수 있음을 보여줌
 - 이전에 본 적 없는 Domain의 이미지를 만날 수 있는 Agent에 적합한 방법
- 이러한 방식은 새로운 Domain을 순차적으로 만나는 현실세계의 지식 축적과 매우 유사
 - 평생 학습(Life-long) 관점에서 Unsupervised learning 활용한 방안이 적합함

Q & A

One shot transition loss function

⟨Phase1: Loss function⟩

$$\mathcal{L}_{REC_B} = \sum_{s \in P(\Lambda)} \|G_B(E_B(s)) - s\|_1 \quad (1)$$

$$\mathcal{L}_{VAE_B} = \sum_{s \in P(\Lambda)} \text{KL}(E_B \circ P(\Lambda) \parallel \mathcal{N}(0, I)) \quad (2)$$

$$\mathcal{L}_{GAN_B} = \sum_{s \in P(\Lambda)} -\ell(\overline{D}_B(G_B(E_B(s))), 0) \quad (3)$$

$$\mathcal{L}_{D_B} = \sum_{s \in P(\Lambda)} +\ell(D_B(\overline{G}_B(\overline{E}_B(s))), 0) + \ell(D_B(s), 1) \quad (4)$$

⟨Phase2: Loss function⟩

$$\mathcal{L}_{REC_A} = \sum_{s \in P(x)} \|T_{AA}(s) - s\|_1 \quad (10)$$

$$\mathcal{L}_{\text{cycle}} = \sum_{s \in P(x)} \|T_{BA}(T_{AB}(s)) - s\|_1 \quad (11)$$

$$\mathcal{L}_{GAN_{AB}} = \sum_{s \in P(x)} -\ell(\overline{D}_B(T_{AB}(s)), 0) \quad (12)$$

$$\mathcal{L}_{D_{AB}} = \sum_{s \in P(x)} +\ell(D_B(\overline{T}_{AB}(s)), 0) + \ell(D_B(s), 1) \quad (13)$$