

Learning Shape Abstractions by Assembling Volumetric Primitives

2023.05.19

- 본 논문은 복잡한 3D Shape을 추상화 하는 방법론을 제시.
- 1990년도 이전에는 cylinder를 이용해 3D 물체를 추상화 하려는 연구가 진행되었음. 그러나 cylinder를 이미지 데이터와 맞추는데 수작업이 많이 드는 문제로 인해 이 방법은 자취를 감춤.
- 동시에 머신러닝을 이용한 3D 추상화가 1990년도에 등장하였고, 이 논문에서는 Unsupervised Learning과 CNN을 사용해 각 Shape을 추상화 함.
- 3D Shape의 전반적인 패턴을 학습하는데 도움을 줄 수 있다고 밝히고 있음.

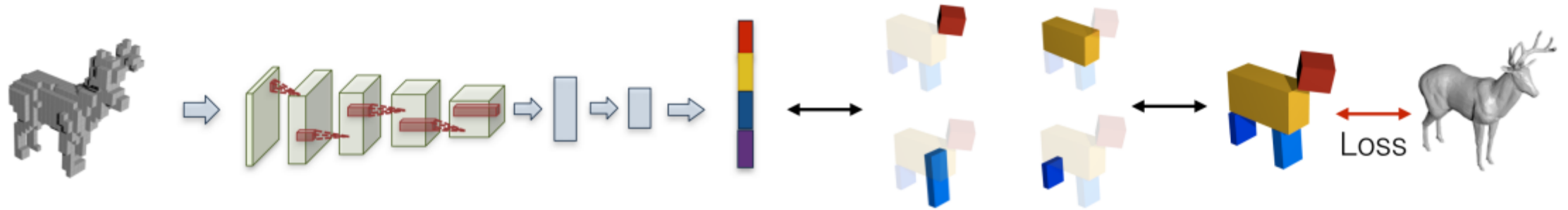


Figure 2: Overview of our approach. Given the input volume corresponding to an object O , we use a CNN to predict primitive shape and transformation parameters $\{(z_m, q_m, t_m)\}$ for each part (Section 3.1). The predicted parameters implicitly define transformed volumetric primitives $\{\bar{P}_m\}$ whose composition induces an assembled shape. We train our system using a loss function which attempts to minimize the discrepancy between the ground-truth mesh for O and the assembled shape which is implicitly defined by the predicted parameters (Section 3.2).

- 입력 신호 I 를 받아들여 목표 객체 O 를 조립하는 문제로 정의
- CNN 모델을 학습하여 최종 모양을 출력하는데 필요한 최대 M 개의 서로 다른 Primitive를 예측
- Unsupervised learning을 수행하며, 목표 객체와 가장 유사한 Primitive 구성에 대한 정보는 미리 제공되지 않음
- 예측된 Primitive 구성을 사용하여 형태를 조립한 경우 목표 형태와 일치하는지를 알려주는 loss 함수를 정의하고, 이 loss를 최적화하여 CNN을 학습
- Primitive Based Representation
 1. 각 Primitive는 튜플 (z, q, t) 로 구성됨.
 2. z 는 shape, q 와 t 는 각각 Rotation, Translation을 의미.
 3. CNN을 통해 예측된 Shape은 아래와 같이 표현할 수 있음.

$$\{(z_m, q_m, t_m) | m = 1, \dots, M\} = h_\theta(I)$$

- Distance Field

Ground truth는 Mesh이고 Predicted shape은 Paramerized된 모양이기 때문에 비교가 어려움.
이를 극복하고자, 본 논문은 Distance Field 을 계산함. 이를 통해 두 개의 Loss를 정의할 수 있음.

$$\mathcal{C}(p; O) = \min_{p' \in O} \|p - p'\|_2$$

- Coverage Loss(L1)

1. Predicted Shape 안에 Target Object가 포함되지 않는다면 이에 대한 페널티를 주고자하는 Loss 함수이다.
2. 이를 만족하는 충분조건은 Predicted Shape의 Distance Field가 Target object 표면의 모든 Point에 대해 0인 경우이다.

$$L_1(\{(z_m, q_m, t_m)\}, O) = \mathbb{E}_{p \sim S(O)} \|\mathcal{C}(p; \cup_m \bar{P}_m)\|^2$$

- Consistency Loss(L2)

1. Coverage Loss와 반대이다.
2. Target object 안에 Predicted Shape가 완벽히 들어가지지 않았을 때 페널티를 주는 Loss 함수이다.
3. 이를 만족하는 충분조건은 Target object 의 Distance Field가 Predicted Shape의 개별 부품 표면의 모든 Point에 대해 0인 경우이다.

$$L_2(\{(z_m, q_m, t_m)\}, O) = \sum \mathbb{E}_{p \sim S(\bar{P}_m)} \|\mathcal{C}(p; O)\|^2$$

- Primitive Representation

1. Primitive의 개수를 결정한다.
2. p_m 은 Primitive의 존재 가능성을 보여주는 지표
3. Z_m 는 Primitive's Dimensions을 결정하는 z_m^s 와 Bernoulli Distribution을 따르는 z_m^e 로 구성
4. 실제로 $z_m^e=0$ 이라면 m 번째 Primitives는 존재하지 않는 것으로 간주

$$\{(z_m^s, q_m, t_m, p_m) | m = 1 \cdots M\} = \mathbf{h}_\theta(I)$$
$$\forall m \ z_m^e \sim \text{Bern}(p_m); \ z_m \equiv (z_m^s, z_m^e)$$

5. learning

- 1) final loss 함수 $L(h_\theta(I), O)$ 를 정의하여 사용
- 2) p_m 에 따라 가능한 z_m^e 샘플링으로 final loss가 산출된다.
- 3) $L(h_\theta(I), O) = E_{\forall m \ z_m^e \sim \text{Bern}(p_m)} L(\{(z_m, q_m, t_m)\}, O)$
- 4) $L(\{(z_m, q_m, t_m)\}, O) = L1(\{(z_m, q_m, t_m)\}, O) + L2(\{(z_m, q_m, t_m)\}, O)$
- 5) 계산을 위해 z_m^s 와 z_m^e 를 z_m 로 통합해야 한다.

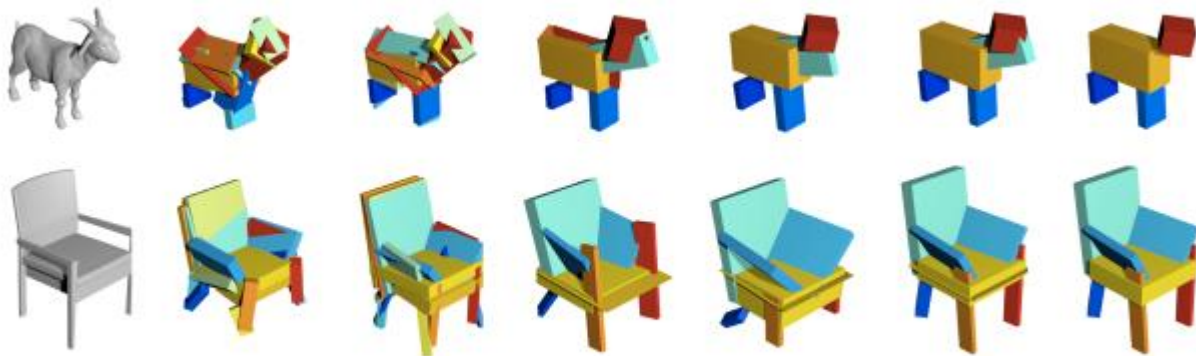
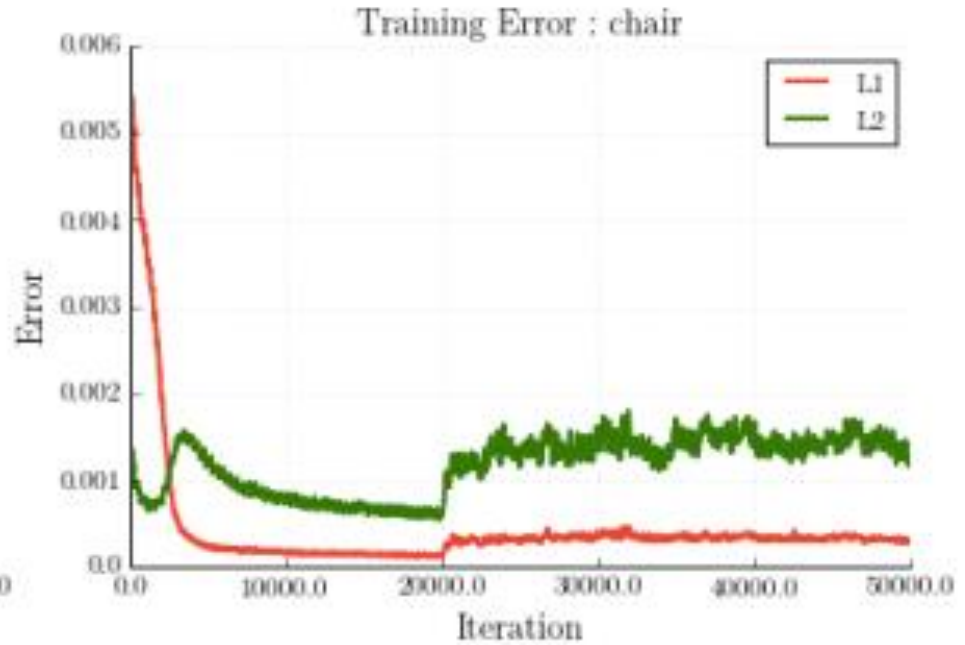
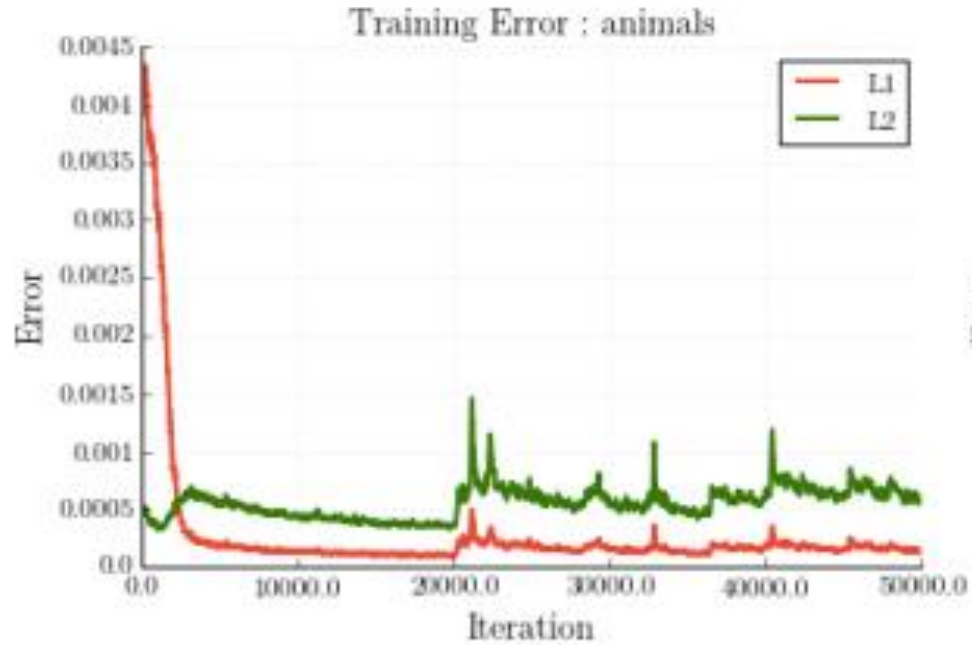


Figure 4: Visualization of the training progression. We visualize the prediction for two instances (shown in column 1) after every 10,000 iterations (left to right, in columns 2-6). The last column shows the result after post-processing to remove redundant parts that overlap significantly with others. The initial training stage (up to 20,000 iterations) uses all primitives but we later allow the network to learn to use fewer primitives and the predictions gradually become more parsimonious.

- Data set : sharp net dataset
- Airplane, chair object category 사용
- 10000번 훈련마다 추상화한 객체 출력
- 마지막 훈련 단계에서 겹치는 Primitive를 제거
- 첫 훈련에서 모든 Primitive를 사용했지만, 점점 적은 Primitive를 사용하도록 학습됨
- 더 적은 Primitive로 객체를 표현할 수 있음



- 반복에 따른 모델의 Loss 값들을 시각화한 그래프이다.
- 초기 훈련 단계(최대 20000회 반복)에서는 부품의 올바른 구성이 학습됨에 따라 손실이 급격하게 감소함
- 20000회 반복이 넘어가는 시점에서 갑자기 튀어오르는 현상을 확인할 수 있음. 이 시점부터 일부 Primitive가 사라지고, 모델이 더 적은 수의 Primitive를 더 잘 사용하는 방법을 학습하기 때문임.