

Unsigned Distance Field를 이용한 Sketch Classification

*김민우, 조남익

서울대학교 전기정보공학부, 뉴미디어통신공동연구소

*mwk0614@ispl.snu.ac.kr, nicho@snu.ac.kr

Sketch Classification using Unsigned Distance Field

*Min Woo Kim, Nam Ik Cho

Department of ECE, INMC, Seoul National University

요약

본 논문에서는 스케치를 unsigned distance field로 변환하여 스케치 클래스 분류 네트워크의 입력으로 사용한다. 그리고 unsigned distance field scaling factor를 제안하여, unsigned distance field에 보존되는 스케치의 전역적인 정보와 국소적인 정보 사이에 상호조정이 가능하게 하였다. 다양한 scaling factor 값에 대해서 실험을 진행하여, 기존 unsigned distance field보다 국소적인 정보가 더 포함되어 있을 때 클래스 분류 성능이 향상되는 것을 확인하였다. 또한 스케치를 고밀도 데이터로 변환하여 사용했을 때 학습이 더 안정적으로 되고, 네트워크가 더욱 합리적인 근거로 스케치를 올바른 클래스로 분류한다는 것을 Smooth Grad-CAM++를 통해서 확인하였다.

1. 서론

스마트폰과 터치패드가 광범위하게 공급되면서, 사람들은 아이디어나 생각을 시각적으로 표현하고 기록하기 위해 스케치를 사용한다. 스케치는 간결하고 추상적이지만 어떤 대상에 대한 직관적인 설명이 가능하기 때문에, 이에 기반한 컴퓨터비전 활용 분야가 많이 연구되고 있다.

합성곱 신경망을 사용한 네트워크가 발전하면서, '스케치 기반 이미지 검색, 스케치로부터의 이미지 합성'과 같은 스케치 활용 분야에 관한 연구는 큰 진전을 이루었다. 하지만 스케치의 특성은 응용에 여전히 어려움을 초래한다. 예를 들면 같은 물체더라도, 그 물체를 그리는 사람에 따라서 스케치가 달라져서 네트워크에 의해 추출된 특성 간 차이가 크다. 그리고 흑백으로만 이루어진 스케치는 RGB 이미지와 색채 및 질감 특성이 매우 달라서, RGB 이미지 도메인과 융합된 활용 분야에서는 큰 도메인 차이 때문에 좋은 성능을 내기가 쉽지 않다. 마지막으로, 스케치는 색채나 질감이 빠져 있는 희소 데이터(sparse data)이므로, 네트워크가 학습할 때 시각적 단서를 얻는 데 어려움이 있다.

이러한 스케치의 특성에 의한 한계를 극복하기 위해 여러 방법이 연구되었다. [1]에서는 스케치의 클래스를 잘 분류하기 위해 네트워크를 학습시키는 과정에서 RGB 이미지를 보조적인 수단으로 활용하였다. 스케치와 RGB 이미지를 '공유 잠재 공간'으로 임베딩 시키고, RGB 이미지의 임베딩 벡터를 스케치의 임베딩 벡터를 학습에 사용하였다. 같은 클래스에 속한 임베딩 벡터는 서로 가깝게, 다른 클래스에 속한 임베딩 벡터는 서로 멀게 만들어져서, 클래스 분류 성능을 향상시켰다. [2]에서는 '스케치 기반 이미지 검색'의 성능을 높이기 위해 '스케치 토큰(sketch token)'을 사용할 것을 제안하였다. RGB 이미지를 스케치 토큰으로 변환하여, 스케치와 RGB 이미지 사이의 도메인 차이를 줄인 후 학습을 진행하였다. 하지만 위 두 방법 모두 스케치에 기반한 응용을 위해, RGB 이미지 혹은 스케치 토큰 등 스케치가 본질이 아닌 데이터의 도움을 받았고, 스케치 관련 응용에 어려움을 초래하는 원인 중 하나인 희소 데이

터에 대한 개선이 이루어지지 않는 것이다.

위와 다르게, [3]에서는 스케치의 희소성을 극복하기 위한 시도가 이루어졌다. 스케치에 대한 unsigned distance field를 연산하여, 스케치를 희소 데이터에서 고밀도 데이터(dense data)로 변환하였다. 하지만 [3]에서 적용한 unsigned distance field는 스케치의 국소적인 특성은 잘 보존하지 못한다는 한계가 있다.

본 논문에서는 [3]에서처럼 스케치로부터 unsigned distance field를 계산하였다. 즉, 희소 데이터인 스케치를 고밀도 데이터로 변환하였고, 이를 클래스 분류 네트워크 학습에 사용하였다. 추가로, [3]에서 사용한 unsigned distance field는 스케치의 국소적인 특성을 잘 보존하지 못한다는 한계를 극복하기 위하여 'unsigned distance field scaling factor(이하 scaling factor)'를 제안한다. 그림 1에서 보는 것과 같이, scaling factor가 작으면 전역적인 특성이 더 강조되고, 크면 국소적인 특성이 강조된다(기존 unsigned distance field는 scaling factor가 1 일 때이다). 즉, scaling factor를 변화시킴으로써 전역적인 특성과 국소적인 특성 중 어떤 특성이 비중이 있게 보존될 것인지 조절할 수 있다. 그 결과, scaling factor가 1인 기존 unsigned distance field보다는 국소적인 특성이 강조될 때, 네트워크가 더 유의미한 특성을 추출한다는 것을 발견하였다.

2. 스케치 클래스 분류 알고리즘

2.1. 스케치를 Unsigned Distance Field로 변환

본 논문에서는 스케치를 고밀도 데이터인 unsigned distance field로 변환하여 네트워크의 입력으로 사용하였다. 스케치의 각 픽셀이 0부터 1까지의 범위로 정규화가 되어 있을 때, 1보다 작은 값을 가지는 픽셀은

0으로 만듦으로써 스케치가 0과 1로만 구성되도록 하였다. Unsigned distance field의 각 픽셀값은, 해당 픽셀에서 픽셀값이 1인 픽셀까지의 거리를 나타낸다.

$$I_i = \min_{p_j} D(p_i, p_j) \text{ for } S_j = 1 \dots \text{수식 (1)}$$

$p_i = (x_i, y_i)$: 픽셀 위치

S_i : 스케치의 p_i 픽셀 강도

I_i : Unsigned distance field의 p_i 픽셀 강도

$$D(p_i, p_j) = \|p_i - p_j\|_2 = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

수식 (1)에서 S는 스케치, I는 unsigned distance field를 의미한다. 본 논문에서는, 수식 (1)을 통해 구해진 unsigned distance field에, 전역적인 정보와 국소적인 정보 사이의 비중을 조정하기 위해 scaling factor를 도입하였다. Scaling factor는 unsigned distance field의 각 픽셀에 지수적으로 적용되고, Scaling factor로 각 픽셀을 조정된 후에 픽셀값이 0부터 1까지의 값을 가지도록 정규하였다. (이하 scaling factor가 적용된 unsigned distance field를 “scaled distance field”라고 명명)

$$I_{s,i} = I_i^d / I_{s,max} \dots \text{수식 (2)}$$

I_s : Scaled distance field

$I_{s,i}$: I_s 의 p_i 픽셀 강도

$I_{s,max}$: I_s 의 픽셀 강도 최댓값

(d : scaling factor, $d \in \mathbb{R}$, $d > 0$)

Scaling factor에 따른 scaled distance field 변화는 그림 (1)과 같다. Scaling factor가 작으면 세부적인 정보보다는 항공기 전체 형태에 관한 정보가 남고, scaling factor가 커지면 항공기 창문과 같은 세부적인 정보가 보존되는 것을 볼 수 있다. 기존 unsigned distance field는 scaling factor가 1일 때에 해당한다.

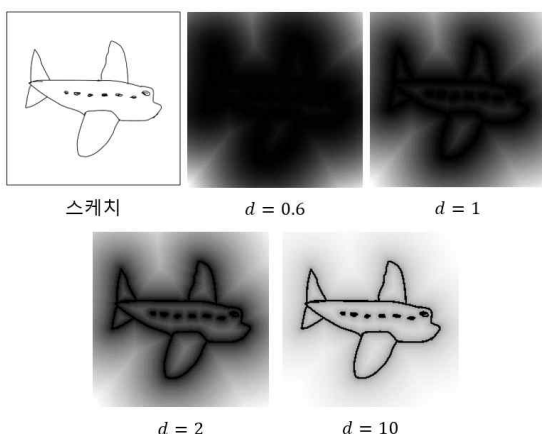


그림 1. Scaling factor에 따른 scaled distance field 변화

2.2. 학습 네트워크 구조 및 손실 함수

스케치를 Scaled distance field로 변환하여 클래스 분류 네트워크의 입력으로 사용하였다. 먼저, scaled distance field로부터 특성을 추출하기 위해 Resnet-34[4]를 적용하였다. 그리고 완전연결신경망의 출력 차원을 데이터셋에 맞게 변형하였고, 손실함수는 다중-레이블 클래스 분류에 많이 사용되는 크로스 엔트로피 손실함수를 사용하였다.

$$L_{CE} = - \sum_{i=1}^c y_i \log(\hat{y}_i) \dots \text{수식 (3)}$$

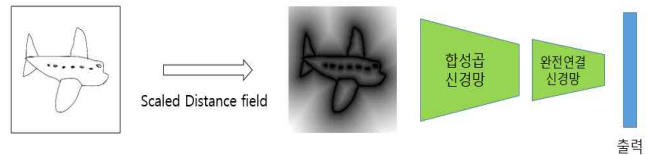


그림 2. 스케치 클래스 분류 알고리즘

수식 (3)에서 c는 분류하고자 하는 클래스의 개수를 나타낸다. \hat{y}_i 는 완전연결신경망을 통과한 벡터에 softmax를 적용시킨 예측값인데, 아랫첨자는 벡터에서 몇 번째 요소인지 의미한다. y_i 는 정답 클래스를 나타낸다. 본 논문에서 설계한 알고리즘을 정리하면 그림 (2)와 같다.

3. 데이터셋

본 논문에서 제안한 알고리즘을 실험하기 위한 데이터셋으로, 스케치 기반 활용 분야에 널리 사용되는 벤치마크 데이터셋인 TU-Berlin 스케치 데이터셋[5]을 사용하였다. TU-Berlin 데이터셋은 250개의 클래스로 구성되어 있고, 각 클래스별로 80개의 이미지가 있어서 총 20000장의 스케치로 구성되어 있다. 본 논문에서는 각 클래스에 속한 스케치를 4:1로 분할하여 각각 학습 및 테스트에 사용하였다.



그림 3. TU-Berlin 데이터셋 예시

4. 실험 결과 및 분석

Scaling factor를 변화시키면서 얻은 다양한 scaled distance field를 스케치 클래스 분류 네트워크에 통과시킴으로써, 각 scaling factor에 따른 성능을 측정하여 비교하였다. 클래스 분류 성능은, 네트워크 출력으로 나온 결과 벡터 중 가장 큰 값을 가지는 인덱스를 네트워크가 예측한 클래스로 하였을 때, 테스트 셋에 대한 예측 클래스 중 정답이 어느 정도 되는지 퍼센트로 측정하였다. 그리고 네트워크 입력으로써 scaled distance field를 사용했을 때, 스케치를 사용했을 때보다 어떤 장점이 있는지 분석하였다.

4.1. Scaling factor 변화에 따른 분석

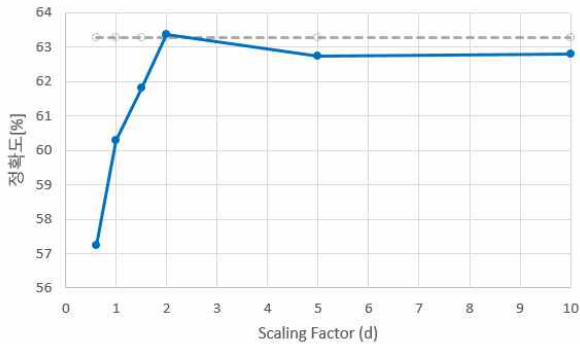


그림 4. Scaling factor에 따른 스케치 클래스 분류 실험 결과. 회색 점선은 네트워크 입력에 스케치 자체가 사용되었을 때의 결과를 나타냄.

Scaling factor	0.6	1	1.5	2	5	10	스케치
정확도 [%]	57.241	60.282	61.799	63.363	62.734	62.798	63.27

표 1. Scaling factor에 따른 스케치 클래스 분류 실험 결과

Scaling factor를 0.6부터 10까지 변화시키면서 클래스 분류 정확도를 측정하였다. Scaling factor가 작을 때 국소적인 정보는 대부분 사라지고 전체적인 형태와 같은 전역적인 정보만 남게 되는데, 그 정도가 지나치면 오히려 클래스 분류 성능이 떨어지는 것을 관찰할 수 있다. 그리고 기존 unsigned distance field보다 scaling factor가 클 때, 성능이 더 향상됨을 알 수 있다. 특정 scaling factor 이상부터는 성능에 뚜렷한 변화가 있지 않았지만, 본 논문에서는 scaling factor가 2일 때를 기준으로 스케치 자체가 네트워크 입력으로 사용되었을 때와 비교하였다. 이 두 가지 경우 사이에 성능이 유의미하게 차이가 나지 않았지만, 네트워크 학습 안정성은 고밀도 데이터인 scaled distance field가 입력으로 사용되었을 때 더 우수했다. 특히 학습 배치가 작을수록, 학습에 사용되는 네트워크가 작을수록 학습 안정성에서 큰 차이가 났다.

4.2. Smooth Grad-CAM++을 이용한 분석

Scaling factor가 2일 때의 scaled distance field와 스케치가 네트워크 입력으로 들어갔을 때, 합성곱 신경망이 입력의 어떤 부분을 보고 클래스 분류를 수행하는지 Smooth Grad-CAM++[6]을 통해서 분석하였다.

그림 5에서 볼 수 있듯이, 입력으로 들어온 이미지를 양초와 파이프라고 판단하면서 scaled distance field가 입력으로 들어왔을 때 네트워크가 더 사물에 집중하여 판단한다는 것을 알 수 있다. 예를 들면 양초의 경우, 스케치가 입력으로 들어갔을 때 물체가 아니라 주변 공간에 집중하는 것을 볼 수 있지만, scaled distance field가 입력으로 들어갔을 때 양초에 집중하고 것을 볼 수 있다.

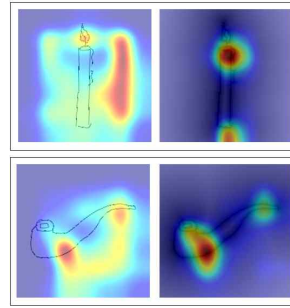


그림 5. Smooth Grad-CAM++로 분석한 네트워크의 클래스 판단 기준(좌: 스케치, 우: Scaled distance field). Scaled distance field가 입력으로 사용되었을 때, 네트워크가 더욱 사물에 근거하여 클래스 분류를 한다는 것을 알 수 있음.

5. 결론

본 논문에서는 희소 데이터인 스케치를 고밀도 데이터인 unsigned distance field로 변환하고, unsigned distance field scaling factor를 도입하여 전역적인 정보와 국소적인 정보를 조정하는 방법을 제시하였다. 이를 바탕으로 너무 전역적인 정보(작은 scaling factor)만을 담고 있는 입력보다는, 국소적인 정보(큰 scaling factor)도 보존하는 입력이 사용되었을 때, 네트워크가 더 유의미한 특성을 추출한다는 것을 알 수 있었다. 그리고 scaled distance field가 입력으로 들어갔을 때 학습 안정성이 더 보장되며, 네트워크가 더 합리적인 근거로 클래스 분류를 진행함을 확인할 수 있었다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2021R1A2C2007220).

참고 문헌

- [1] Hua Zhang et al, "SketchNet: Sketch Classification with Web Images", Conference on Computer Vision and Pattern Recognition(CVPR), 2016
- [2] Li Liu et al, "Deep Sketch Hashing: Fast Free-hand Sketch-Based Image Retrieval", Conference on Computer Vision and Pattern Recognition(CVPR), 2017
- [3] Wengling Chen et al, "SketchyGAN: Towards Diverse and Realistic Sketch to Image Synthesis", Conference on Computer Vision and Pattern Recognition(CVPR), 2018
- [4] Kaiming He et al, "Deep Residual Learning for Image Recognition", Conference on Computer Vision and Pattern Recognition(CVPR), 2016
- [5] M. Eitz et al, "How do humans sketch objects?" ACM Trans. Graph., 31(4):44-52, 2012.
- [6] Daniel Omeiza et al, "Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep Convolutional Neural Network Models", Intelligent Systems Conference(IntelliSys), 2019